Interests and epistemic integrity in science.
A new framework to assess interest influences in scientific research processes.

Abstract

Non-epistemic interests (e.g., financial interests, political interests) sometimes influence scientific decisions (e.g., hypothesis acceptance, theory choice). For instance, pharmaceutical giant Merck manipulated clinical trial data in order to make sure that these data confirmed the safety of one of its products (Vioxx), as this served the company's short-term commercial interests. The latter is obviously unacceptable. But why exactly is it unacceptable? One way to account for this is on the basis of the full ideal of purity. According to this ideal, scientific decision-making should be pure, i.e. unaffected by non-epistemic interests. Although this ideal is questionable in light of earlier philosophical work, some philosophers of science still hold on to it, or to a less strict version of it. In part 1 of this dissertation, it is argued that it is better to fully abandon the ideal of purity. In part 2, an alternative ideal to assess interest influences in science is proposed: the ideal of epistemic integrity. A new concept of epistemic integrity is spelled out and systematically defended. Furthermore, the new concept is not only used to analyze the Vioxx debacle, but also to identify unacceptable interest influences in aerospace science and climate science, and to explain why exactly these interest influences are unacceptable.

Contact:
jan.dewinter@ugent.be
Centre for Logic and Philosophy of Science, Department of Philosophy and moral sciences, Ghent University, Blandijnberg 2, 9000 Ghent, Belgium

Nederlandse vertaling:
Belangen en epistemische integriteit in wetenschap

Verschijningsdatum:
19/11/2014

Administratief verantwoordelijk promotor:
Prof. dr. Erik Weber, Universiteit Gent

Promotor:
Dr. Jeroen Van Bouwel, Universiteit Gent

Doctorale begeleidingscommissie:
Prof. dr. Hans Radder, Vrije Universiteit Amsterdam
Prof. dr. Maarten Van Dyck, Universiteit Gent

Leescommissie:
Prof. dr. Tom Claes, Universiteit Gent
Dr. Rogier De Langhe, Universiteit Gent
Prof. dr. Hans Radder, Vrije Universiteit Amsterdam
Prof. dr. Julian Reiss, Durham University
Prof. dr. Maarten Van Dyck, Universiteit Gent

Decaan:
Prof. dr. Marc Boone, Universiteit Gent

Rector:
Prof. dr. Anne De Paepe, Universiteit Gent

Faculteit Letteren & Wijsbegeerte

Jan De Winter

Interests and epistemic integrity in science

A new framework to assess interest influences in scientific research processes

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

2014

Preface

A few years ago, I read a renowned article written by Justin Bekelman, Yan Li, and Cary Gross, called 'Scope and impact of financial conflicts of interest in biomedical research: A systematic review'. The article reveals a statistically significant correlation between industry sponsorship of biomedical research and pro-industry conclusions (Bekelman, Li & Gross 2003). I started wondering where this correlation comes from. A first, obvious explanation was that pharmaceutical companies manipulate research and/or its results in order to arrive at favorable conclusions, that is, conclusions that can help them sell their products and make a profit. In that case, we should seriously worry about the fact that the financial interests of pharmaceutical companies increasingly permeate biomedical science (e.g., Bekelman, Li & Gross (2003) report that industry's share of total investments in biomedical research grew from 32% in 1980 to 62% in 2000).

But it is also possible that the correlation is due to pharmaceutical companies being better than government-funded agencies at detecting in early stages of research the most promising drugs for treating a certain condition. Pharmaceutical companies would then be better than government-funded agencies at avoiding expensive clinical trials for inferior drugs, and they would mainly conduct research on drugs that will turn out to be most effective and safe. This is an alternative explanation for the finding that industry-sponsored research is more likely to draw favorable conclusions. If this would be the right explanation, then the growing influence of industry and its financial interests on biomedical research would not be a problem. On the contrary: it would make the biomedical enterprise more efficient, as it would be able to deliver better medicines for less money (due to resources being primarily allocated to research on the most valuable drugs).

Depending on which explanation we accept, the conclusion is different. In fact, the two explanations lead to opposite conclusions: 'financial interests corrupt biomedical science' versus 'financial interests advance biomedical science'. I considered the tension between these two conclusions to be an instance of a more general tension: the tension between the idea that science should not be affected by financial interests, political interests, career interests, etc. on the one hand, and the idea that such interests play a valuable role in science on the other hand.

Aroused by the desire to develop a more advanced, consistent picture of interests in science, I started thinking about a number of questions relating to this topic. How do different kinds of interests influence different kinds of decisions in science? Which of these influences are valuable and why? Do interests in science cause certain problems, and if so, which problems exactly? What are the best ways to deal with these problems? Because financial interests become more and more important in science, and also because our dependence on scientists' expertise increases, these questions are today more relevant than ever. Therefore, I have searched for answers to these questions in the past five years. The main results of that quest are presented in this dissertation.

Before I start, I would like to add a word of thanks to all those who have contributed. Firstly, I am grateful to the Research Foundation – Flanders for financially supporting my research, first through research project 3G003109 (from January to September 2009), and from October 2009 to September 2013 through a Ph.D. fellowship. Secondly, I owe special thanks to the supervisor of my doctoral research, Erik Weber, and the co-supervisor, Jeroen Van Bouwel, for the opportunities they gave me, for guiding me through my research, and for their useful advice and valuable comments on this dissertation. Thirdly, I want to thank all others who have offered helpful comments on earlier versions of (parts of) this dissertation: Rogier De Langhe, Dries De Winter, Laszlo Kosolosky, Bert Leuridan, Hans Radder, Julian Reiss, Sigrid Sterckx, Maarten Van Dyck, and several anonymous reviewers of journals. Also thanks to several people at conferences and workshops for suggestions and discussions. For the design of this dissertation, I thank Gitte Callaert. Finally, I want to thank the members of the reading committee: Tom Claes, Rogier De Langhe, Hans Radder, Julian Reiss, and Maarten Van Dyck.

Contents

Chapter 1
Introduction

1.1 Conceptual framework

1.1.1 The concept of interests

A first thing to be done in a dissertation on interests in science, is a clarification of the term 'interests'. What does this term stand for? Let me propose a characterization of interests that is taken for granted throughout this dissertation.

An interest can be thought of as a relation between an agent (which could be an individual, a social group, an institution, etc.) and a potential state or circumstance, more specifically, the following relation: the realization of the potential state is by the agent regarded as beneficial in one way or another. For instance, when I speak of the health interests of a patient, I refer to potential health states that the patient considers beneficial. When I speak of the financial interests of a company, I refer to potential financial circumstances that the company considers beneficial. An action serves an interest if it contributes to the realization of the potential state that is regarded as beneficial.

Note that interests differ from goals. In fact, goals form a subcategory of interests; all goals are interests, but not vice versa. A goal requires that there is a potential state or circumstance of which the realization is regarded as beneficial in one way or another, and therefore it is an interest. But more is required: for an interest to be a goal, effort must be directed towards the realization of the potential state or circumstance. Hence, not all interests are goals in my terminology. If no effort is made to realize the relevant state or circumstance, then this is not considered a goal.

A category of interests that is especially relevant for the sciences is the category of interests in knowledge, or epistemic interests. An epistemic interest is here understood as an interest concerning the potential state of knowing something, which is regarded as beneficial in one way or another. These are interests that usually (partly or fully) motivate scientific inquiry: someone wants to know something – whether it is out of sheer intellectual curiosity or because such knowledge will serve other (i.e. non-epistemic) interests – and this causes him to pursue a certain line of research. The results of the research are supposed to serve the person's epistemic interest; they are supposed to contribute to the realization of the potential state of knowledge wished for.

Of course, scientific research could also have unexpected results. It may not deliver the knowledge wished for, but instead, say, the finding that such knowledge cannot be attained as easily as initially expected. In that case, the research does not serve the epistemic goal that motivated it. This does, however, not mean that the research does not serve any epistemic interests. If the unexpected result (i.e. the finding that the desired knowledge cannot be attained as easily as initially expected) is considered valuable knowledge as well, then the research serves one's interest in *such* knowledge, even though serving this interest is not what the research originally aimed at. The interest served is then a post factum interest, that is, an interest which is recognized after it is served.

The distinction I draw between epistemic and non-epistemic interests does not entail that they can always be strictly separated. My framework allows epistemic interests to be closely connected to non-epistemic interests. This is in line with a point made by Philip Kitcher. As Kitcher has indicated, scientists do not just aim at discovering plain truth:

> Nobody should be beguiled by the idea that the aim of inquiry is merely to discover truth, for, as numerous philosophers have recognized, there are vast numbers of true statements it would be utterly pointless to ascertain. The sciences are surely directed at finding *significant* truths. (Kitcher 2001, p. 65)

Kitcher distinguishes two kinds of significance: practical and theoretical significance.[1] A true statement is *practically* significant if knowing it increases the probability of attaining practical goals. A true statement is *theoretically* significant if knowing it is intrinsically valuable (Kitcher 2001, p. 65). So the sciences are, in Kitcher's view, directed at generating knowledge that is either useful for practical goals, or intrinsically valuable, or both. In my terminology: in the sciences, there is an epistemic interest in knowledge that is either useful for non-epistemic interests, or intrinsically valuable, or both.

[1] Kitcher also uses the term 'epistemic significance' as a synonym for 'theoretical significance' (see Kitcher 2001, p. 65).

This points at a potential connection between epistemic and non-epistemic interests. Certain kinds of knowledge may be considered valuable – and are therefore pursued in the sciences – (partly) because of their usefulness for non-epistemic interests. Hence, the epistemic interest in such knowledge may (partly) stem from the non-epistemic interests that such knowledge would serve.

I should note that not only the public non-epistemic goals that most people consider acceptable in science, such as the aim to promote public health, may motivate epistemic interests in science, but also more controversial non-epistemic interests, such as the financial interests of a company, or political interests. Such interests are often served by particular kinds of knowledge – for instance, the financial interests of a pharmaceutical firm are served by scientific evidence suggesting that one of its products is safer or more efficient than competing drugs – and this may cause the potential beneficiary to be interested in such knowledge (= epistemic interest).

1.1.2 The concept of values

A second concept that I have to delineate is the concept of values. This is because I will use the existing literature on values in science in order to reveal certain interest influences in science. The concept of values is closely related to the concept of interests, but the two are not the same. Let me offer a characterization of values that I take for granted throughout this dissertation.

Values differ from interests in that they are not relations between agents and states, but properties of entities. More specifically, a property is a value for an entity of a certain kind when it is regarded as beneficial that an entity of that kind has the property. For instance, speed is a value for a race car since it is regarded as beneficial that a race car is fast. Fairness is a value for society if we regard it as beneficial that society is fair. Empirical adequacy is a value for a theory if we consider it beneficial that a theory is empirically adequate.[2] I start from this characterization of values because (1) I think it captures quite well what is meant by this concept in the philosophical literature that I will use, and (2) it is appropriate for my purposes. It is inspired by a characterization offered by Ernan McMullin: "A property or set of properties may count

[2] Note that this characterization involves a relation between agents and properties: an agent regards it as beneficial that an entity of a certain kind has the relevant property. But that does not mean that a value is a relation between an agent and a property. The relevant property is the value, not the relation between the property and the agent. For example, if Peter regards it as beneficial that a race car is fast, then speed is the value in this case, and Peter holds this value. The value is not the relation between Peter and speed.

as a value in an entity of a particular kind because it is desirable for an entity of that kind" (McMullin 1983, p. 5).

It should be remarked that people can disagree on whether or not a certain property is a value for a certain kind of entity; people can hold different values. For instance, while one person may consider diversity a value for society, someone else may prefer a homogeneous society.

A second remark is that entities can often have a certain property to a higher or lower degree. Speed, fairness, empirical adequacy, and diversity are all properties that come in degrees; a race car can be faster or slower, a society can be more fair or less fair, and more diverse or less diverse, and a theory can be more empirically adequate or less empirically adequate. Usually, when properties that come in degrees are accepted as values for the relevant kinds of entities, then it is regarded as beneficial that the entities have the properties to a degree that is as high as possible, and the higher the degree, the better.

If we take the example of the race car, one may have the following objection: as a race car gets faster, it becomes harder to control, and so a faster race car is not necessarily a better one. This objection does, however, not imply that it is not regarded as beneficial that a race car is as fast as possible, but only that the value of speed can conflict with other values, such as the controllability of the car. If scoring higher on the value of speed would not imply scoring lower on the value of controllability, then the faster race car *would* be considered the better one. So when I say that usually it is regarded as beneficial that an entity has a value to a degree that is as high as possible, then I mean: usually it is regarded as beneficial that an entity has a value to a degree that is as high as possible, *all other things being equal*.

The reason why I say 'usually', is that this is not necessarily *always* the case. For instance, someone can (consistently) consider diversity a value for society, and at the same prefer moderate diversity over extreme diversity, all other things being equal. As the person prefers a society with a degree of diversity above a certain threshold degree, over a society with a degree of diversity below this threshold, he holds the value of diversity, even though he does not consider it beneficial that a moderately diverse society becomes even more diverse.

What can we now, in light of my conceptual framework, say about the potential relations between values and interests? Firstly, values can give rise to certain interests. When we consider certain properties valuable for a certain kind of entity, then this may lead to an interest in adjusting an entity of this kind so that it acquires the valuable properties. Or we may want to replace an entity that doesn't manifest certain values by one that does. For instance, if we think that a race car is too slow (that is, it does not sufficiently manifest the value of speed), we may want to tune the car in order to make it faster, or replace it by another one which is faster (that is, we want to realize the state in which we have a race car that sufficiently manifests the value of speed). It is clear that interests such as these depend on what we value; if we would not consider speed a value for race cars, we would not be interested in tuning or replacing our race car in order to have a faster one.

Secondly, values can depend on interests. For example, racing drivers value speed for race cars because they want to win races, and in order to satisfy that interest, their car needs to be as quick as possible. A car manufacturer may value speed because building a fast race car serves its commercial interests: if it can develop a race car that is remarkably fast, then more people will be prepared to buy this car, or other cars from the same manufacturer. If fast race cars would not serve these interests (e.g., because all racing drivers would have to use the same car, and potential buyers would not be affected by speed), and if speed would not serve any other interests, then the relevant agents would probably not consider speed a valuable property for race cars.

It should be noted that the influence of interests on values could be conscious as well as unconscious. If someone's valuing a certain property for a certain kind of entity is due to the fact that certain interests are better served by entities *with* the relevant property, then he may be aware of this; he may even refer to these interests in justifying why he considers the property valuable. But it is also possible that the influence of certain interests is more obscure. Suppose, for instance, that a car manufacturer thinks that it regards speed as valuable for race cars only because speed is needed to win races and that its commercial interests have nothing to do with this. Still, it is possible that the manufacturer's commercial interests are causally relevant for the fact that it considers speed a value for race cars; it is possible that once fast cars stop serving these interests (say, because speed or winning races doesn't sell anymore), the manufacturer stops holding the value of speed, or that, had these interests not been served by fast race cars from the beginning, the manufacturer would never have started valuing speed in the first place.

1.2 Main assumptions

1.2.1 Science serves non-epistemic interests

Now that I have spelled out a conceptual framework for this dissertation, I should turn to its main assumptions. A first assumption is that the sciences do not only serve epistemic interests, but also a variety of non-epistemic interests. Let me explain this assumption on the basis of several examples.

Research in the health sciences has delivered products and treatments that help us tackle various diseases. These innovations have had a major impact on human health, and still have. Due to the health sciences and its products (e.g., new medicines, new ways to diagnose disease), we can cure diseases that were incurable before, recovery processes can be accelerated, the development of certain health conditions can be prevented more effectively, and so on. This has contributed to an increase of life expectancy. It is obvious that these achievements would not have been possible, or at least not to the same degree, without the health research that preceded them (Bush 1945; Lichtenberg 2003). This means that health research has served certain non-epistemic interests, namely, people's interest to live a long and healthy life. (Living a long and healthy life is the potential state that they regard as beneficial; cf. section 1.1.1.)

Research in the agricultural sciences has delivered techniques to substantially increase food production, ranging from the use of water, machinery, fertilizers, and insecticides, to genetic engineering. Recently, attention has also been drawn to some alternative approaches in the agricultural sciences, which fall under the name 'agroecology', and which enable us to not only increase food production, but also to alleviate rural poverty and mitigate climate change (De Schutter 2011). So research in the agricultural sciences serves certain non-epistemic interests: the interest to increase food production, to alleviate rural poverty, and to mitigate climate change. (Increased food production, less rural poverty, and mitigated climate change are potential states that are regarded as beneficial.)

Social benefits such as better public health, increased life expectancy, and having sufficient food to feed the population are often used to justify public funding of science: because science delivers these benefits, the government should support it (see, for instance, Vannevar Bush's (1945) famous report to U.S. President Harry S. Truman, in which he indicates how and why the government should support science). Another interest that is often used to justify public funding of science is national security. Science enables us to develop weapons and techniques that are very useful to protect the nation from enemy attacks, and this is a reason to allocate government resources to such research (Bush 1945).

Serving social interests is not limited to the health sciences, the agricultural sciences, and military science; other sciences serve certain social interests as well. Think, for instance, of sociology, which enables us to predict and anticipate the social consequences of certain conditions or social measures, meteorology, which delivers more or less accurate weather forecasts, or seismology, which offers estimates of the probability of an earthquake at a particular location at a particular time. All these things are considered valuable to society.

One may have the impression that this very brief overview of some of the fruits of science only concerns the applied sciences, that is, sciences that aim at solving practical problems in medicine, agriculture, engineering, etc., but not the fundamental sciences, that is, sciences that aim at the advancement of knowledge without the prospect of an immediate practical use (e.g., biology, chemistry, physics). However, the latter sciences serve a variety of non-epistemic interests as well. They could contribute to the aforementioned social benefits by delivering knowledge that is used in applied research. Applied research often builds on the knowledge generated in the fundamental sciences, and as such, the latter indirectly serve the social interests that applied research serves (better public health, national security, etc.).

An example of fundamental research that was (partly) justified on the basis of social interests is the Human Genome Project, which aims at mapping and sequencing the genome of our species. This project would bring improvements in diagnostic testing and partial gains in coping with certain diseases, and by enhancing understanding of basic biology, it would *possibly* contribute to significant breakthroughs in the treatment or prevention of diseases in a distant future. These potential medical benefits were emphasized in public defense of the Human Genome Project (Kitcher 2001, pp. 4-5).

Scientific research also serves private interests of those involved in it. Research and development (R&D) of commercial products delivers profits to the private companies that have invested in it. Nowadays, universities, too, can make money from commercial research. Several countries have enacted legislation that permit universities to obtain patents on products that are developed using government funding (Siepmann 2004; Mowery & Sampat 2005, p. 123), and these patents enable them to demand compensations from anyone who wants to use the patented products. Hence, R&D of commercial products can serve not only the financial interests of private companies, but also of universities. Another private interest that research serves, is the objective of universities and researchers to maintain/obtain government funding in the future. Research delivers products such as publications in scientific journals and academic prizes, and these can be used to claim future research funding from the government. Researchers also pursue such products in order to be eligible for a promotion or tenure.

1.2.2 Interests influence science

A second assumption of this dissertation is that serving interests is not just an *ad hoc* aspect of science, something which happens at the end of the research process by accident. I assume that from the beginning of the scientific process, serving certain interests is anticipated by those supporting and performing the research. I have already mentioned that Vannevar Bush referred to social interests in justifying public funding for science. This justification was the basis of the establishment of the National Science Foundation, a U.S. government agency that supports fundamental research. Hence, the prospect that research would serve certain interests is what motivated the government to allocate resources to it. Private companies, too, fund research because they expect it to serve certain interests, typically, the financial interests of their owners or shareholders. They invest in research because they believe this will deliver financial rewards that exceed their investments in research, so that profits can be made. Researchers themselves are driven by interests as well – whether it is sheer intellectual curiosity, the desire to make a difference in the world, or the desire to gain status; most, if not all of them, perform research because of the interests they hope it to serve.

Interests do not only influence decisions on whether or not to support or perform research, but also decisions on *which* research is supported/performed. Private companies select the research that they expect to meet consumer wishes and deliver the highest profits (and thus serve their financial interests best), non-governmental organizations select the research that they expect to contribute most to the realization of the organization's goals, and academic researchers' decisions on which research to pursue are based on, e.g., which topics they are most curious about, which lines of research they expect to deliver the results that are most beneficial to society or most interesting to the academic community, or which research they expect to contribute most to the advancement of their careers.

Once the research has started, interests *continue* to play an important role. Epistemic and non-epistemic interests affect reasoning processes within scientific research. They play in decisions on whether or not to accept a certain hypothesis, in the process of choosing between rival theories, and in the construction and evaluation of explanations. I admit that this claim may be more controversial than my claims on the role of interests in the phases preceding research (see above): while most people will accept that non-epistemic interests play an important role *before* research (in decisions concerning the allocation of resources), the claim that they also play an important role *during* research may be more controversial. The idea that scientific inquiry *is*, or at least *should be* pure, in the sense of being independent of factors such as the ideological or financial interests of the researcher(s), seems deeply rooted in the minds of a lot of people (Proctor 1991, pp. 3-4). Because my view conflicts with this idea, I support my view in part 1 of this dissertation. As such, I hope to eradicate the idea that scientific research should be free from non-epistemic interests. I will come back to this in section 2.3, where I present the aims of this dissertation.

Now consider the phases *after* research. Once research is finished, the results are usually communicated to others. Interests influence which results are communicated to whom. An example is when academic researchers communicate their findings to the academic community (typically by publishing articles in academic journals) but not to the broader public because the former contributes to their status in the academic community and serves their career interests, while informing the broader public does not serve these interests, or at least not to the same degree. Not only the interests of the knowledge providers have an influence on who is informed, but also the interests of potential audiences. Suppose there is a question that only a small community of scientists is curious about – maybe outsiders do not even understand the question – and that a member of this community develops an answer to this question. The fact that only the members of the community are interested in such an answer, and that the answer does not serve the interests of outsiders, at least not directly, can then be a reason to only inform the former group. So who is informed and who is not depends on the interests of the informer and/or his potential audiences. But interests can also cause someone to actively keep the results of one's research hidden for certain groups. Trade secrets are a well-known example: companies keep a certain design or process secret (typically by having their employees sign an agreement that they will not inform outsiders of the design or process) in order to have a competitive advantage over rival companies.

Finally, consider the application of scientific research results. Here, the story is more or less similar to the selection of research projects: private companies select the applications that they expect to meet consumer wishes and deliver the highest profits, non-governmental organizations select the applications that they expect to contribute most to the realization of the organization's goals, and governments apply research results in ways that they think will serve the interests of their citizens best, or of the specific groups they represent, or they choose certain applications because they expect these to serve their own interests best (e.g., by attracting the most potential voters).

1.2.3 Troublesome interest influences

So interests can play in different phases of the scientific process. While this is often acceptable or even desirable, there are also cases in which a particular interest influence is troublesome. This is a third assumption of this dissertation. I borrow an example from Biddle (2007) to illustrate this.

When pharmaceutical giant Merck reported the results of research on the painkiller Vioxx, it didn't mention all cardiovascular events in the groups taking Vioxx. Take, for instance, the large VIGOR study (Vioxx Gastrointestinal Outcomes Research), in which Vioxx was compared with naproxen with respect to gastrointestinal outcomes. A paper on this study was published in the New England Journal of Medicine. The paper mentioned 17 heart attacks in the Vioxx group. But later, it turned out that data relevant to cardiovascular risks had been deleted from the final draft of the VIGOR study. Among these data were three additional heart attacks that occurred in the Vioxx group (so actually, there were 20 heart attacks in the Vioxx group). At least two of the authors of the paper were aware of the three additional heart attacks before the paper was submitted to the New England Journal of Medicine (Biddle 2007, p. 28).

A similar problem occurred with respect to another study on Vioxx, known as 'Advantage'. In the Advantage study, Vioxx was also compared with naproxen. A paper on this study that was written by Merck employees[3] and published in Annals of Medicine, stated that five patients suffered from heart attack or sudden cardiac death in the Vioxx group, as compared to one in the naproxen group. This difference was not statistically significant. But at least one, and possibly three more deaths in the Vioxx group resulted from cardiovascular problems. If one of these deaths had been classified as resulting from cardiovascular problems, then the difference in cardiovascular events in the Vioxx group and the naproxen group would have been statistically significant (Biddle 2007, p. 29).

Biddle attributes the inadequacies in Merck's characterization and reporting of data to its short-term financial interests (Biddle 2007, p. 30). Merck was the owner of Vioxx. If research would show that there were significant cardiovascular risks attached to Vioxx, this would damage the sales of this drug, and thus Merck's revenues from it. To make sure that this would not happen, the data on cardiovascular events occurring in Vioxx groups were altered. In that way, Merck's (short-term) financial interests were protected.

The particular influence that interests had in this case on research/communication is clearly troublesome. The data should not have been altered in order not to jeopardize the sales of Vioxx. Merck's short-term financial interests caused these illegitimate alterations of data, and so the influence of these interests is problematic.

1.3 Two challenges for philosophy of science

My conceptual framework and assumptions present two challenges for philosophy of science. A first challenge is to describe and analyze the different roles that different kinds of interests play in science. We have seen that there are various interest influences in the scientific process. Philosophers of science should study these influences and make them explicit. This is the first challenge for philosophy of science.

In part 1 of this dissertation, I will contribute to meeting this challenge. More specifically, part 1 discusses the roles of interests in reasoning processes within scientific research (and so it doesn't concern the phases before and after research). It

[3] Jeffrey R. Lisse of the University of Arizona is the lead author of the paper, but he was paid to put his name on it (Biddle 2007, p. 29).

shows how different kinds of interests affect the acceptance of hypotheses, theory choice, and the evaluation of explanations in science. One of the main purposes is to show that non-epistemic interests play a more prominent role in reasoning processes within scientific research than most people tend to think. I will do this by attacking two versions of the ideal of purity: (1) the full ideal of purity, which states that ideally, non-epistemic interests do not play a role in scientific decisions such as hypothesis acceptance and theory choice, and (2) the partial ideal of purity, which states that ideally, non-epistemic interests do not play a *direct* role in such scientific decisions. My criticism on the latter version will be the most extensive one, since it is more controversial. Together, my criticisms reveal that non-epistemic interests form an integral part of scientific reasoning.

Once we acknowledge this, a second challenge arises: to distinguish *acceptable* interest influences in science from *unacceptable* interest influences in science. The Vioxx case shows that the influence of certain interests on the research process can sometimes be troublesome. Philosophers of science should specify when it is troublesome and when it is not. For those particular interest influences that are regarded as problematic, it should be clarified *why* they are problematic. This is the second challenge for philosophy of science.

The aim of part 2 of this dissertation is to contribute to meeting the second challenge. The focus is on *epistemically* problematic interest influences, and so I will not consider interest influences that are only for non-epistemic reasons problematic. An example of an epistemically problematic interest influence is the interest influence in the Vioxx case discussed above. Biddle (2007) characterizes the inadequacies in Merck's characterization and reporting of data as epistemic failings. He does, however, not sufficiently explain *why* exactly they are epistemic failings. He claims that "the research was organized in such a way as to allow short-term commercial interests to compromise epistemic integrity" (Biddle 2007, p. 21). But he does not explicate what he means by 'epistemic integrity', nor does he clarify why short-term commercial interests compromised it.

In part 2 of this dissertation, an explication of the concept of epistemic integrity is developed, and I clarify why epistemic integrity was compromised by short-term commercial interests in the Vioxx case. Several other examples of interests compromising epistemic integrity are discussed as well. In these discussions, my concept of epistemic integrity is applied to examples from aerospace science and climate science. The purpose is to show how this concept can be used (1) to identify epistemically problematic interest influences in the sciences, and (2) to explain *why* these interest influences are epistemically problematic.

1.4 Overview of this dissertation

Now that I have touched on and clarified some topics and questions that will be addressed in this dissertation, let me offer a more systematic overview of the parts and chapters in it. The first part deals with the question which roles different kinds of interests play in reasoning processes within scientific research. Chapter 2 starts by presenting a common view on this matter: the full ideal of purity. A criticism on this view is developed on the basis of Richard Rudner's argument that the scientist qua scientist makes value judgments. Next, I present a less strict version of the ideal of purity – the partial ideal of purity – which avoids my criticism on the basis of Rudner's argument. Furthermore, I discuss some insights from the existing philosophical literature that challenge the ideal of purity. It is argued that while these insights may successfully undermine the *full* ideal of purity, they do not undermine the *partial* ideal of purity.

Nevertheless, I think the partial ideal of purity is untenable as well. The remainder of part 1 aims to show this. Chapters 3-5 reveal that the scientific explanation of phenomena is often affected by interests in ways that have not been considered in chapter 2. The focus of chapters 3-4 is on explanation in technological science, that is, sciences that have human-made, rather than natural objects as their ultimate study object (Hansson 2007). I discuss the role of interests in the development and evaluation of mechanistic artifact explanations (explanations that account for an artifact behavior by describing the underlying mechanism) in chapter 3, and chapter 4 reveals how epistemic and non-epistemic interests operate in the development and evaluation of explanations in software engineering. Chapter 5 shows that the kinds of interest influences identified in chapters 3-4 are not unique to technological science. Similar interest influences can be found in other sciences. More specifically, chapter 5 reveals how epistemic and non-epistemic interests can play in the explanation of phenomena in the social sciences, biology, and the medical sciences. Chapter 5 also clarifies how the findings of chapters 3-5 exactly undermine the partial ideal of purity.

The second part aims to develop a criterion to determine when interest influences in science become epistemically problematic. A concept of epistemic integrity that can function as such a criterion is developed in chapter 6. Several possible definitions of this concept, some of which are inspired by earlier writings on research integrity, are discussed. A final, new definition of epistemic integrity is shown to avoid the shortcomings of the other definitions. It forms the basis for chapters 7-10.

In chapter 7, I show that my concept of epistemic integrity is a good explication of 'epistemic integrity of the research process'. I do this by arguing that it meets Rudolf Carnap's four requirements of explication: (1) it is similar to our common sense notion of epistemic integrity of the research process, (2) it is exact, (3) it is fruitful, and (4) it is as simple as requirements (1), (2), and (3) permit. It is in chapter 7 that my concept of epistemic integrity is applied to the Vioxx case. I show how this concept can be used to explain why the influence of short-term commercial interests is epistemically problematic in the Vioxx case.

The purpose of chapters 8-9 is to reveal how the concept of epistemic integrity can be used to identify epistemically problematic interest influences in different fields of science. In chapter 8, the concept of epistemic integrity is applied to a case from aerospace science: the U.S. Space Shuttle Program. It is demonstrated that the epistemic integrity of different kinds of research processes in the Space Shuttle Program was damaged. I explain this damage by the financial and career interests of those involved in the research processes; epistemic integrity may have been compromised by these interests. In that way, the particular influence of these interests is shown to be epistemically problematic. Chapter 9 discusses the epistemic integrity of research processes in climate science. The focus is on the Intergovernmental Panel on Climate Change (IPCC). Two controversies concerning the work of the IPCC are analyzed in terms of epistemic integrity. More specifically, the epistemic integrity of the work by the IPCC that has been the target of criticism is assessed. Where epistemic integrity was compromised, an explanation in terms of the non-epistemic interests of the persons involved is offered, suggesting that non-epistemic interests influenced the relevant research process in an epistemically problematic way.

Finally, in chapter 10, I summarize some of the main results of this dissertation. I clearly indicate how it contributes to addressing the two challenges presented in section 1.4. Some routes of further research needed to fully tackle these challenges, are described.

Part 1

THE ROLES OF INTERESTS WITHIN SCIENTIFIC RESEARCH

Chapter 2
The ideal of purity

This chapter starts with a standard, common view on the roles of non-epistemic interests within scientific research, which I call the full ideal of purity (section 2.1). While this view seems mainly common among non-philosophers, I show that there are also some philosophers of science whose views correspond with the full ideal of purity. I criticize this ideal by building on an argument which is known as Rudner's argument (section 2.2). Next, a less strict ideal for science is described: the partial ideal of purity (section 2.3). This less strict ideal avoids the objection developed in section 2.2. Furthermore, I argue that insights from the existing literature on values in science, which challenge the full ideal of purity, do not undermine the partial ideal of purity (section 2.4). Nevertheless, the latter ideal should be rejected as well I think, as I will argue in chapters 3-5. In those chapters, a new criticism on the partial ideal of purity will be developed. These discussions should result in a more accurate picture of interest influences in reasoning processes within scientific research, which will be used as a basis for part 2. But first, let me describe the common view which I will attack.

2.1 The full ideal of purity

It is often stated that certain decisions in science – decisions such as whether or not one accepts a certain hypothesis, which theory one chooses out of a set of rival theories, how one explains a certain phenomenon, etc. – should in principle not be affected by non-epistemic interests such as political interests or financial interests. Rather, these decisions should be made on the basis of the degree to which the relevant hypothesis/theory/explanation manifests truth-seeking values such as empirical adequacy and internal consistency. The influence of non-epistemic interests on such scientific decisions is regarded as undesirable; something to be reduced as much as possible. I call this view the full ideal of purity (Proctor 1991, pp. 3-4). It states that ideally, scientific decisions are pure, in the sense of being unaffected by non-epistemic interests.

This is the view of a lot of people. Let me give an example of an article in which such a view is expressed. 'Funding food science and nutrition research: Financial conflicts and scientific integrity' is an article which was published in the journals 'Nutrition Reviews' and 'The American Journal of Clinical Nutrition'. It is an article which has 16 authors, and it was written for the International Life Sciences Institute North America Working Group on Guiding Principles. The following passage appears in the article:

> It has been said that "scientific 'truth' is the primary aim that all should pursue in the jungle of academic-industry interactions" (1). The point of scientific endeavor, in the first place, is and should be, the pursuit of truth—nothing more, nothing less—irrespective of financial or other interactions. It goes without saying that seekers of truth must not impose preconceptions on the method or result of their search: they must not have ulterior motives. (Rowe et al. 2009, p. 1285)

To put it in my terminology (see section 1.1), this passage suggests that the point of scientific endeavor is to satisfy epistemic interests, and that this quest should be independent of non-epistemic interests (financial or other). This corresponds with the full ideal of purity.

While such views are common mainly among non-philosophers (also see Douglas 2009, p. 1; Proctor 1991, pp. 3-4; Ruphy 2006, p. 192), they are less common among present-day philosophers of science. There are, however, still some philosophers of science whose view corresponds with what I have here called the full ideal of purity. Take Stéphanie Ruphy. She calls her view 'empiricism all the way down'. According to this view, scientists may use background assumptions and various constitutive values (e.g., empirical adequacy, simplicity, external consistency) to make scientific decisions. But ideally, the background assumptions should be justified on the basis of constitutive values, and the constitutive values that are not epistemically rewarding in *all* branches of science (e.g., simplicity, external consistency, novelty) should be justified on empirical grounds. The only two constitutive values that need no further justification in Ruphy's view, as their epistemic virtues are assumed to be universal, are empirical adequacy and internal consistency. She puts it as follows:

> I propose thus to distinguish between a very minimal, stabilized set of values whose cognitive virtues are universal (this set would be restricted to empirical adequacy

> and internal consistency), and a larger, unstabilized set that would include values whose cognitive virtues are not universal. Simplicity may be a good guide to empirical success in certain branches of physics at a certain stage of their development, but not in economics or biology today, where ontological heterogeneity is a better bet. Requiring external consistency is sometimes epistemically rewarding, sometimes not. And the same goes for novelty. In any case, I see no good reason why the epistemic worth of these values could not be established (by inductive arguments in particular), or challenged, on empirical grounds. *Let us then be an empiricist all the way down:* not only background assumptions can and should be criticized by appealing to constitutive values, but what counts as constitutive (i.e. cognitive) values (except of course the minimal set mentioned earlier) is also susceptible of being revised on empirical grounds. (Ruphy 2006, p. 212)

Ruphy thinks that in this way, that is, by establishing non-universal constitutive values on empirical grounds, it is in principle possible to free the appeal to constitutive values in scientific practice from political and social interests:

> What I'm contesting thus is only Longino's claim that the appeal to constitutive values in an ideal process of mutual criticism could not free itself from social or political motivations. It can. And not because there are such things as "universally applicable criterion of epistemic worth", but because the actual epistemic merit of what is taken as constitutive values by a scientific community is itself susceptible to be critically discussed on purely empirical grounds. (Ruphy 2006, p. 213).

So, in a nutshell, Ruphy's view is that the influence of political and social interests on scientific decisions can and should be eliminated by justifying the background assumptions and constitutive values used in such decision-making ultimately on purely empirical grounds. This view can be regarded a specific version of the full ideal of purity.

Another philosopher of science who endorses the full ideal of purity is Sheldon Krimsky. In 'Science in the private interest: Has the lure of profits corrupted biomedical research?', he criticizes the increasing influence of financial interests on biomedical research. One of the norms for science he uses to support this criticism is the norm of disinterestedness. He describes this norm as follows:

> It requires that scientists apply the methods, perform the analysis, and execute the interpretation of results without considerations of personal gain, ideology, or fidelity to any cause other than the pursuit of truth. Of course, this concept is highly idealistic and quite antithetical to the practice of science. Scientists are not neutral to the outcome of a study in which they may have much at stake. Scientists pose

> conjectures. Positive results may be publishable, whereas negative outcomes are generally not. Such a scenario means that scientists are not disinterested in the outcome and that they would prefer that the results support their conjecture. [...]
>
> Although scientists are not disinterested in the outcome of their investigations, they must behave as if they were disinterested. They must not allow their bias to affect how they approach their experimental inquiry and how they interpret the results. (Krimsky 2003, p. 77)

This passage reveals that Krimsky is a proponent of the full ideal of purity: according to Krimsky, scientific decision-making should be as if it were only driven by epistemic interests, and it should not be affected by non-epistemic interests.

2.2 Rudner's argument and beyond

I think the full ideal of purity is untenable. This can be shown on the basis of an argument developed by Richard Rudner. His argument is that before a hypothesis can be accepted in science, a value judgment has to be made, that is, the judgment that the evidence supporting the hypothesis is *strong enough* to warrant its acceptance (Rudner 1953). The value involved is the property to be supported by a certain kind of evidence, namely, the kind of evidence that we would regard as scientific. This is a property that comes in degrees; a hypothesis can be more or less supported by scientific evidence. The value *judgment* is then whether the hypothesis manifests the value of being supported by scientific evidence *to a sufficient degree*.

Such a judgment depends on non-epistemic interests. An example can help me clarify this. Suppose scientists consider the hypothesis that a certain toxic ingredient of a drug is not present in lethal quantity.[1] We want the scientists to only accept this hypothesis if it is supported by very strong scientific evidence. The reason is that if the scientists would wrongly accept the hypothesis, then this could cause the death of several potential users of the drug. Because the users' interest to stay alive is a much more fundamental interest than the interest, say, not to suffer from minor ailments like acne or itchiness, stronger evidence is required to accept that the toxic ingredient is not present in lethal quantity than to accept that the drug does not cause such minor ailments.

---

[1] This hypothesis is also offered as an example by Rudner (1953, p. 2).

That is basically Rudner's point, but I think the actual picture is more complex; non-epistemic interests affect hypothesis acceptance in more complicated ways. Let me clarify this on the basis of the example. The stronger the evidence should be, the more time it usually takes to gather it; more tests are needed to obtain enough evidence. If we assume then that the drug can only be brought to market once scientists have accepted that the toxic ingredient is not present in lethal quantity, then this implies that the more evidence is required, the longer it takes to bring the drug to market. This could damage the health interests of the patients who urgently need the drug. Furthermore, requiring more evidence increases the costs of research. So requiring more evidence also harms the financial interests of those who have to pay for gathering the evidence (note that this could, indirectly, be the patient, since higher research costs may lead to a higher price of the drug).

It is on all of these non-epistemic interests (not just the users' interest to stay alive) that the decision as to how much evidence is exactly needed to accept the hypothesis, depends. If the degree of evidential support for the hypothesis that the toxic ingredient is not present in lethal quantity could be considerably increased in a very quick and very cheap way, then we may consider this necessary before it can be accepted, given the harmful consequences of wrongly accepting this hypothesis. If, on the other hand, it would be very time-consuming and expensive to marginally increase an already high degree of evidential support, then the current degree of evidential support may be considered sufficient to accept the hypothesis, and we may not consider it beneficial that more time and resources are spent on gathering additional evidence.

In general, when scientists decide whether or not to accept a certain hypothesis, they have to ask whether the amount of evidence they have is sufficient, and for this, they have to consider the different non-epistemic interests that would be served/harmed by rightly/wrongly accepting the hypothesis and by postponing this decision in order to gather additional evidence. In that way, non-epistemic interests influence whether or not scientists accept a certain hypothesis.

A possible response is that scientists can avoid this interest influence by restricting their task to determining the probability of a hypothesis, or the degree to which it is supported by scientific evidence. The decision whether or not to *accept* the hypothesis could be left to others (non-scientists). But, as Rudner (1953, p. 4) rightly points out, this would not avoid that scientists have to make a decision on the acceptance of a hypothesis. The hypothesis would then be that the likelihood of hypothesis *h* is p, or that *h* is supported by a particular kind of evidence with a particular strength. Scientists would have to decide whether or not to accept *such* hypotheses (i.e. hypotheses on probability or strength of evidence), and for that, they would have to consider the relevant non-epistemic interests just as well.

Decisions on the acceptance of hypotheses are inevitable in science. These decisions have a potential impact on how well our non-epistemic interests are served. Scientists have to take this impact into consideration. When deciding whether or not to accept a hypothesis, they have to consider the relevant non-epistemic interests, and these considerations should affect their judgment.

What are the implications of this argument for the ideal of purity? I have argued that the decision on whether or not to accept a hypothesis in science should not be detached from non-epistemic interests. Hence, we should reject the full ideal of purity, which states that such scientific decisions *should* be detached from non-epistemic interests.

2.3 The partial ideal of purity

One way to avoid this objection is to opt for a less strict version of the ideal of purity, which allows scientists to consider non-epistemic interests in judging whether or not the evidence in favor of a hypothesis is sufficient to accept it. Such a view has been defended by Heather Douglas (2009). Her view is based on a distinction between epistemic criteria and values. Douglas mentions two epistemic criteria: internal consistency and predictive competence. She distinguishes these criteria from values, including ethical values (which focus on the good or the right), social values (which arise from what is valued by a particular society), and cognitive values (which help scientists with their cognition in science) (Douglas 2009, pp. 92-94).

Furthermore, Douglas makes a distinction between two roles that values can play: a direct and an indirect role. Values play a direct role when they act as stand-alone reasons to motivate a decision (just as evidence is a stand-alone reason for accepting a claim). They play an indirect role when they are used to determine whether the available evidence is sufficient to make a certain decision. More specifically, in the indirect role, they are used to weigh the potential consequences of a wrong decision, and when these consequences are considered dire, this leads to more evidence being required (Douglas 2009, pp. 96-97). Douglas states that in certain kinds of decisions in science, namely, decisions on how to characterize data, on how to interpret evidence, and on the acceptance or rejection of scientific theories, values are not allowed to play a direct role, but only an indirect role (Douglas 2009, pp. 102-103).

This corresponds with the following view on interest influences on scientific decisions. The only role that non-epistemic interests are allowed to play in scientific decision-making on the characterization of data, the interpretation of evidence, and theory acceptance, is the role of being considered in judgments on whether or not the available evidence is sufficient to make the decision. Such an indirect influence of non-epistemic interests on scientific decisions is regarded as acceptable. But when non-epistemic interests have a direct influence on decisions concerning the characterization of data, the interpretation of evidence, or theory acceptance, that is, when scientists use these non-epistemic interests as stand-alone reasons to motivate the decision, then their influence is *not* acceptable. I call this view the partial ideal of purity. It consists of the following theses:

(1) An interest *indirectly* influences a decision if it is used to determine whether the available evidence is sufficient to make that decision.
(2) Indirect influences of non-epistemic interests on scientific decisions are harmless (and even desirable).
(3) An interest *directly* influences a decision if it is used as a stand-alone reason to motivate the decision.
(4) Direct influences of non-epistemic interests on decisions concerning the characterization of data, the interpretation of evidence, or theory acceptance are unacceptable in science.
(5) Such influences can and must be eliminated.

2.4 Earlier work on values in science

In the next chapters, I will attack the partial ideal of purity. But first, I want to link this ideal to the existing philosophical literature on values in science. A lot of philosophers of science have worked on the topic of values in science, and their work contains some insights that challenge the ideal of purity. What I wish to argue in the current section is that while these insights may successfully undermine the *full* ideal of purity, they do not undermine the *partial* ideal of purity.

2.4.1 Underdetermination of scientific theory

A lot of writings on values in science build on the idea of underdetermination of scientific theory. So let me start from there. Underdetermination arguments, which are often linked to the original formulation of underdetermination by Pierre Duhem (1914/1954) and W.V.O. Quine (1951), support the view that empirical evidence alone is not sufficient to choose between scientific theories. According to Stanford (2009), there are two fundamental varieties of such arguments. The first is that when we falsify an empirical prediction derived from a certain hypothesis, then we do not know whether the fault lies with this particular hypothesis, or with one of the many background assumptions (e.g., about the operation of instruments) that we needed to generate the failed prediction. A hypothesis can never be tested in isolation; we need additional assumptions and auxiliary hypotheses to generate predictions. If a prediction then turns out to be falsified, then this indicates that there is something wrong with the collection of hypotheses and assumptions that we used to derive the prediction. But which part of this collection – which particular hypothesis or background assumption – is problematic remains undetermined. So it is not because we derived a failed prediction from a hypothesis, that we need to abandon the hypothesis; we can also abandon any of the other elements used. Stanford (2009) calls this idea 'holist underdetermination'.

The second kind of underdetermination he distinguishes is contrastive underdetermination. Contrastive underdetermination concerns the thought that whenever a theory is confirmed by a body of evidence, there might be other theories which are confirmed by that same body of evidence (Stanford 2009). A now-classic illustration of this idea has been offered by Bas van Fraassen. Let me recapitulate it. Newton's theory of mechanics and gravitation assumes that there is an absolute space to which motion is relative. Newton added the hypothesis that the center of gravity of the solar system is at rest in absolute space. He noted, however, that if this center of gravity would move at a constant velocity in absolute space, the appearances would be no different. This means that we can generate an endless amount of theories which are all consistent with the same appearances: take Newton's theory and add the postulate that the center of gravity of the solar system has constant absolute velocity 0 m/s; take Newton's theory and add the postulate that the center of gravity of the solar system has constant absolute velocity 1 m/s; repeat this step for an endless amount of other velocities. The resulting theories are all consistent with the same body of empirical evidence (van Fraassen 1980, p. 46).

Both kinds of underdetermination arguments challenge the following view on theory choice: first, we should develop as many theories about a set of phenomena as we can come up with, and then, we should test them until all theories but one have failed certain empirical tests; the only theory that hasn't failed on any tests, is the one we should choose. Holist underdetermination shows that it is not because a theory has failed on an empirical test, that we should dismiss the theory; maybe there is nothing wrong with the theory, but with the background assumptions we made in linking the empirical test to the theory. Contrastive underdetermination shows that it is naïve to think that we will ever end up with only one theory that hasn't failed on any tests; whenever there is one theory that has passed all tests, we can always come up with other theories that would have passed exactly the same tests.

2.4.2 Thomas Kuhn on theory choice

How, then, should we choose between theories? It has been argued that in science, different values (for theories) are used to make such choices – not only the value of empirical adequacy or (predictive) accuracy, but other values, such as simplicity or fruitfulness, as well. Thomas Kuhn (1977) mentions five characteristics of a good scientific theory: a theory should (1) be accurate (i.e. in demonstrated agreement with the results of existing experiments and observations), (2) be consistent with itself and with other currently accepted theories applicable to related aspects of nature, (3) have broad scope (that is, the theory's consequences should extend beyond the particular observations, laws, or subtheories it was initially designed to explain), (4) be simple (that is, bring order to an otherwise confused set of phenomena), and (5) be fruitful (that is, the theory should disclose new phenomena or relationships). According to Kuhn, these five criteria for evaluating the adequacy of a theory form, together with other criteria of the same sort, the shared basis for theory choice (Kuhn 1977, pp. 321-322).

Kuhn stresses that we should not think of the criteria as rules that determine theory choice; they do not provide a shared algorithm for choosing between theories. Rather, we should think of them as values, which *influence* theory choice. Kuhn presents his desiderata as part of a shared list of properties that scientists have to consider in choosing between theories. The outcome of considering these properties may, however, differ from scientist to scientist (Kuhn 1977, p. 331). The reasons Kuhn gives for this are that the interpretation of a certain desideratum may differ from scientist to scientist, and that different scientists may attach different relative weights to the various desiderata. Both kinds of differences could result in a different theory choice (Kuhn 1977, pp. 324-325).

Thus, theory choice does not only depend on which desiderata are used, but also on how these desiderata are interpreted and weighed. The latter is, according to Kuhn, influenced by the characteristics of the individuals making the choice (Kuhn 1977, p. 324). Note that this view provides a first hint at some potential influences of interests on theory choice: the personal interests of a scientist could affect how he interprets the desiderata for theory choice and which relative weights he attaches to these desiderata, and this could, in turn, affect which theory he ends up choosing.

2.4.3 The epistemic/non-epistemic distinction

The basics of Kuhn's view are later endorsed by Ernan McMullin (1983). McMullin reworks Kuhn's list of values a bit and ends up with predictive accuracy, internal coherence, external consistency, unifying power, fertility, and simplicity as desiderata for theories. He agrees with Kuhn that such desiderata operate as values, and that different scientists, committed to the same set of values, could end up choosing different theories because they can disagree on how well a certain theory scores on a certain value, and because they may attach different relative weights to the different values (McMullin 1983, pp. 15-17). Where McMullin goes further than Kuhn is in calling the values that are appropriate for theory assessment 'epistemic' and linking them to truth. He puts it as follows:

> Such characteristic values I will call *epistemic*, because they are presumed to promote the truth-like character of science, its character as the most secure knowledge available to us of the world we seek to understand. An epistemic value is one we have reason to believe will, if pursued, help toward the attainment of such knowledge. I have concentrated here on the values that one expects a good *theory* to embody. But there are, of course, many other epistemic values, like that of reproducibility in an experiment or accuracy in a measurement. (McMullin 1983, p. 18)

McMullin makes a distinction between these values and other, non-epistemic values on the basis of a correspondence theory of truth:

> The decision as to whether a value is epistemic or non-epistemic in a particular context can sometimes be a difficult one. But the grounds on which it should be made are easy to specify in the abstract. When no sufficient case can be made for saying that the imposition of a particular value on the process of theory choice is likely to improve the *epistemic* status of the theory, that is, the conformity between theory and world, this value is held to be non-epistemic in the context in question. (McMullin 1983, p. 19)

Larry Laudan (1984) calls values such as empirical accuracy, coherence, and simplicity 'cognitive values', and distinguishes them from moral values. Contrary to McMullin, he does not base this distinction on the idea that the values of the former category are conducive to truth or to the conformity between theory and world. The distinction is instead based on the (rough) idea that the cognitive values represent properties of theories "which we deem to be constitutive of 'good science'" (Laudan 1984, p. xii). The cognitive values figure in Laudan's reticulated model, which states that theories, methods, and aims (including cognitive values) are involved in mutually supportive relationships in science. Elements of all three categories can change (Laudan 1984, Chap. 4). From this, we can infer that truth is not *the* ultimate, fixed aim of science, and hence, that cognitive values need not necessarily be truth-conducive in Laudan's view.[2]

Helen Longino (1983, 1990) makes a distinction between constitutive and contextual values which is similar to Laudan's distinction. Constitutive values are, in her view, "generated from an understanding of the goals of science", and they are "the source of the rules determining what constitutes acceptable scientific practice or scientific method" (Longino 1990, p. 4). For instance, if we take explanation of the natural world to be the goal of science, then the constitutive values are determined on the basis of an understanding of what counts as a good explanation (e.g., satisfying criteria such as truth, accuracy, simplicity, predictability, and breadth). Contextual values belong, on the other hand, "to the social and cultural environment in which science is done" (Longino 1990, p. 4). Contextual values are "group or individual preferences about what ought to be" (Longino 1990, p. 4), and they include personal, social, and cultural values. Longino uses this distinction to argue that contextual values are incorporated in scientific reasoning. Her argument is that background assumptions are needed in evidential reasoning (i.e. reasoning between data and hypotheses), and these background assumptions sometimes reflect certain contextual values (Longino 1990, pp. 215-216).

2.4.4 Criticism

Regardless of whether values such as empirical adequacy, consistency, and simplicity are called 'epistemic', 'cognitive', or 'constitutive', the very idea that they could/should be distinguished from other values, such as moral or political values, has been criticized. For instance, Phyllis Rooney (1992) questions the distinction by arguing that how a scientist perceives the simplicity, fruitfulness, or coherence of a theory – that is, how well the theory scores on the so-called 'epistemic' values – is determined by all sorts of 'non-epistemic' factors. An example she offers is the perception of the simplicity of the linear-hormonal model. This model posits that prenatal hormone levels causally influence brain organization, that the latter causally influences gender-linked behavior, and that while the environment may influence gender-linked behavior as well, this influence is independent of hormone levels and brain organization (Longino 1990; Magnus 2005). This explains biological determinism (i.e. the assumption that behavior is determined by biological factors), which is, according to Rooney, one of the reasons why

[2] Also see Laudan (2004), in which Laudan explicitly argues that some of the cognitive virtues, namely attributes of scope and generality, have no epistemic foundation, since manifesting these virtues is neither necessary nor sufficient for the truth of a statement.

the linear-hormonal model was considered simple. The view of biological determinism was, however, motivated by gender dimorphism (i.e. the assumption that men and women are different), and the latter was based on the social and cultural values of a society based on gender order and hierarchy. Hence, it was partly because of the 'non-epistemic' values of those developing the linear-hormonal model, that this model was perceived as manifesting the 'epistemic' value of simplicity (Rooney 1992, p. 18).

Another criticism on the epistemic/non-epistemic distinction has been developed by Longino (1996). She starts by describing two lists of values for theories. On the one hand, she recapitulates Kuhn's original list of cognitive values, consisting of accuracy, (internal and external) consistency, simplicity, breadth of scope, and fruitfulness (see above). Next, she shows that Kuhn's list is not the only possible one. She presents an alternative list of values for theories, which contains, besides empirical adequacy (Kuhn's accuracy), also the values of novelty, ontological heterogeneity, mutuality of interaction, applicability to current human needs, and diffusion of power (Longino 1996, pp. 41-50).

Longino links the latter values to the feminist cognitive aim of revealing gender and/or the activities of those gendered female, which is also a sociopolitical aim. More specifically, she claims that the values of the second list can be justified by the fact that theories manifesting them are more likely to increase the visibility of women, female-identified phenomena, and gender relations. The first list does not seem to have that same progressive, feminist potential. In fact, Longino argues that at least some of the values of the first list (external consistency, simplicity, and fruitfulness) often serve anti-progressive, gender-oppressive ends. Accordingly, which values one uses to evaluate theories reflects certain political values, which suggests that the former values are not purely cognitive. This undermines the dichotomy between cognitive and non-cognitive values (Longino 1996, pp. 50-55).

(Note that none of the criticisms on the distinction between epistemic and non-epistemic *values* undermine the distinction I made between epistemic and non-epistemic *interests*. I have characterized epistemic interests as interests in knowing something, and non-epistemic interests as all other interests. The purpose of this distinction is to show that interests other than epistemic interests affect scientific decisions. For this purpose, a strict separation between epistemic and non-epistemic interests is not required. Epistemic interests may reflect certain non-epistemic interests, and vice versa. I do not presuppose that such links are absent.)

2.4.5 Values in science and the ideal of purity

How do these discussions on values in science exactly challenge the ideal of purity? These discussions suggest that there are at least three ways in which non-epistemic interests can potentially affect theory choice:

(1) Non-epistemic interests can influence which properties one regards as values for theories (which are then used to evaluate theories and choose between them).
(2) Non-epistemic interests can influence how the values, used to choose between theories, are interpreted.
(3) Non-epistemic interests can influence the relative importance or weights one attaches to the different values in choosing between theories.

It seems hard, if not impossible, to exclude such potential interest influences from scientific practice. So far, no one has been able to develop a complete and viable picture of how theories ought to be evaluated which ensures that theory choice is unaffected by non-epistemic interests. Therefore, requiring that scientific decisions such as theory choice are unaffected by non-epistemic interests, is too demanding. Since the full ideal of purity *does* require that, we can conclude that it is too demanding. Hence, earlier work on values in science undermines the full ideal of purity.

It does, however, not undermine the partial ideal of purity. In none of the three potential interest influences described above, non-epistemic interests are used as stand-alone reasons to choose a certain theory. So these interest influences do not involve a *direct* influence of non-epistemic interests on theory choice (see section 2.3). Instead, non-epistemic interests *indirectly* influence theory choice, by influencing which values are used in theory evaluation, how these values are interpreted, or how they are weighed. Since the partial ideal of purity only condemns certain *direct* influences of non-epistemic interests, it is not challenged by the apparent inevitability of the three potential *indirect* interest influences on theory choice described above.

2.4.6 Janet Kourany on values in science

Finally, I should mention Janet Kourany's 'Philosophy of science after feminism'. In that book, Kourany attacks the ideal of value-free science (which corresponds with what I have here called the full ideal of purity), and defends an alternative view: the ideal of socially responsible science. According to the ideal of socially responsible science, scientific research should be controlled through and through by sound epistemic values as well as by sound social values (Kourany 2010, pp. 68-71). It seems then that Kourany allows non-epistemic interests to play a more prominent role in science than adherents of the ideal of purity would allow. But I think that Kourany's arguments only undermine the *full* ideal of purity, and not the *partial* ideal of purity. To support this point, let me consider one of Kourany's key arguments in favor of her ideal of socially responsible science.

In order to argue that scientific research should be controlled through and through by sound social values, Kourany uses the example of Carolyn West's psychological research program concerned with the problem of domestic violence in the United States. The aim of this program is "to uncover the similarities in intimate-partner violence within the black and white communities of the United States without negating the experiences of black women and simultaneously to highlight the differences within the black and white communities without perpetuating the stereotype that black Americans are inherently more violent than other ethnic groups" (Kourany 2010, p. 69). According to Kourany, this complex aim requires research on domestic violence to change in multiple ways. More specifically, it requires (1) that the definition of partner violence is broadened so that it includes psychological, emotional, verbal, and sexual abuse as well as physical abuse, (2) that the ways in which violence is measured are changed – from merely counting violent acts and measuring their severity to taking into account the contexts, motives, and outcomes of the violent acts, using a combination of qualitative and quantitative research methods, including listening to the voices of battered women, (3) that measurement scales are revised so that they reflect more than the experiences of white European Americans, (4) that within-group differences in the black and white communities are investigated in order to determine whether apparently racial differences are not simply socioeconomic differences instead, and (5) that participants are integrated into every stage of the research process, from planning to implementing, interpreting, and disseminating results (Kourany 2010, pp. 69-70).

Kourany identifies two social values that play in West's research program: 'women deserve to live without fear of violence from domestic partners' and 'black women deserve the same opportunities as white women to live in such partnerships'. She argues that these social values imply that the similarities in domestic violence between the black and white communities should be highlighted in the relevant research, and that any dissimilarities between these communities should be explained in terms of social differences such as racism and poverty, insofar as this is empirically possible (Kourany 2010, p. 70). So according to Kourany, social values shape the research program. She also specifies which aspects of the research program are exactly affected:

> It affects not only research questions but also, as we have seen, such aspects of research as concepts (e.g., the concept of 'partner violence' itself); measurement scales and techniques; methods of subject selection; strategies of data collection, analysis, and interpretation; and even methods of publishing and disseminating results. It may even affect other central aspects of the research process, such as

> consideration of the consequences of error and setting acceptable levels of risk (see, e.g., Douglas 2000). (Kourany 2010, pp. 72-73)

This passage can be used to argue against the full ideal of purity. More specifically, the recognition of the role of social values in considering the consequences of error and in setting acceptable levels of risk corresponds with Rudner's argument, and this argument shows that the full ideal of purity must be rejected, as we have seen in section 2.2. The above passage does, however, not show that the partial ideal of purity must be rejected as well. It suggests that non-epistemic interests affect decisions concerning the characterization of data, the interpretation of evidence, and theory acceptance by influencing which concepts and methods are used to characterize data and to interpret evidence, and by influencing how much evidence is needed to accept a theory. So non-epistemic interests affect the relevant decisions via certain conceptual and methodological choices, and via judgments on the sufficiency of evidence. Such indirect interest influences are acceptable according to the partial ideal of purity. Accordingly, the view that the interest influences identified by Kourany are acceptable does not undermine the partial ideal of purity.

Chapter 3
Mechanistic artifact explanation[1]

Although the partial ideal of purity avoids certain problems of the full ideal of purity (see chapter 2), I think we should reject it as well. A criticism on the partial ideal of purity will be developed in this and the next chapters. In order to develop such a criticism, I will discuss certain interest influences that have not been considered so far in this dissertation. More specifically, I will discuss certain influences of epistemic and non-epistemic interests in the construction and evaluation of explanations in science. Since these influences are maybe most evident in the technological sciences, I will draw on examples from these sciences to clarify the roles of interests in explanatory tasks (in chapters 3 and 4). Chapter 5 will show that similar interest influences can also be found in other sciences. Furthermore, chapter 5 will reveal how my findings exactly challenge the partial ideal of purity.

3.1 Seeking a mechanistic artifact explanation

To elucidate how interests could affect explanation, let me start with a kind of explanation constructed and used in many technological sciences: mechanistic artifact explanation, i.e. an explanation that accounts for an artifact behavior by describing the

[1] This chapter is based on De Winter (2011), which is an article on a pragmatic account of mechanistic artifact explanation. I have included parts of that article in this dissertation because I think that it has some important (unexpected) implications with respect to the roles of non-epistemic interests in science (which are not reported in De Winter (2011)). These implications are discussed at the end of this chapter (sections 3.4 and 3.5).

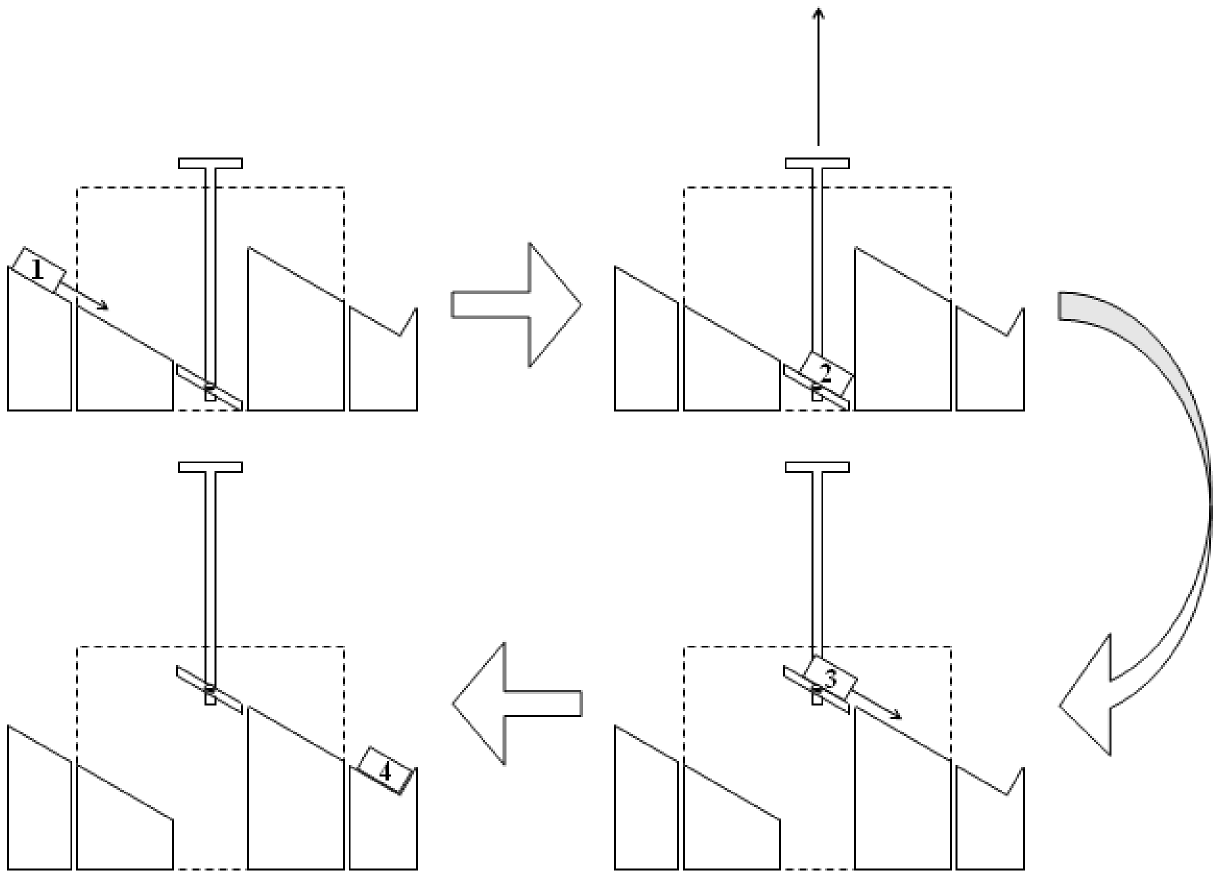

Figure 3.1 Simple mechanism.

underlying mechanism. Following Glennan (2002, p. S344), a mechanism can be defined as follows:

> A mechanism for a behavior is a complex system that produces that behavior by the interaction of a number of parts, where the interactions between parts can be characterized by direct, invariant, change-relating generalizations.[2]

In order to illustrate and clarify my claims, I will use an example of a simple mechanism. The mechanism is depicted in figure 3.1. When one puts a wooden box on the left block of the artifact (**1**), then it slides down the slope until it is on the foot of the elevator (**2**). When one pulls the handle, the box moves up (from **2** to **3**), after which it slides down the second slope, to finally drop in the cup on the right hand side of the box-moving device (**4**). The explainee (= the person who seeks an explanation) does not

[2] I use Glennan's (2002) definition of mechanisms because I think it captures a number of important aspects of mechanisms (which are spelled out in de Ridder (2006)), and because I see no reason to reject this definition. (De Ridder (2006) does formulate some criticisms on Glennan (2002), but these focus on Glennan's characterization of a *mechanical model*, and not on his notion of a *mechanism*.)

know what happens within the dotted line. He knows that when one puts a box on the left block, and pulls the handle after a while, the box will turn up on the right hand side of the artifact, but he does not know the inner workings of the box-moving device. However, he is interested in these inner workings, and therefore, he asks for an explanation that explains the behavior of the artifact by describing the underlying mechanism of this behavior. In other words, he asks for a mechanistic artifact explanation of the transportation of the box. I believe that different mechanistic artifact explanations can be constructed for this artifact behavior.

3.2 Top-down and bottom-up

De Ridder (2006) distinguishes two strategies to create mechanistic artifact explanations: the top-down strategy and the bottom-up strategy. He offers the following definitions:

> *Top-down strategy:* take the behavior to be explained and decompose it into more basic sub-behaviors, reiterate this step if possible, it should become clear how the complex behavior being explained is realized by simpler behaviors in a specific spatiotemporal configuration, and for all the sub-behaviors, indicate which component(s) take(s) care of them.
> *Bottom-up strategy:* name the structural components of the artifact and give information about their physicochemical make-up and spatial configuration, show how their physicochemical features and configuration result in various behaviors and then describe how these behaviors, in their spatiotemporal configuration, together make up the behavior to be explained. (de Ridder 2006, p. 87)[3]

[3] Similar explanatory strategies are presented in Bechtel & Richardson (1993, p. 18). The top-down strategy is Bechtel & Richardson's synthetic strategy applied to an artifact. The bottom-up strategy, on the other hand, is not entirely the same as Bechtel & Richardson's analytic strategy applied to an artifact. The difference is that the bottom-up strategy includes giving information about the physicochemical make-up of the components, while the analytic strategy does not. Cummins (1975) also presents two explanatory strategies, the analytical strategy and the subsumption strategy, that are similar to, respectively, the top-down and the bottom-up strategy. For the differences between Cummins's explanatory strategies on the one hand, and the top-down and bottom-up strategy on the other hand, see de Ridder (2006, p. 87n).

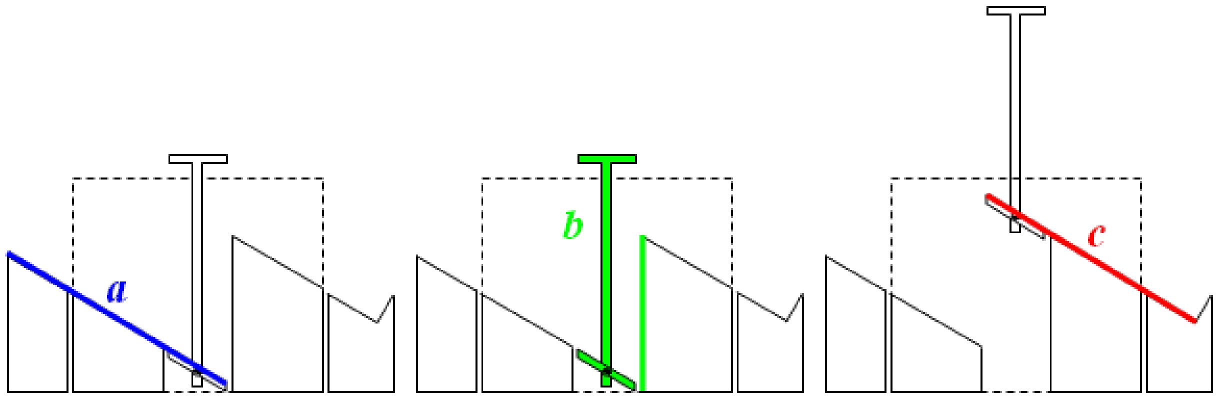

Figure 3.2 Artifact components to which the top-down explanations refer.

Before I illustrate these strategies, it should be noted that whether or not the first step of the top-down strategy (decompose the behavior into more basic sub-behaviors) can, and thus should, be reiterated, is relative to context in de Ridder's view. We can infer this from the following quote.

> It would be absurd if the [...] guideline ['break the function down as fine as possible'] were interpreted as requiring that every functional decomposition end in fundamental physics. The relevant sense of 'as fine as possible' must be more pragmatic. It seems plausible that certain sub-behaviors will count as basic or atomic for particular engineering disciplines. Which will vary across different fields; where mechanical engineering will accept, say, the strength of a particular alloy as given, materials engineering will be interested in how this strength is realized by the behaviors of elements in the alloy. (de Ridder 2007, p. 83)

This implies that which sub-behaviors are not further decomposed in a top-down explanation can differ from context to context, and thus, that a top-down explanation can include less or more information.

Let me construct two different top-down explanations of the transportation of the box in the box-mover. The first makes abstraction of the forces at work, the second does not. The first top-down explanation decomposes the transportation of the box into three sub-behaviors: the box's movement from **1** to **2** (see figure 3.1) during time interval [t_1, t_2[, the upward movement of the elevator with the transportation of the box to **3** during [t_2, t_3[, and the box's movement from **3** to **4** during [t_3, t_4[.The first sub-behavior is taken care of by slope *a* (see figure 3.2); the second sub-behavior is taken care of by elevator system *b*; and the third sub-behavior is taken care of by slope *c*. This is the first top-down explanation.

The second top-down explanation includes more information. Just as the first one, it decomposes the transportation of the box into three sub-behaviors: the box's movement from **1** to **2** during time interval [t_1, t_2[, the upward movement of the elevator

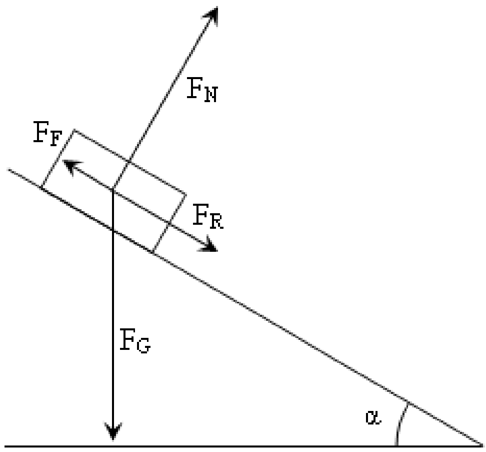

Figure 3.3 Forces exerted on a box on a slope.

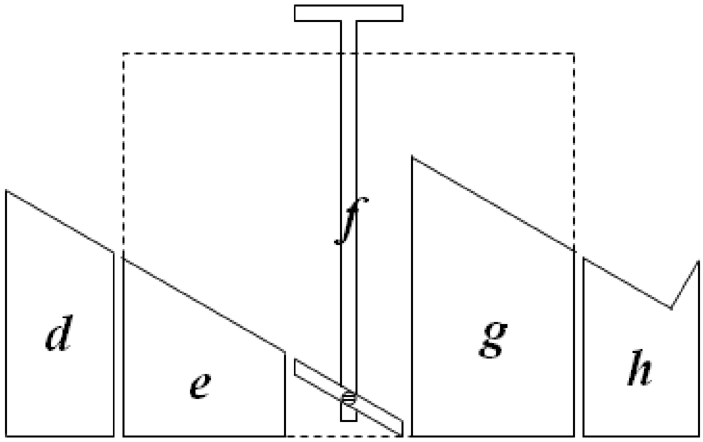

Figure 3.4 Artifact components to which the bottom-up explanation refers.

with the transportation of the box to **3** during [t_2, t_3[, and the box's movement from **3** to **4** during [t_3, t_4[. But it also decomposes each of these behaviors. Slope *a* causes the box to move from **1** to **2** by reacting to gravitational force F_G (see figure 3.3) with normal force F_N, and by reacting to the resultant F_R (= $F_G \cdot \sin \alpha$) of these two forces with a frictional force F_F that is smaller than F_R, so that the total force F, which is parallel to the surface of the ramp and equal to $F_R - F_F$, is larger than 0. Because $F > 0$, the box slides down slope *a*, from **1** to **2**. The box's movement from **2** to **3**, and from **3** to **4**, can be decomposed in similar ways. To reduce the complexity of this chapter, I do, however, not explicate such decompositions here, but they are assumed to be included in the second top-down explanation.

Some may claim that only an explanation such as the second, more extensive top-down explanation is a full explanation of the transportation of the box from **1** to **4**, and that an explanation that does not reveal the forces at work is not a real explanation. I return to this point in section 3.3, showing that the first, more concise top-down explanation can be a complete and fully satisfying explanation as well.

Now consider a bottom-up explanation of the transportation of the box. Such an explanation might look as follows. The box is a 0,5 kg oak box. Components *d*, *e* and *g* (see figure 3.4) are oak blocks that each have a lubricated 30-degree slope. Components *f* and *h* are also made of oak. The foot of *f* has a 30-degree slope that is lubricated, and is strongly screwed down the rod, which is held upright by the surrounding box. The top of *h* is a lubricated 30-degree slope with a barricade at its end. (Of course, several other features should be added, such as the length, width, and height of each component, but I will stop here because summing up all these features is not necessary to clarify how a bottom-up explanation might look.) In which behaviors do the physicochemical and spatial features of the components of the box-mover result?

This is revealed in the second part of the bottom-up explanation. Because the ramps in the box-mover are constituted by oak components, they do not collapse when a 0,5 kg box is on top of them (if the constituting components were made of, say, whipped cream, then the ramps would collapse). Instead, they react to the weight F_G of the box with normal force F_N (see figure 3.3). F_R is the resultant of F_G and F_N, and is equal to $F_G \cdot \sin \alpha$. Since F_G = 4,9 N and α = 30°, F_R is equal to 2,45 N. The box slides down the slope if F_R exceeds the friction between the slope and the box (F_F). In the box-mover, it is guaranteed that F_F is smaller than 2,45 N because the slopes and the box are lubricated (lubrication reduces friction). As a consequence, the box moves down the slopes in the box-mover, from **1** to **2**, and from **3** to **4**. Because *f* is made of oak and because the foot is strongly screwed down the rod, it is solid enough to carry the weight of the box if the handle is pulled. As *g* is made of oak, it is solid enough to prevent the box from sliding down the foot of *f* as long as it is next to *g*. Because *f* can carry the weight of the box, and because *g* prevents the box from sliding down *f*'s foot, we know that when the box is on the foot of *f* and the handle is pulled, the box moves from **2** to **3**.

So far, I have explicated physicochemical and spatial features of the different components of the box-moving device, and I have shown how these features result in various behaviors. One thing remains to be done for our bottom-up explanation to be complete: it should be clarified how these various behaviors make up the behavior to be explained, that is, the transportation of the box. This can easily be done by specifying the temporal organization of the different behaviors. The box's movement from **1** to **2** occurs during time interval $[t_1, t_2[$, its movement from **2** to **3** occurs during $[t_2, t_3[$, and its movement from **3** to **4** occurs during $[t_3, t_4[$. This organization constitutes the transportation of the box from **1** to **4** during $[t_1, t_4[$.

What are the differences between top-down and bottom-up explanation of the transportation of the box? A salient difference is the fact that the bottom-up explanation does not only refer to sub-behaviors of the behavior to be explained and the components taking care of these behaviors, as the two top-down explanations do; it also mentions physicochemical and spatial properties of the artifact components, while the top-down explanations do not.

A second difference is that the artifact components to which the top-down explanation refers, differ from the artifact components to which the bottom-up explanation refers (compare figure 3.2 to figure 3.4; the components are marked out differently). This is due to a difference in perspective. When one uses the top-down strategy, one takes a behavioral or functional perspective: one marks out the artifact components on the basis of the sub-behaviors of the behavior to be explained. If several pieces contribute to the same behavior, then there is, from a functional perspective, a tendency to take them together as one more-piece component. When one uses the bottom-up strategy, one takes a structural perspective, which means that one identifies the artifact components before one knows what they do. From this perspective, one will typically pick out one-piece components; these components cannot be taken together on the basis of the behavior to which they contribute because one does not know the relevant behaviors yet. Let me call artifact components that are identified from a functional perspective, functional components or F-components, and artifact components that are identified from a structural perspective, structural components or S-components. Since top-down explanations correspond with a functional perspective, they refer to F-components, and since bottom-up explanations correspond with a structural perspective, they refer to S-components.[4]

3.3 Ideal explanation in different contexts

So there are at least three mechanistic artifact explanations of the transportation of the box in the box-mover. Does the fact that the bottom-up explanation reveals more about the artifact than the top-down explanations imply that the bottom-up explanation is always the best explanation? I do not think so. I think that for each explanation, there are contexts in which it is ideal. Mechanistic artifact explanations are useful in at least two kinds of contexts: the context of use and the context of design. First, let us consider the context of use.

3.3.1 Context of use

If we know the underlying mechanism of the box transportation, then we can draw several conclusions about how the box-mover will react to certain manipulations. For instance, we know that if we hold the box-mover upside down, it will not display the behavior it is supposed to display; and if we pull the handle before the box is on the foot of the elevator, then the box will end up under the elevator. We can draw these conclusions from all the mechanistic explanations of the box transportation discussed. But some conclusions can only be drawn from the bottom-up explanation. For instance, we know that we can drive (metal) screws in the parts of the box-mover (e.g., for attaching the box-mover to a wooden bottom plate) because we know that these parts are made of oak, and not of metal or some other hard material. The fact that the parts are made of oak is only revealed by the bottom-up explanation.

[4] I do not assume that F- and S-components are necessarily different, because often, they are not. My point is that in a top-down explanation, we can be sure that the components referred to are F-components, while it is not guaranteed that they are also S-components. In a bottom-up explanation, the components referred to are necessarily S-components, but they are not necessarily F-components.

We can draw more conclusions about an artifact from a bottom-up explanation of the behavior of this artifact, but this comes with a cost. Suppose we want to draw conclusions that are not only relevant to the artifact under consideration, but also to artifacts with other physicochemical/spatial properties. We should then make abstraction of the physicochemical and spatial properties, and only focus on the factors that the different artifacts have in common. Top-down explanations do that, contrary to bottom-up explanations, and therefore, the conclusions we draw from them are more likely to be relevant for artifacts with other physicochemical/spatial properties as well. So while bottom-up explanations are more adequate if our goal is to draw as many conclusions as possible about a certain specific artifact, top-down explanations are more adequate for drawing conclusions that are relevant for as many artifacts as possible.

But how much information should exactly be included in a top-down explanation? This depends on what the explanation will be used for. Consider the first top-down explanation of the box transportation. Once we know this explanation, we can predict what will happen if we turn the box-mover upside down, or if we pull the handle before the box is on the foot of the elevator; we do not need the additional information that the second top-down explanation provides to make such predictions. This is because we are familiar with falling objects and with objects sliding down slopes.

But sometimes, we want to make predictions concerning situations with which we are not familiar. Suppose that, before the first man traveled in space, one wanted to know how box-movers would behave in an orbital space station. The explainee could then not be familiar with the situation about which he wanted to make a prediction. As a result, the first top-down explanation could not help him out: knowing the sub-behaviors of the box transportation on earth, and the components taking care of these behaviors, would not be sufficient to predict what would happen in space. The explainee would need more information, and such information is provided by the second top-down explanation. The second top-down explanation reveals that slope *a* causes the box to move from **1** to **2** by reacting to F_G with normal force F_N. Since slope *a* does not react to F_G with normal force F_N in an orbital space station, our explainee could conclude, on the basis of the second top-down explanation, that the box would not move from **1** to **2** in the space station.

It is clear that in the latter context, the second top-down explanation is more adequate than the first, since it enables the explainee to satisfy his epistemic interest (predicting how box-movers behave in an orbital space station), contrary to the first top-down explanation. Therefore, the second top-down explanation is the better one in that context. But in the former context (in which the explainee only wants to make predictions concerning situations with which he is familiar), the first top-down explanation is the better one I think. The reason is that, in this context, the additional information that the second top-down explanation provides is redundant: the explainee's interests can be satisfied without it. Hence, the time and effort needed to gather and use this information would be a waste of time and energy for the explainee. If we assume that the explainee wants to use his time and energy efficiently (= non-epistemic interest),[5] then this means that the first top-down explanation is more adequate – it serves the explainee's interests better.

3.3.2 Context of design

Mechanistic artifact explanations are also useful in the context of design. Suppose we want to design an artifact that performs the same behavior as some pre-existing artifact. A mechanistic artifact explanation of the behavior of the pre-existing artifact can then be very helpful. Suppose we want to build a fireproof box-mover that is as flat as possible (= non-epistemic interest). Due to the new requirements, we do not consider the physicochemical and spatial features of the original box-mover's components very relevant for our goal (we cannot just use oak because the box-mover should be fireproof, and the slopes should be reconsidered because the box-mover should be as flat as possible), and therefore, the explanation should not explicate these physicochemical and spatial features. So the bottom-up strategy drops out. What is useful in this context, is to know the sub-behaviors of the box transportation and the components taking care of these sub-behaviors. Such information is very useful for our goal because these sub-behaviors and components can be used to realize the box transportation in the new box-mover as well. Although the components of the new box-mover will have different physicochemical and spatial properties, it can, just as the original box-mover, transport the box by making it slide down two slopes and by using an elevator to transport the box from the first slope to the second. As both top-down explanations refer to these sub-behaviors and components, the question is: which top-down explanation is most appropriate?

To know the minimal height of the box-mover (it should be as flat as possible), one should determine the optimal degree of slope (the lower the degree of slope, the flatter one can make the box-mover, but if the degree of slope is too low, the box does not slide down anymore). To determine this optimum, the generalizations to which the second top-down explanation refers ($F_R = F_G \cdot \sin \alpha$, and $F = F_R - F_F$) are very useful. If the material of the box and of the ramp is given, then we can, on the basis of these generalizations, determine a minimal degree of slope that guarantees the sliding down of the box. So it seems that the second top-down explanation of the box transportation

[5] Note that in practice, this interest is often connected to financial interests: usually, an investor (e.g., the government, a private firm) has to pay for the time and energy of scientists, which entails that the efficient use of their time and energy serves the investor's financial interests.

is most useful, and thus most adequate, as it refers, contrary to the first top-down explanation, to generalizations that can help us in reaching our goal. This is often the case, but there are also contexts in which the first top-down explanation is most adequate.

Suppose the explainee has a lot of experience with slopes, and knows, for different combinations of materials (e.g., a copper box on a zinc ramp, an aluminum box on a glass ramp, etc.), the lowest degree of slope that guarantees the sliding down of the box. Due to this background knowledge, the explainee does not need the generalizations to which the second top-down explanation refers to reach his goal, that is, to make a fireproof box-mover that is as flat as possible. The additional information that the second top-down explanation provides would then be redundant, and the first top-down explanation would provide all the information needed, and thus be most adequate in this context (assuming that the explainee has a non-epistemic interest in not losing time and energy on the acquisition of redundant information).

It is also possible that the job of the explainee is not to build the new box-mover by himself, but to delegate sub-tasks to experts. Once he knows that a box can be transported by making it slide down two slopes, and by using an elevator to transport it from the first slope to the second, he can give slope experts the assignment of designing the two slopes, and experts in elevator systems the assignment of designing the elevator system. No additional information is needed, which means that the first top-down explanation is fully satisfying, and thus most adequate.

So top-down explanations are often most adequate for developing an artifact that performs the same behavior as a pre-existing artifact, and how much information should exactly be included in the top-down explanation can differ from context to context. In still other contexts, the most promising explanatory strategy to create an artifact that displays the same behavior as a pre-existing artifact, is the bottom-up strategy. This is the case when a description of the physicochemical make-up and spatial configuration of the components of the pre-existing artifact is a good starting point for the new design. If the current context resembles the context in which the pre-existing artifact was designed (same requirements, same material costs, etc.), then a lot of physicochemical and spatial features of the components of the original artifact can be used for the new artifact as well. In such cases, knowing these features is useful. Because a bottom-up explanation explicates them, contrary to top-down explanations, it is more adequate in such cases.

The first part of a bottom-up explanation (description of the physicochemical make-up and spatial configuration of the S-components of the artifact) is a first suggestion on how an artifact performing the desired behavior might look. We may, however, still be critical of the physicochemical make-up and spatial configuration of the components of the original artifact, and ask questions such as 'Why is the artifact made of the materials it is made of?' The second and third part of a bottom-up explanation (demonstration of how the physicochemical make-up and spatial configuration of the artifact components result in various behaviors and a description of how these behaviors result in the behavior to be explained) help one to answer such questions. For instance, the bottom-up explanation of the box transportation reveals that the artifact components are made of oak *in order to be able to carry the weight of the box without collapsing*. Such information is not only useful to legitimize the physicochemical or spatial property under consideration, but it is also relevant if one decides to change it: e.g., given that the components of the original box-mover are made of oak in order to be able to carry the weight of the box, we know that the new material should be solid enough to carry the weight of the box as well.

3.3.3 Limitations

We have seen that there are various interests that can be served by mechanistic artifact explanations. There are, however, also contexts in which non-mechanistic explanations are most adequate. Suppose that the question 'Why is the box transported from the one end of the box-mover to the other end?' is not motivated by the interest in the underlying mechanism of this behavior, but by the desire to know why the box-mover did not work properly a minute ago, while it does work properly at this moment. The explanatory interest in the box transportation can then be motivated by the desire to know what has changed. An explanation that perfectly serves this epistemic interest is: the box is transported because someone repaired the handle of the box-mover, which was broken a minute ago. This is not a mechanistic artifact explanation as it does not explicate any mechanism, and it outperforms the aforementioned mechanistic artifact explanations because it reveals what has changed in the last minute, while the mechanistic artifact explanations do not.

So it is wrong to think that offering a good explanation of an artifact behavior means offering a mechanistic explanation of this behavior. Although mechanistic artifact explanations can serve various interests, there are also interests that they do not serve, such as the desire to know what has changed. For these interests, mechanistic artifact explanations are inadequate, and therefore bad explanations.

3.4 Construction and evaluation of mechanistic artifact explanations

Section 3.3 shows that a mechanistic artifact explanation that is ideal and fully satisfying in one context, may be insufficient in another, and that a mechanistic artifact explanation that only includes relevant information in one context, may include a lot of redundant information in another. Hence, whether or not an explanation is ideal is relative to context. The reason is that explanations are always sought in specific contexts, in which there are particular interests (e.g., the central epistemic or non-epistemic goal that one wants to reach in this context, but also other interests, such as the non-epistemic interest not to waste time and energy), and it is on the basis of these interests that the explanations are evaluated; the better these interests are served by the explanation, the better the explanation.[6] Let me, on the basis of this finding, develop a general picture of the construction and evaluation of mechanistic artifact explanations in the technological sciences.

It is not the case that technological scientists first construct several different mechanistic artifact explanations for an artifact behavior and then use the interests to be served to compare and evaluate the resulting explanations (as I have done above). Rather, the interests to be served are considered from the beginning, before an explanation is constructed. When the explainer decides which explanatory strategy to pursue, the top-down or the bottom-up strategy, he does this on the basis of the interests to be served, more specifically, on the basis of what kind of explanation he expects to be most adequate for the explainee's (epistemic and non-epistemic) interests.[7] If, for example, the explainee wants to predict how a class of artifacts, with different physicochemical and spatial properties but which behave more or less the same, will behave in certain situations, then the explainer will opt for the top-down strategy. If the explainee wants to make predictions for which he needs information about the physicochemical and spatial properties of a specific artifact, the explainer will use the bottom-up strategy.

During the construction of the explanation, interests keep playing. Suppose, for instance, that in the context of design the explainer starts with the top-down strategy to check whether the mechanism of the original artifact can also be used for the new

[6] This corresponds with Pitt's (2009) account of technological explanation. According to Pitt, which explanation should be offered, depends on the audience for whom the explanation is meant, on what would satisfy them, and on the specific why- and/or how-questions the explanation is supposed to answer.

[7] Note that the explainer, that is, the person who constructs the explanation, and the explainee will often be one and the same person.

design. Once he sees that it can, he may shift to the bottom-up strategy in order to check whether the physicochemical and spatial properties of the original artifact are also relevant for the new design. This move is justified because if they are, knowing these properties could significantly facilitate the design process (hence, this move is justified on the basis of the non-epistemic interest to save time and energy insofar as possible). Suppose, however, that the explainer finds out that they aren't; the artifact to be designed needs totally new physicochemical and spatial properties in order to meet design requirements. Instead of gathering more information about the physicochemical and spatial properties of the original artifact, the explainer will then immediately shift back to the top-down strategy (which does not require that such information is gathered). This shift is justified because no time and energy should be wasted on searching for useless information (i.e. useless for the goals pursued).

During the construction of a top-down explanation, interests are also used to decide where to stop. When the explainer has decomposed a behavior into certain sub-behaviors, he has to decide whether or not to further decompose these sub-behaviors into even more basic sub-behaviors. This depends on whether this would serve or damage the interests of the explainee. If it would serve the explainee's interests, the sub-behaviors are further decomposed; if it would damage his interests (e.g., by wasting his time), they are not.

So interests play an important role in the construction of mechanistic artifact explanations. They are also used in the evaluation of such explanations. While mechanistic artifact explanations are being developed, intermediate results are constantly being evaluated, and this evaluation is based on how well the current explanation serves the interests of the explainee. Is the information currently included sufficient for the realization of the explainee's goals, or should more information be incorporated? Is all the information currently included relevant for the explainee's goals, or should certain parts be removed? If the explanation is sufficient to accomplish the explainee's mission and if it only includes relevant information, it is considered a good explanation, as it is fully adequate.

3.5 Non-epistemic interests and the acceptance of mechanistic artifact explanations

The picture presented in this chapter suggests that non-epistemic interests are allowed to have a direct influence on the decision whether or not one accepts a certain mechanistic artifact explanation. Let me first specify what I mean by a direct influence of an interest on a decision. An interest has a direct influence on a decision if it is used as a stand-alone reason to motivate that decision (see section 2.3). The phrase 'using *x* as a stand-alone reason to motivate a decision' does not mean that *x* is the *only* reason for the decision, or that it is *sufficient* to make the decision; it means that *x* is used as a reason, *in itself*, to motivate the decision. This does not exclude that one also has other, additional reasons for making this decision. Additional reasons may even be necessary for the decision to be considered a justified decision. The point is that the reasons can be thought of as *separate* reasons, and not just as part of an overarching argument (although they may be part of such an overarching argument). Such separate reasons can be weighed against each other in order to arrive at a justified decision. It should also be stressed that a direct influence of a non-epistemic interest on the decision whether or not one accepts a certain explanation does not imply that the non-epistemic interest acts as an ulterior motive. The explainer may be very open about the fact that he uses the non-epistemic interest as a stand-alone reason to motivate his decision. The relevant interest is then definitely not ulterior.

Let me now, in light of my discussion of mechanistic artifact explanation, argue that non-epistemic interests are allowed to have a direct influence on the decision whether or not one accepts a certain explanation. Suppose someone wants to design a new artifact and that he wants to spend as little time and energy as possible on the design. He thinks that a mechanistic artifact explanation of the behavior of an existing artifact can help him in satisfying these non-epistemic interests. Therefore, he (= the explainee) consults someone who can offer him such an explanation (= the explainer). Once the explainer has constructed a mechanistic artifact explanation of the relevant artifact behavior, there are different ways in which he can proceed: he can decide to include more information in the explanation, to remove parts from it, or he can switch to another explanatory strategy (from top-down to bottom-up or vice versa). He can also decide to accept the explanation as it is, and present it in its current form to the explainee. If the explainer considers the current explanation most adequate for the explainee's non-epistemic interests, then he is allowed to use this as a stand-alone reason to accept the explanation as it is. This implies that non-epistemic interests are allowed to have a direct influence on the decision to accept a certain mechanistic artifact explanation.

Some may object that I have mixed things up. They may have a division in mind between two stages. In the first stage, the explainer decides whether or not he accepts a certain explanation, and in the second, he assesses the adequacy of the accepted explanation for the explainee's non-epistemic interests. The idea is that non-epistemic interests can act as stand-alone reasons in the second stage, but not in the first. In the first stage, the explainer should only consider epistemic factors, such as epistemic interests and the values of accuracy and internal consistency, and non-epistemic interests should not have a direct influence on decisions in this stage. Since decisions in the first stage concern the acceptance of explanations, the conclusion of this line of reasoning is that non-epistemic interests should not have a direct influence on the acceptance of explanations. The objection against my point of view is then that by claiming that such a direct influence is allowed, I have mixed up the stage in which one decides whether or not one accepts an explanation, and the stage in which one assesses the adequacy of an accepted explanation for the explainee's non-epistemic interests.

But although such a two-stage approach may sometimes be acceptable, I reject the idea that it is mandatory. In my view, scientists should be allowed to operate in one single context in which both epistemic and non-epistemic interests directly influence decisions on the acceptance of explanations. To support this view, it is useful to point at the fact that none of the mechanistic artifact explanations of the behavior of the box-mover offered in this chapter is entirely accurate. This is because the box is never in position **3** (see figure 3.1). In fact, the box starts sliding down the foot of the elevator as soon as the top right corner of the box is higher than ramp *g* (see figure 3.4); it does not wait until the elevator is completely lifted (as in position **3** in figure 3.1). Since all of the mechanistic artifact explanations above state that at a certain time the box is in position **3**, they are all (partly) inaccurate. The reason why I have not eliminated the inaccuracy is that this would make the explanations more complex. I would have to spend more time and effort on formulating the explanations, and the reader would have to spend more time and effort on reading them. Although such additional time and effort would make the explanations more accurate, it would not make these explanations more fit for the purpose of making certain philosophical ideas clear to the reader. Therefore, I have opted for the inaccurate mechanistic artifact explanations instead of replacing them by more accurate ones.

I think that similar situations can occur in scientific practice. Technological scientists may opt for inaccurate explanations because these explanations are more adequate for the explainee's interests than more accurate explanations. Suppose, for instance, that someone seeks a mechanistic artifact explanation of the behavior of the box-mover (= epistemic interest) because this will help him in designing a new box-mover (= non-epistemic interest). Suppose that one of the explanations offered above is adequate for the latter interest, and that eliminating the aforementioned inaccuracy would not make the explanation even more adequate for this interest. It would only make the explanation more complex, so that more time and effort would be needed for the explainee to understand the explanation. Since the explainee has an interest in saving time and energy (= non-epistemic interest), the explainer decides to accept the relatively inaccurate explanation, and he presents it to the explainee. So in deciding whether or not to accept the explanation, the explainer does not only look at epistemic factors such as accuracy, but he also considers the non-epistemic interests to be served. The fact that the inaccurate explanation serves these interests best is used as a reason, in itself, to accept the explanation. So non-epistemic interests directly influence the acceptance of an explanation in this case.

There is an easy strategy to transform such a direct influence into an indirect influence: one can regard the value of simplicity (in the sense of requiring as little time and effort as possible) as an epistemic value which is adopted because of the explainee's non-epistemic interest to save time and energy. Because simplicity is an epistemic value, it is allowed to have a direct influence on the acceptance of explanations. The non-epistemic interest to save time and energy then *indirectly* influences the acceptance of explanations, via the epistemic value of simplicity. But if we would allow this, then why not allow a direct influence of the non-epistemic interest to save and energy as well? It seems that both the direct and indirect influence come down to the same thing in this case. Hence, if one allows the latter, one should also allow the former.

The proponent of the two-stage approach can avoid this objection by disallowing not only the direct influence, but also the indirect influence via the value of simplicity. Following Douglas (2009), he may hold that only the values of accuracy and internal consistency are true epistemic values, and that the value of simplicity is not (see section 2.3). Accordingly, in the first stage, in which the explainer decides whether or not he accepts an explanation, he is only allowed to use the epistemic values of accuracy and internal consistency, and he should neither use non-epistemic interests nor the value of simplicity as stand-alone reasons in this context. But then, the explainer in the case at hand would not end up with the most adequate mechanistic artifact explanation. In the first stage, he would have to reject this explanation because it is inaccurate, and because non-epistemic reasons would not be allowed to overrule such an epistemic argument in this stage. So the best explanation, that is, the explanation that serves the explainee's interests best, would never reach the second stage. Since explainers in technological science should be able to arrive at the best explanations, we should allow them to use more than the values of accuracy and internal consistency in deciding whether or not to accept an explanation. Once we allow them to also use values like simplicity, we should also allow them to use non-epistemic interests as stand-alone reasons, as we have seen. The conclusion is that non-epistemic interests are allowed to have a direct influence on the acceptance of explanations.

Chapter 4
Explanations in software engineering[1]

The fact that interests play an important role in the construction and evaluation of explanations is not idiosyncratic to mechanistic artifact explanation. In this chapter, I offer an analysis of explanatory practice in software engineering,[2] in which several different kinds of explanation are considered. I will indicate how interests give rise to certain specific explanation-seeking questions, and how both influence the construction and evaluation of explanations in software engineering. At the end of this chapter, I will use my findings to argue that non-epistemic interests are allowed to have a direct influence on the acceptance of explanations in software engineering.

---

[1] This chapter is based on De Winter (2010), which is an article on a pragmatic account of explanations in software engineering. The reason why I have included parts of that article in this dissertation is the same as the reason why I have included parts of De Winter (2011) (see previous chapter): the article has some important (unexpected) implications with respect to the roles of non-epistemic interests in science. These implications (which are not reported in De Winter (2010)) are discussed at the end of this chapter (sections 4.5 and 4.6).

[2] In this dissertation, it is assumed that software engineering is a technological science. It should, however, be noted that there is a lot of controversy about the scientific status of computer science (see Tedre (2011) for an overview), of which software engineering is a branch. Therefore, I want to remark that the assumption that software engineering is a science is in fact a stronger assumption than strictly necessary for the purpose of this chapter. As it only concerns explanatory practice in software engineering, I only need to assume that explaining in software engineering is scientific. The latter assumption may provoke significantly less controversy.

4.1 Explanation-seeking questions

Let me start by giving some theoretical background. According to the erotetic model of explanation, which I will take for granted here, explanations can be thought of as answers to why-questions. A framework of question formats that covers all why-questions I will consider in this chapter, is formulated by Van Bouwel & Weber (2002), and contains the following formats (each accompanied by an example):

(P-contrast) Why does object *a* have property P, rather than property P'?[3]
e.g., 'Why does the software procedure *GenerateSchedule* have a very long worst-case execution time, rather than a short worst-case execution time?'
(T-contrast) Why does object *a* have property P at time t_1, but property P' at time t_2?
e.g., 'Why does the software procedure *GenerateSchedule* have a very long worst-case execution time, while ten minutes ago, it had a very short worst-case execution time?'
(O-contrast) Why does object *a* have property P, while object *b* has property P'?
e.g., 'Why does the software procedure *GenerateSchedule1* have a very long worst-case execution time, while the software procedure *GenerateSchedule2* has a very short worst-case execution time?'
(plain fact) Why does object *a* have property P?
e.g., 'Why does the software procedure *GenerateSchedule* have a very long worst-case execution time?'

Questions that fit one of the first three formats, are contrastive questions. According to Van Bouwel & Weber (2008), such questions can be motivated by (amongst others) surprise and a therapeutic or preventive need. A question is motivated by surprise if the reason for asking it, is that one wants to explain why things turn out to be different from what we expected. A question is motivated by a therapeutic or preventive need if it is motivated by a desire to know why things are different from how we want them to be, and by the need to know how the correspondence between the actual and the desired state of affairs can be achieved. Possible incentives for asking a plain fact question, are sheer intellectual curiosity, a desire to have information that enables us to predict whether and in which circumstances similar events will occur in the future, and a desire to causally connect object *a* having property P to events with which we are

[3] P and P' are mutually exclusive properties.

a	Conference 1	Conference 2			
	p	u			
	q	v			
	r	w			
b	$p - q$	$q - r$	$r - u$	$u - v$	$v - w$
	$p - q$	$q - r$	$r - v$	$u - v$	$v - w$
	$p - r$	$q - u$	$r - w$	$u - w$	
	$p - r$	$q - v$		$u - w$	
	$p - u$	$q - w$			
	$p - v$				
	$p - w$				

Table 4.1 Teams and games.

more familiar (Van Bouwel & Weber 2008, p. 170). Later in this chapter, we will meet more possible incentives for raising explanation-seeking why-questions.

4.2 Example

In order to clarify how interests result in certain specific explanation-seeking questions, and how both affect the construction and evaluation of explanations in the field of software engineering, I will use an example of a software engineering task, which I consider representative of projects in software engineering. Let me describe it.

There are two conferences in a sports league, each containing three teams, as is depicted in table 4.1a. Every team has to play at least one game against every other team. Teams of the same conference have to play two times against each other, while two teams of different conferences have to compete only once. Table 4.1b presents all the games that have to be played. A schedule has to be generated that meets the following conditions: (1) the schedule contains no more and no less than the games in table 4.1b, (2) a team cannot play more than one game per day, and (3) the games are distributed over as few days as possible ((1) and (2) take priority over (3)).

A software procedure should fulfill the task of generating the schedule. This software procedure should do this in such a way that the content of the schedule is as random as possible. In other words: if n is the number of schedules that meet conditions (1) to (3), then for any schedule that meets conditions (1) to (3), the probability that it is generated by the procedure should be equal to 1/n. The user should be able to execute the procedure by pressing a button, but only after all games on the current schedule have been finished. When the current schedule contains games that haven't been played yet, pressing the button should cause the appearance of the message 'You can only generate a new schedule after finishing all games on the current schedule.'

Several problems might occur while developing such a computer program, and a strategy I consider very useful in solving some of these problems, is to ask certain explanation-seeking questions. The usefulness of this strategy is made clear in the next sections, in which I give examples of interests and explanation-seeking questions that might turn up in the process of writing and debugging software, and pay attention to the nature of the explanations that answer these questions.

4.3 Contrastive questions

In order to come to a first explanation-seeking question, I suggest an algorithm that might allow our software procedure to randomly generate a schedule that satisfies the three conditions mentioned earlier. The first step of the algorithm is the random permutation of the set of games presented in table 4.1b.[4] We can call the resulting array of games *array1*. The second step fills a calendar with days on which games will be played. For each day *d*, the procedure runs through *array1*, checking for each element whether it can be added to *d*. An element can only be added if two requirements are met. The first is that it is a game, and the second is that for both teams, this game should be the first they have to play on *d*. If and only if these two requirements are met, the game is added to *d* and removed from *array1* (replaced by something that is not a game). When all elements of *array1* have been checked, *d* is added to the calendar, and if *array1* still contains games, a new day is created. This process is repeated until all games are part of the calendar. Then, the calendar is converted into a schedule that is accessible for users.

It is possible that one has made some mistakes in implementing this algorithm. In that case, the procedure will probably not have the anticipated result. Suppose an engineer has tried to implement the algorithm, but made a typing error in the program line that should bring about the adding of a game to day *d* if possible. When he tests his procedure, he finds out that it does not work properly. Of course, the engineer does not know what causes the perceived error (otherwise, he would have corrected the typing

[4] An algorithm that can be used to shuffle the games, is Knuth's modern version of the Fisher-Yates algorithm (Radu 2008).

error before testing). In order to debug the procedure, he asks why the procedure led to the error, and not to the desired result (P-contrast question). The answer to this question is that the procedure should contain a program line that actually adds a game to the calendar if possible, while his procedure does not. This explanation is complete if the explainer assumes that the procedure would have worked properly if it contained such a program line. It has the following format:

> Object *a* has property P, rather than P' because it does not have the properties D_1, ..., D_n (with $n \geq 1$).

It is assumed that object *a* would have property P' if it had properties D_1, ..., D_n. I call the combination of this assumption and the explanation format, the P-contrast explanation-type.

In order to solve P-contrast questions, it can sometimes be useful to ask other explanation-seeking questions. To demonstrate this, we can look at how our engineer attains a button that allows the user to generate a new schedule if and only if all games on the current schedule have been finished. Suppose a button with the label 'generate schedule' has already been inserted. Suppose further that when the last game of the regular season is finished, the value of a variable *V* stocked at a certain location in a database, turns from 0 into 1. When a new schedule is generated, the value of *V* turns from 1 into 0. Because up till now, all cases in which the value of *V* is equal to 1, correspond with the cases in which all games on the schedule have been finished, and all cases in which the value of *V* differs from 1, correspond with the cases in which the schedule still contains games that have not been played yet, the consequences of pressing the button can be determined on the basis of whether or not the value of *V* is equal to 1. The engineer knows this and assigns to the pressing of the button a procedure that generates a new schedule if and only if the value of *V* is equal to 1. While testing his program, he finds out that the pressing of the button has the desired consequences.

Later, the engineer decides to extend his program with some post-season events. The games at the post-season events should be added to the schedule, and only when all these games are over, the user should be able to generate a new schedule. During a first attempt at implementing the extension, the engineer deletes the program line that shifts the value of *V* from 0 into 1 when the last game of the regular season is finished, but forgets to insert a line that brings about this shift when the last game of the last post-season event is finished. He discovers that the pressing of the 'generate schedule' button never leads to a new schedule, even when there are no non-played games on the schedule remaining. An answer to the T-contrast question 'Why does the program have, at this moment t_1, the property of never generating a new schedule when the 'generate schedule' button is pressed, while at time t_2 (before extending the program), it had the property of generating a schedule if (1) the 'generate schedule' button was pressed and (2) all games on the schedule were finished?' can help him solve this problem.

The T-contrast can be explained by the fact that at t_2, a program line that leads to the shift of *V*'s value when the last game on the schedule is finished, was present, while such a program line is absent at t_1. This explanation can be fitted into the following format (without changing its meaning):[5]

> Object *a* has property P at time t_1 but P' at time t_2 because it had properties $D_1, \ldots, D_n$ (with $n \geq 1$) at t_2, while these properties are absent at t_1.

In combination with the assumption that object *a* would have property P' at time t_1 if it had properties $D_1, \ldots, D_n$ at t_1, I will call this format the T-contrast explanation-type.

The T-contrast explanation brings the engineer closer to an answer to the question why the pressing of the button does not lead to a new schedule when it should, that is, when all games on the schedule are over (P-contrast question). The answer is that the program should contain a program line that changes the value of *V* into 1 when the last game on the schedule is finished, while it doesn't (P-contrast explanation). The benefit in asking the T-contrast question 'Why doesn't object *a* work properly at this moment t_1, while it did at time t_2?', is that it draws attention to those factors that have been changed during interval $[t_2, t_1]$. Since the program worked properly at time t_2, these factors are more likely to cause the malfunctioning, as is illustrated by the previous example. The fact that the value of *V* did not get changed anymore, caused the malfunctioning, and not, say, a typing error in the program line that converts the calendar into an accessible schedule. One need not pay attention to this last possibility because if this program line would not perform its function, the program would not have worked at time t_2 either, while it did. A T-contrast question can thus function as a tool to address the engineer's attention to those factors that are most likely to be part of an accurate answer to a P-contrast question.[6]

[5] The explanation can be reformulated into: The program has, at this moment t_1, the property of never generating a new schedule when the 'generate schedule' button is pressed, while at time t_2, it had the property of generating a new schedule if (1) the 'generate schedule' button was pressed and (2) all games on the schedule were finished, because at t_2, it had the property of containing a program line that leads to the shift of *V*'s value when the last game on the schedule is finished, while this property is absent at t_1.

[6] The fact that a T-contrast explanation helps the explainer to answer a P-contrast question, does not mean that a P-contrast explanation is in itself insufficient to answer such a question. The P-contrast explanation is a complete answer to the P-contrast question, and this answer should not contain any additional information. The T-contrast explanation is the preferential and complete answer in another context, that is, when a T-contrast question is being addressed.

An O-contrast question can serve the same purpose. Suppose the engineer did not extend the original program with some post-season events, but started a new program, in which he reuses the program code that allows for the random generation of a schedule. The same error as with the T-contrast might arise (pressing a button does not lead to a new schedule when it should). Given that the engineer now has two programs, one that does not generate a schedule at the appropriate times, and one that does, he can ask the O-contrast question 'Why does the new program never generate a new schedule, while the original program generates a new schedule at the appropriate times?' This question draws the engineer's attention to those factors that the two programs do not share (and makes abstraction of all other factors), which are the factors that are most likely to be part of an accurate answer to the P-contrast question 'Why does the new program never generate a new schedule, instead of generating a new schedule at the appropriate times?' Thus, both T- and O-contrast questions can facilitate the debugging process.

The preferential answers to O-contrast questions are O-contrast explanations. An O-contrast explanation has the following format:

> Object *a* has property P, while object *b* has property P' because *b* has properties D_1, ..., D_n (with $n \geq 1$) which object *a* does not have.

An explanation that fits this format is an O-contrast explanation if it is assumed that object *a* would have property P if it had properties D_1, ..., D_n.

4.4 Plain facts

Not only contrastive questions can help one to solve P-contrast questions. To illustrate this, we can return to an earlier stage in the development of the computer program, that is, before our engineer assigned a procedure to the button with the label 'generate schedule', and after he corrected the typing error in the program line that should bring about the adding of a game to the calendar if possible. At this stage, the engineer is still testing the (alleged) implementation of the algorithm I described at the beginning of the previous section. While doing this, he faces a problem: in two tests, his generator generates the two schedules presented in table 4.2. Because schedule 1 contains less days then schedule 2, the engineer knows that his procedure cannot guarantee that a generated schedule meets the third condition (the games are distributed over as few days as possible).

Schedule 1

Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
$p-q$	$p-v$	$p-u$	$u-v$	$p-r$	$q-v$	$u-v$
$v-w$	$u-w$	$r-v$	$q-r$	$q-u$	$p-r$	$r-w$
$r-u$	$q-r$	$q-w$	$p-w$	$v-w$	$u-w$	$p-q$

Schedule 2

Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8	Day 9
$q-r$	$q-v$	$p-r$	$p-q$	$p-u$	$p-r$	$p-v$	$q-w$	$p-q$
$u-v$	$r-u$	$v-w$	$u-v$	$v-w$	$u-w$	$u-w$	$r-v$	
$p-w$		$q-u$	$r-w$	$q-r$				

Table 4.2 Schedules.

The following P-contrast question arises: ‘Why does the software procedure, when executed several times, occasionally result in schedules with more than seven days, instead of always resulting in schedules with exactly seven days?’[7] To know where to search for an answer to this question, the engineer wonders what causes the malfunctioning: a problem with the algorithm, or a problem with the conversion of the algorithm into a software procedure. If the algorithm is problematic, he should develop a new algorithm, and if not, he should correct the implementation error(s). So in order to know how to continue the debugging process, the engineer needs to know whether his algorithm is problematic. This he can find out by trying to explain the shortcomings of schedule 2 by clarifying how the execution of the algorithm could have caused them. If such an explanation is possible, the algorithm is problematic, and if not, there’s something wrong with its implementation.

One of the shortcomings of schedule 2 is that the second day contains only two games.[8] The engineer then asks the plain fact question ‘Why does the second day of schedule 2 contain only two games?’, and tries to answer it by an explanation that clarifies how the execution of the algorithm led to this plain fact (explanation by algorithm). The format of such an explanation is:

[7] Seven days is the minimum of days in a schedule, since each team has to play seven games (see table 4.1b), and a team cannot play more than one game per day (second condition). The fact that a schedule containing seven days is possible, is shown by schedule 1.

[8] Each day should contain three games. We know this because 21 games (see table 4.1b) should be distributed over seven days (see previous footnote), and one day cannot contain more than three games. If there would be more than three games (each between two teams) on a day, there should be more than six teams, since one team cannot play more than one game per day (second condition), and this is not the case (see table 4.1a).

Object *a* has property P because:
At step n, event e_n occurred
(Step n + 1 was the next step because condition $c_{(n+1)}$ was satisfied)
At step n + 1, event $e_{(n+1)}$ occurred
(Step n + 2 was the next step because condition $c_{(n+2)}$ was satisfied)
At step n + 2, event $e_{(n+2)}$ occurred
...
(Step n + m was the next step because condition $c_{(n+m)}$ was satisfied)
At step n + m, event $e_{(n+m)}$ occurred
(with $n \geq 1$; and $m \geq 0$)

An explanation by algorithm assumes that the execution of a certain algorithm caused the fact that object *a* has property P, and explains this fact by referring to events that (1) were instructed by the algorithm, and (2) are relevant to object *a*'s acquiring property P. The reason why the sentences that explain why a certain step was next, are bracketed, is that they are only necessary in cases in which a condition, specified by the algorithm, had to be satisfied for this step to be next.

An explanation by algorithm of the fact that the second day of schedule 2 contains only two games, might go as follows. After the second game was added to the second day of the calendar, the focus moved to the next game of *array1*. Because this game could not complete the second day without leading to a schedule that doesn't meet the three aforementioned conditions,[9] it wasn't added to the second day. Then, the focus moved to the next game of *array1*. This game could not complete the second day either, so that, again, the focus moved to the next game without the second day being extended with a third game. This process continued until all games of *array1* had been checked, after which the composition of the third day started. Thus, no third game was added to the second day before the process of constituting this day ended. Similar explanations can elucidate why days 6 to 9 contain less than three games.

Because the shortcomings can be explained by the execution of the original algorithm, our engineer does not have to control whether his software procedure is a correct implementation of the original algorithm, and can refine his P-contrast question into 'Why does the algorithm, when executed several times, occasionally result in schedules with more than seven days, instead of always resulting in schedules with

[9] All games other than '*p* – *w*' could not complete the second day because they contain teams that already had a game to play on the second day. '*p* – *w*' could not complete the second day either, because it was added to the first day, and interconference games should appear only once on the schedule.

exactly seven days?'[10] An answer to this question is that the algorithm permits that games that together exclude the possibility of a third or a second game being added to one of the days that do not yet contain three games, are added to the definitive calendar. For example, the first two games of the second day of schedule 2 exclude together with the third game of the first day, that a third game is added to the second day.

An algorithm that avoids this mistake, is the following. First, the interconference games are randomly shuffled. Next, for each game of the resulting array, it is checked whether it can be added to the first day. If so, the game will be added to the first day, if not, it will be the first game of a day that does not contain games yet. This step results in a calendar consisting of seven days, with the first day containing three interconference games, and the next six days containing each only one (interconference) game. Then, days 2 to 7 are completed by those intra-conference games that can be added to them without leading to a schedule that does not meet the three aforementioned conditions. Next, for each day, the three games on it are randomly shuffled, and so are the days of the calendar. This calendar is converted into a schedule that is accessible for users. When this algorithm is implemented correctly, one obtains a software procedure that satisfies the requirements stated above.

Answering plain fact questions can also serve other purposes than satisfying the desire to know whether a certain algorithm caused an error. One such purpose is related to the development of a new program by using components of an outdated program. If our engineer prefers creating a new program *a* over updating an old program *b*, and he wants to reuse those components of *b* that together have the function of generating a new schedule, he has to know which part of the original program *b* fulfills this function. The corresponding knowledge-seeking question can be reformulated into the plain fact explanation-seeking question 'Why does the original program *b* have the capacity to generate schedules?'

The answer the engineer is looking for, is a set of reusable software components that are non-redundant parts of the set of all factors that bring about the explanandum. The reason why he wants his answer to contain only reusable software components, and not, say, physical laws, non-reusable software, etc., is that he wants to obtain software that he can use while writing the new program. The reason why the software components should not be redundant, is that the engineer does not want to reuse more than needed.

In most (if not all) cases in which the question of why program *b* has property P, is motivated by the desire to pass on property P to a new program by reusing parts of *b*, the optimal answer will be an inus explanation. According to John Mackie, "what is

[10] Notice that by answering this question, one also answers the original, non-refined P-contrast question.

typically called a cause, is an inus condition, or an individual instance of an inus condition ...” (Mackie 1974, p. 64), with an inus condition being an ‘insufficient but non-redundant part of an unnecessary but sufficient condition’ for the effect to occur. For example, if $(D_1 \wedge D_2) \vee (D_3 \wedge D_4)$ is necessary for object *b* to have property P, while both $D_1 \wedge D_2$ and $D_3 \wedge D_4$ are sufficient for *b* to have property P, then D_1, D_2, D_3 and D_4 are possible causes of the fact that object *b* has property P. An inus explanation is, then, an explanation that refers to some of the factors that are insufficient but non-redundant parts of a conjunction that is not only unnecessary and sufficient for object *b* to have property P, but also true. The format of this kind of explanation is:

Object *b* has property P’ because of factors $D_1, \ldots, D_n$ (with $n \geq 1$).

I can now indicate why in most cases in which the question ‘Why does program *b* have property P?’ is motivated by the desire to pass on property P to a new program *a* by reusing parts of *b*, the optimal answer will be an inus explanation. We already saw why the explanation should only refer to software components that are not redundant: because the engineer does not want to reuse more than needed. These software components will not be sufficient for the program to work in a certain way, because there will always be some non-software-related background conditions (e.g., certain physical laws being active) that have to be met for any computer program to work. The software components are part of a sufficient condition for the occurrence of the explanandum, because, if otherwise, the explanandum would not occur (while it does). In most cases, this sufficient condition will not be necessary to cause the explained effect, since this effect could also be caused by using other algorithms, or other implementations of the same algorithm. The software components that constitute the explanation can thus be considered to be insufficient but non-redundant parts of an unnecessary but sufficient condition for the original program to have the property one wants to pass on to the new program. That the unnecessary but sufficient condition should be true, is evident, since the engineer wants his explanation to point at factors that actually (and not just possibly) caused the fact that program *b* has property P.

4.5 Construction and evaluation of explanations

We have seen how in software engineering, one central non-epistemic interest (the desire to have a software procedure that performs a certain task) can give rise to certain epistemic interests (e.g., the desire to know how to debug the procedure, the desire to know which components of an earlier computer program to reuse), and how these epistemic interests can result in certain explanation-seeking questions. The interests and questions determine what kind of explanation one will construct. If one wants to know how to debug a software procedure, one asks a P-contrast question, which is answered with a P-contrast explanation. In order to construct such an explanation, one has to look for properties that the software procedure currently does not have, and that could eliminate the problem. We have also seen that, in order to answer P-contrast questions, the best strategy can sometimes be to first address other kinds of explanation-seeking questions, such as T-contrast questions, which are answered with T-contrast-explanations, or O-contrast questions, which are answered with O-contrast explanations. It can also sometimes be helpful to ask a plain fact question and try to answer it with an explanation by algorithm, that is, when one wants to know whether a problem is due to the algorithm or to its implementation. One will then check whether certain shortcomings could be explained in terms of the algorithm, regardless of its implementation. Another possible epistemic interest we met is the goal to know which part of an earlier computer program to reuse. One then asks a plain fact question, and tries to answer it with an inus explanation that indicates which components to reuse.

This brief sketch of explanatory practice in software engineering indicates that interests play an important role in the construction of explanations: they determine which questions are asked and what type of explanation is constructed. As such, they determine the kind of information that is sought and included in the explanation.

Interests are also important in the evaluation of explanations. Let me clarify this with an example. Suppose that a software procedure does not work properly and that one wants to fix it, and that one has constructed an explanation of the error that reveals a property that the procedure currently does not have, and that would eliminate the error if it would be implemented in the procedure. This property is all the explanation refers to. As it enables the explainee to satisfy his interest, that is, to fix the procedure, he will consider it a good, fully adequate explanation. Even though a lot more information could be included in the explanation (e.g., the entire mechanism leading up to the error could be described), he will not consider the former, less extensive explanation insufficient, incomplete, or inadequate, since including more information is neither necessary nor useful for his interest to be served. In fact, adding such information would only make the explanation less adequate, as it would then contain a lot of redundant information on which the explainee probably does not want to spend any time and energy (assuming that he has a non-epistemic interest in using his time and energy efficiently).

Table 4.3 presents the epistemic and non-epistemic interests at play in the different contexts we met in this chapter. Each row corresponds with a certain kind of context, in which there are certain interests, and in which a certain kind of explanation-seeking question is asked. The non-epistemic interests to have a software procedure that

Non-epistemic interests	Epistemic interests	Kind of question	Kind of explanation
The desire to have a software procedure that performs a certain task; the desire to use one's time and energy efficiently	The desire to know how to debug the procedure	P-contrast question	P-contrast explanation
	The desire to know the answer to a P-contrast question (in order to know how to debug the procedure)	T-contrast question	T-contrast explanation
	The desire to know the answer to a P-contrast question (in order to know how to debug the procedure)	O-contrast question	O-contrast explanation
	The desire to know whether an error is caused by the algorithm (in order to know how to debug the procedure)	Plain fact question	Explanation by algorithm
	The desire to know which parts of a computer program *b* with property P' should be reused to pass on P' to program *a*	Plain fact question	Inus explanation

Table 4.3 Interests, question-formats and explanation-types.

performs a certain task, and to use one's time and energy efficiently, are present in all contexts. The right hand column shows what kind of explanation is best in each kind of context. Thus, it indicates what kind of explanation will be constructed and evaluated as most adequate given the engineer's non-epistemic interests, his epistemic interest, and the question he asks. So for instance, the first row reveals that a P-contrast explanation will be constructed and considered most adequate in a context in which (1) an engineer wants to have a software procedure that performs a certain task, (2) he wants to use his time and energy efficiently, (3) he wants to know how to debug the current procedure, and (4) he asks a P-contrast question.

The table shows that there is not one explanation-type which is always the best, but that different explanation-types are most adequate in different contexts. We can infer from the table that which explanation-type is most adequate, depends on the kind of question asked and the underlying epistemic interest.

It also depends on the underlying non-epistemic interests (hence the left hand column). This can be demonstrated as follows. Suppose that the explainee would not want to have a software procedure that performs a certain task, and that he would not bother spending his time and energy on activities that are irrelevant for the development of such a procedure. Suppose his main goal would be to kill time. Thus, the non-epistemic interests in table 4.3 would be replaced by the non-epistemic interest to kill time. In that case, the explanation-types in the right hand column would be too concise; e.g., an explanation that only reveals properties which are relevant for eliminating a certain error (P-contrast explanation) would be less adequate for the explainee's interests than an explanation that also describes the entire mechanism leading to the error. The explainee would, due to his desire to kill time, prefer an explanation of the latter kind. Thus, how things (e.g., an error) are to be explained depends on the presence/absence of the non-epistemic interests in the left hand column of table 4.3. The conclusion is that in software engineering, the construction and evaluation of explanations is not only influenced by the kinds of questions asked and the underlying epistemic interests, but also by the non-epistemic interests of the explainee.

4.6 Non-epistemic interests and the acceptance of explanations

The picture presented in this chapter suggests that non-epistemic interests are allowed to have a direct influence[11] on a software engineer's decision to accept a certain explanation. When software engineers want to develop a software procedure that satisfies certain requirements (= non-epistemic interest), they often use explanations that can help them with this, and they usually do not want to spend more time and resources on such explanations than needed (= non-epistemic interest). When such an explanation is most adequate for the engineer's non-epistemic interests (that is, it is most useful for developing the relevant procedure and it does not waste time or resources), the engineer can use this as a stand-alone reason to accept the explanation. His non-epistemic interests then have a direct influence on the acceptance of the explanation. Such a direct influence should not be regarded as dangerous or harmful, on the contrary. It is exactly because engineers use non-epistemic interests to guide their choices that certain goals are accomplished most efficiently and that no resources are wasted. This promotes technological progress. Therefore, direct influences of non-

[11] An interest has a direct influence on a decision if it is used as a stand-alone reason to motivate that decision (see section 2.3).

epistemic interests on decisions concerning the acceptance of explanations should be permitted in software engineering.

Chapter 5
Interests and explanation in the sciences

We have seen that interests, including non-epistemic interests, play certain roles in the construction and evaluation of mechanistic artifact explanations and of explanations in software engineering. Interests such as specific design goals and the desire not to waste time and energy, influence how much and what kind of information is included in an explanation of an artifact behavior or of an artifact property. We have also seen that technological scientists are allowed to use non-epistemic interests as reasons to motivate the acceptance of explanations.

We can then conclude that scientific decisions in the technological sciences should not be 'pure', in the sense that they should not be detached from the non-epistemic interests of certain stakeholders. This conflicts with the full ideal of purity (see section 2.1). But my findings also contest the partial ideal of purity. In section 2.3, I characterized the partial ideal of purity as consisting of the following theses:

(1) An interest *indirectly* influences a decision if it is used to determine whether the available evidence is sufficient to make that decision.
(2) Indirect influences of non-epistemic interests on scientific decisions are harmless (and even desirable).
(3) An interest *directly* influences a decision if it is used as a stand-alone reason to motivate the decision.
(4) Direct influences of non-epistemic interests on decisions concerning the characterization of data, the interpretation of evidence, or theory acceptance are unacceptable in science.
(5) Such influences can and must be eliminated.

At the end of the two previous chapters, in sections 3.5 and 4.6, I argued that direct influences, as characterized by (3), of non-epistemic interests on decisions concerning the acceptance of explanations should be allowed in the technological sciences. This conclusion challenges theses (4) and (5) of the partial ideal of purity. Therefore, I think that we should not only reject the full ideal of purity, but the partial ideal of purity as well.

Some may, however, think that I move too quickly. Perhaps the partial ideal of purity is untenable as an ideal for technological science, but not for other scientific disciplines. The current chapter aims to show that interest influences similar to those identified in technological science in the previous chapters, can also be found in other scientific fields. I will consider the social sciences (section 5.1), biology (section 5.2), and the medical sciences (section 5.3). I use my findings on interest influences in these fields to argue that the partial ideal of purity is untenable as an ideal for science *in general* (section 5.4).

5.1 The social sciences

Social phenomena can be explained in a variety of ways: one can explain a social phenomenon by showing that the rational choices of agents that are driven by self-regarding desires result in the phenomenon, by describing how the actual biographies of particular individuals (who may have behaved irrationally and/or on the basis of other-regarding desires) culminated in the social phenomenon, by referring to another social phenomenon that caused the first one, by describing the social or cultural function of the phenomenon, etc. While some pick out one of these approaches as the one that is always the best one, others prefer explanatory pluralism in the social sciences, i.e. the view that different kinds of explanation are legitimate in this field. To support the latter view, several philosophers of science have argued that different kinds of explanation serve different interests. I think such pragmatic defenses of explanatory pluralism can be used to argue that non-epistemic interests influence decisions in explanatory contexts in the social sciences. So I start by briefly recapitulating some of those defenses.

Jackson & Pettit (1992) make a distinction between individual-level explanations (or micro-explanations) and structural explanations (or macro-explanations). The former explain social phenomena in terms of individuals, and the latter in terms of supraindividual entities, such as institutions, groups, norms, aggregate statistics, etc. To support the view that both kinds of explanation are legitimate in the social sciences, Jackson & Pettit contend that they provide different kinds of information, which are complementary. They state that a macro-explanation offers “modally contrastive information”, that is, it “focuses on similarities between the actual world and other possible worlds; it takes us to a distance at which we can discern constancies across the actual way things are and the ways things might have been" (Jackson and Pettit 1992, p. 15). A micro-explanation, on the other hand, offers "modally comparative information", which means that it "focuses on the differences between the actual world and other possible worlds; it homes in the particularities of the actual case" (Jackson and Pettit 1992, p. 15). This view supports explanatory pluralism in the social sciences: because micro- and macro-explanations offer different kinds of information, they are both legitimate kinds of explanation, and we should not assume that one of the two is always preferable. Which kind of explanation is preferable in a particular context depends, according to Jackson & Pettit, one one's specific perspective and purpose in that context (Jackson & Pettit 1992, p. 16).

Elsewhere, Philip Pettit discusses rational-choice explanation of social behavior, that is, an explanation of such behavior by reference to rational self-regard. According to Pettit, rational choice theory presupposes that people are (1) self-centered, i.e. they do what they do as a result of their own desires or utility functions, and (2) self-regarding, in the sense that the more a desire bears on one's own advantage, the stronger it is. He argues that the second presupposition conflicts with common sense. But that does not lead him to the conclusion that rational-choice explanations should be rejected in the social sciences. Pettit rejects rational-choice explanations as explanations of the emergence or continuation of patterns of behavior, but not as explanations of the resilience of patterns of behavior. He contends that a rational-choice explanation can be used to explain why a certain pattern of behavior is robust under various contingencies. The underlying idea is that even when people's behavior is not dictated by self-regard, but instead, e.g., by cultural perceptions and expectations, self-regard can still be *virtually* present. This is the case when selfish interests function as alarm bells, which ring whenever these interests are possibly compromised, prompting the relevant agent to consider personal advantage. Such virtual presence of self-regard makes certain patterns of behavior resilient under possible disturbance or drift according to Pettit. His conclusion is then that rational-choice explanations of social behavior are highly useful to explain why this behavior is resilient under possible disturbance or drift. Therefore, rational-choice explanation is, in his view, a legitimate kind of explanation in the social sciences (Pettit 1995, 2000).

Satz & Ferejohn (1994) discuss rational-choice explanation in the social sciences as well. They argue against the view, which they claim is common among philosophers, that rational-choice explanations are necessarily individualist, psychological explanations, i.e. explanations of social behavior in terms of the mental states of individuals and their interactions. They claim that rational-choice explanations can also be structural explanations, which account for social behavior in terms of irreducible relational or structural properties. More specifically, they argue that a rational-choice explanation can explain the stability of a pattern of behavior by spelling out structural conditions that constrain possible behaviors. This corresponds with Pettit's view that a rational-choice explanation is useful to clarify why a pattern of behavior is resilient (see above). The difference with Pettit's view is that according to Satz & Ferejohn, rational-choice explanations can also be legitimate psychological explanations, and they ask for a background theory that specifies in which contexts such psychological rational-choice explanations are appropriate (Satz & Ferejohn 1994, p. 87). They suggest that this depends on the interests at play:

> Whether individual or structural accounts of social phenomena are appropriate depends, we believe, on the purpose of the explanation. For some purposes, the appropriate focus is on individual agency and choice. For many social-science questions, however, the appropriate focus is on how social structures and features of the agent's environment exert constraints on her action. This is not an a priori feature of social-science explanation. Rather, it is a feature of the types of questions in which social scientists are typically interested. In contemplating a reduction, we need always to focus on the question of what purposes our explanation is intended to serve. (Satz & Ferejohn 1994, pp. 83-84)

The view that which kind of explanation is appropriate depends on the specific interests and purposes that the explanation has to serve, is also endorsed by Uskali Mäki. He puts it as follows:

> [T]he account of explanatory relevance cannot be based on an account of causal relevance only. It would most naturally be given in terms of explanatory pragmatics, in terms of the explainers' and explainees' interests and purposes. While macrosociologists and ministries of justice will be interested in explaining crime rates in terms of unemployment rates, individual psychologists and social workers will want explanations given in terms of theoretically understood individual biographies. Different kinds of explanatory information serve different purposes. (Mäki 2002, pp. 255-256)

These positions constitute a pragmatist argument for explanatory pluralism in the social sciences: because different kinds of explanation provide different kinds of information, which are useful for different kinds of interests, they are all legitimate, and we should not assume that only one of them is appropriate in the social sciences. Such a pragmatist argument for explanatory pluralism in the social sciences has been developed in more detail by Van Bouwel & Weber (2008).

Van Bouwel & Weber make a distinction between intentional explanations (which explain a social phenomenon in terms of the intentional states and actions of one or more individuals), structural explanations (which explain a social phenomenon by referring to another social phenomenon that caused the first one), and functional explanations (which explain a social phenomenon by describing its function). Furthermore, structural explanations that refer to *proximate* causes are distinguished from structural explanations that refer to *remote* causes. Van Bouwel & Weber propose four criteria to assess different explanations of a social phenomenon: accuracy (i.e. their correctness, or their relation with reality), adequacy (i.e. how well they serve the interests they are supposed to serve), efficiency (i.e. the amount of work that is needed to construct the explanation), and appropriateness (i.e. their relevance for the explanation-seeking question asked). These criteria determine, according to Van Bouwel & Weber, how good an explanation is (compared to other explanations) within a certain context. It is shown that different kinds of explanation are the best in different contexts, depending on the particular interests of the explainee and the specific kind of explanation-seeking question that is addressed in the context (Van Bouwel & Weber 2008).

So one of Van Bouwel & Weber's criteria to evaluate explanations in the social sciences is adequacy for the relevant interests. They sum up several interests that explanations of social phenomena could serve: the desire to know how the social phenomenon was produced given antecedents via spatiotemporally continuous processes, the desire to know whether and in which circumstances similar events will occur in the future, the desire to know how object *a* having property P is causally connected to events with which we are more familiar, and the desire to know why things have been otherwise than we expected them to be. These are all epistemic interests, but Van Bouwel & Weber also mention two non-epistemic interests: the desire to anticipate actions of persons/groups, and the desire to change a property of an object.[1] Depending on which of these epistemic and non-epistemic interests one wants to serve in an explanatory context, an explanation of a social phenomenon is regarded as a (relatively) good or a (relatively) bad explanation.

But I think that what we have seen so far doesn't capture the whole spectrum of interest influences on how good an explanation is within a certain context. First, consider Van Bouwel & Weber's third criterion to evaluate explanations: efficiency. This criterion is, I think, the second criterion (adequacy) in disguise. Efficiency can be reformulated as adequacy for the (non-epistemic) interest not to carry out more work

---

[1] Van Bouwel & Weber call these two interests *epistemic* interests. But in my framework (see section 1.1.1), they are *non-epistemic*, because they are not interests to know something. The potential state that is regarded as beneficial in these interests, is not the state of knowing something, but, respectively, the state in which certain actions are anticipated, and the state in which a certain property of an object is changed. The reason why Van Bouwel & Weber and I use different concepts of epistemic interests, is that our aims are different (their aim is to defend explanatory pluralism in the social sciences, and my aim is to discuss interest influences in the social sciences). Van Bouwel & Weber's concept of epistemic interests is not adequate for the aims of this dissertation, and therefore I use a different concept.

than needed. It is because social scientists usually have this interest, that efficient explanations are usually preferable. However, if they would not have this interest, then efficiency may not be an explanatory virtue. Suppose someone wants to construct an explanation of a social phenomenon just to kill time. In such a context, a relatively inefficient explanation, which requires more work, may be preferable, since such an explanation is more adequate for the interest to kill time. So which explanation is best, depends on the presence/absence of the interest not to carry out more work than needed.

Van Bouwel & Weber's fourth criterion – relevance for the explanation-seeking question asked – can also be used to bring certain interest influences to light. Van Bouwel & Weber show that a certain explanation of a social phenomenon, which is a good, relevant answer to a certain explanation-seeking question concerning the phenomenon, can be irrelevant for another explanation-seeking question concerning the same phenomenon. For instance, a functional explanation of object a having property P may be a good and relevant answer to the explanation-seeking question 'What is the function of object *a* having property P?', but not to the explanation-seeking question 'What is the cause of object *a* having property P?' Therefore, whether an explanation is appropriate in a certain explanatory context, depends on the specific explanation-seeking question to be addressed in the context (Van Bouwel & Weber 2008).

But for appropriateness for a certain explanation-seeking question to be an explanatory virtue in a certain context, it is important that it is legitimate to ask that question in the first place. This depends on the interests that are present in the context. For instance, if one wants to legitimize or criticize the existence and/or properties of certain institutions or patterns of behavior (= non-epistemic interest), then it is legitimate to ask what their function is, because a functional explanation of the existence/properties of these institutions/patterns can help one satisfy this non-epistemic interest (Van Bouwel & Weber 2008, p. 178). But in a context in which one has other interests, interests for which a functional explanation would be useless, it is not legitimate to seek a functional explanation (by asking what the function of the relevant social phenomenon is). This points at an influence of interests on which explanation is best: the interests that are present in a certain context determine what kind of explanation-seeking question would be appropriate to ask, and this question in turn determines which kind of explanation would be appropriate. Thus, interests affect, via explanation-seeking questions, which kind of explanation is best.

My general conclusion is that how good an explanation is within a certain context, is highly dependent upon the different interests, including non-epistemic interests, that are present in the context. But do social scientists actually use non-epistemic interests in evaluating explanations? And should we allow them to use such interests as stand-alone reasons in decisions concerning the acceptance of explanations?

The explicit use of non-epistemic interests in explanatory contexts seems significantly less common in the social sciences than it is in technological science. While technological scientists are often driven by certain non-epistemic goals (e.g., designing a new artifact), and tailor their explanations to these specific goals (see chapters 3 and 4), social scientists often work in explanatory contexts in which non-epistemic interests are less dominant, and in which one approach to explanation is taken for granted. Nevertheless, I think that even in the social sciences, there are contexts in which non-epistemic interests play much the same roles as they play in technological science.

Consider, for instance, contexts in which an external party (e.g., a government body) consults social scientists as scientific advisors on how to realize certain non-epistemic social goals (e.g., to reduce crime rates). In order to develop decent scientific advice, the consulted scientists will use scientific explanations of the relevant social phenomena. If the relevant non-epistemic goal is, say, to reduce crime rates, then they will build on explanations of current crime rates. But not every kind of explanation will be equally helpful. For instance, if the purpose is to implement a policy that leads to lower crime rates, then an explanation in terms of unemployment rates is much more useful than a psychological explanation in terms of individual biographies. The social scientists will then focus on those explanations which are most helpful. They favor such explanations on the ground that they are most useful for the non-epistemic goal on which they have to give scientific advice.

Let me now turn to the question whether social scientists should be allowed to use non-epistemic interests as stand-alone reasons in decisions concerning the acceptance of explanations. My answer is that this should be allowed. For instance, if social scientists are consulted to develop decent scientific advice on how to reduce crime rates, then they are allowed to accept an explanation of current crime rates on the ground that it is most useful for the non-epistemic goal to reduce crime rates. Such a direct influence of non-epistemic goals on the acceptance of explanations leads to an optimal realization of the relevant non-epistemic goals, and therefore, it should be allowed.

In order to further support this view, consider two ways in which the opposite view (i.e. that a direct influence of non-epistemic interests on the acceptance of explanations should *not* be allowed in the social sciences) could be developed. Firstly, one can start from a distinction between the decision whether or not one accepts an explanation, and the decision how good/adequate an accepted explanation is, assuming that non-epistemic interests are only allowed to have a direct influence on decisions of the latter kind, and not on decisions of the former kind. Decisions of the former kind should be based solely on the accuracy of the explanation.

Perhaps the easiest way to challenge such a position is on the basis of Pettit's view on rational-choice explanation (see above). As we have seen, Pettit holds that rational choice theory rests on an implausible presupposition, namely, that people are self-regarding. Suppose, in line with this view, that social scientists have developed a rational-choice explanation of a social phenomenon which presupposes that people are self-regarding with respect to this phenomenon, and that this presupposition is shown to be inaccurate, in the sense that we have evidence which demonstrates that people are not self-regarding with respect to the relevant phenomenon. If accuracy would then be the only criterion that is allowed to guide decisions on the acceptance of explanations in the social sciences, then the scientists' conclusion would have to be that they do not accept the rational-choice explanation.

But, as Pettit has shown, even though rational choice theory rests on an implausible presupposition, rational-choice explanations can still be highly useful to explain the resilience of social phenomena. Suppose this also applies to the case at hand: although the rational-choice explanation is inaccurate, it is still adequate to explain the resilience of the relevant social phenomenon. So in a context in which one is interested in the question why this phenomenon is resilient under possible disturbance or drift, the rational-choice explanation can be a good, fully adequate answer. Therefore, I think that social scientists should be allowed to accept the rational-choice explanation, despite its inaccuracy. This shows that in deciding whether or not to accept an explanation, social scientists should be permitted to look at more than the accuracy of the explanation.

This brings us to a second version of the idea that social scientists should not be allowed to use non-epistemic interests as stand-alone reasons for accepting an explanation. This version states that in motivating decisions on the acceptance of explanations, social scientists are allowed to use other values besides accuracy, such as adequacy, efficiency, and appropriateness. The idea is that these values are epistemic values, and that decisions on the acceptance of explanations should be made solely on the basis of such epistemic values; non-epistemic interests should not directly influence such decisions. The view under consideration does allow non-epistemic interests to *indirectly* influence such decisions, since the values of adequacy, efficiency, and appropriateness are linked to non-epistemic interests, as we have seen above.

But once we acknowledge this, it no longer makes sense to prohibit a direct influence of non-epistemic interests on the acceptance of explanations. Let me explain. The view under consideration allows the following indirect interest influence. One of the reasons that a social scientist uses to motivate the acceptance of an explanation *e* is that he considers *e* highly adequate, and he considers *e* highly adequate because *e* serves the non-epistemic goal it is supposed to serve very well. This seems to come down to the same thing as using the judgment that *e* serves the relevant non-epistemic goal very well as a reason, in itself (not via the value of adequacy), to motivate the acceptance of *e*. The latter involves a direct influence of a non-epistemic interest on the acceptance of an explanation. As it comes down to the same thing as the indirect influence, it does not make sense to allow the indirect influence but not the direct influence.

5.2 Biology

Now consider a second discipline: biology. Ingo Brigandt (2013) endorses different kinds of pluralism in biology. The first states that in the field of biology, different explanations may be of a different type. Brigandt distinguishes two types of explanation: mathematical explanation and causal-mechanistic explanation. In mathematical explanations, individual phenomena are explained as being instances of quantitative generalizations. Such explanations can be found in evolutionary biology and ecology. He states that "[m]athematical models in these fields typically represent the dynamics of biological systems, such as change in gene frequencies, the size and structure of populations, or the geographical distribution of several species" (Brigandt 2013, p. 73). Causal-mechanistic explanations, on the other hand, account for phenomena in terms of mechanisms. A mechanism consists of "several entities, often different kinds of entities, and their causal interactions, such as mechanical or chemical interactions" (Brigandt 2013, p. 73). In a causal-mechanistic explanation, the outcome state of a token mechanism or the type of behavior that is regularly produced by a type of mechanism, is explained by a scientific representation of relevant aspects of the mechanism. Causal-mechanistic explanations are, according to Brigandt, widespread in physiology, developmental biology, and molecular biology (Brigandt 2013, p. 73).

A second, broader kind of pluralism that Brigandt mentions is the view that "biology is characterized by and in fact needs a plurality of methods, theories, and explanatory perspectives, as opposed to a unique and overarching approach" (Brigandt 2013, p. 78). The idea is that biology consists of a diversity of fields and theoretical approaches (e.g., evolutionary biology, physiology, developmental biology) because reduction to one fundamental field (e.g., molecular biology) is impossible. Moreover, even within one field, different concepts, methods, modes of reasoning, and explanatory approaches are required (Brigandt 2013, p. 78).

Both kinds of pluralism can be defended on the ground that different kinds of explanation/theoretical approaches are needed to explain different phenomena. But one can also argue, as Brigandt does, that "different explanatory frameworks may be used for a single phenomenon, because there are different legitimate explanatory aims and questions that different scientists can pursue in the study of the phenomenon" (Brigandt 2013, p. 88). The idea is that different explanatory frameworks are needed to satisfy the different explanatory aims that motivate and guide research on the phenomenon.[2]

The different explanatory frameworks that are legitimate to approach a certain phenomenon in biology may be complementary, but this is not *necessarily* the case. It is possible that different legitimate explanatory models make conflicting claims about a phenomenon (Brigandt 2013, p. 88). Let me clarify this.

Explanations in biology often make idealizations by abstracting away from certain aspects of the phenomenon to be explained. They focus on the most important aspects of this phenomenon and leave out irrelevant details. The purpose of idealization is to make the situation represented less complex, and to permit quantitative models that are mathematically more tractable. An example is the assumption of some models in population genetics that populations consist of an infinite number of organisms (even though actual populations are finite), which is made for the sake of mathematical simplicity (Brigandt 2013, pp. 82-83).

Whether a certain idealization is legitimate depends on the aim that the explanation is supposed to serve. An explanation should pick out those factors that are relevant for this aim, and it is allowed to ignore empirical details that are irrelevant for this aim (Brigandt 2013, p. 83). So if empirical fact f is relevant for explanatory aim a_1, but not for explanatory aim a_2, then ignoring f is legitimate with respect to a_2, but illegitimate with respect to a_1. Suppose then that we have two explanations of a certain phenomenon p, one which refers to f and satisfies aim a_1, and one which makes an idealization that conflicts with f and satisfies aim a_2. As both explanations satisfy a certain explanatory aim (a_1 and a_2 respectively), they are both legitimate explanations of p, even though they make incompatible claims with respect to f. Brigandt gives the following example to illustrate this idea:

> For example, descriptions in developmental biology break the development of a model organism species into normal stages (e.g., molt-to-molt intervals in insects). While this is useful for explaining development, the distinct explanatory aim of accounting for the evolution of development may require a different representational framework because developmental biology's normal stages obscure natural variation in development, phenotypic plasticity, and the way in which developmental stages are created and transformed in evolution (Love 2009, 2010, this issue). Explaining the development of a species and explaining the evolution of this

[2] This pragmatic argument for the need of different explanatory frameworks in biology is also offered by Steel (2004, p. 67). For a pragmatic defense of the view that mechanistic explanations (i.e. explanations in terms of the underlying mechanism) and inferential explanations (i.e. explanations in terms of generalizations from which the explanandum can be inferred) are both legitimate in biology, see Gervais & Weber (2013).

> species' development are two different explanatory aims about the same empirical object, and different representations of this object are needed to meet the respective aims. (Brigandt 2013, p. 83)

The explanatory aims guiding research in biology that Brigandt has in mind seem to be epistemic; one wants to know something about a certain phenomenon and this epistemic interest determines how the phenomenon is to be explained and which idealizations are legitimate. But non-epistemic interests can play in research contexts in biology as well. In chapter 1, I already mentioned that the Human Genome Project was (partly) justified on the basis of the expectation that it would serve the non-epistemic interest to more effectively cope with certain diseases (see section 1.2.1). Other non-epistemic interests that can play in biology are the desire to change certain properties of certain organisms (e.g., via breeding or genetic modification), the desire to have more effective tools for improving animal health and well-being, the desire to anticipate biological events, etc. There are different ways in which such non-epistemic interests can be present in research contexts in biology. For instance, an organization may financially support a certain biological research project because it expects this project to contribute to a certain non-epistemic aim, and the organization's continued support may depend on how well the project meets this expectation. It is also possible that a biological scientist postulates a certain non-epistemic interest (e.g., contributing to animal health and well-being) as his personal goal, which drives his research (even though financial support of this research does not depend on how well the research serves this goal).

The question is then: Should we allow such non-epistemic interests to have a direct influence on the acceptance of explanations in biology? In other words: Is it acceptable for biological scientists to use non-epistemic interests as stand-alone reasons for (not) accepting an explanation? I think it is. In deciding whether or not to accept an explanation, biological scientists should not use accuracy as the only criterion, because then they could not accept explanations which make inaccurate, but useful idealizations. Accordingly, they should be allowed to look at more than accuracy. My proposal is that they should be allowed to also consider how useful the explanation is for the interests it is supposed to serve, and this could include non-epistemic interests, as we have seen. The judgment that an explanation serves a certain non-epistemic interest very well can then be a stand-alone reason to motivate the decision to accept the explanation. This should be permitted because it leads to an optimal realization of the interests at play in the relevant research context.

5.3 The medical sciences

A last discipline I will discuss in this chapter is the field of the medical sciences. De Vreese, Weber & Van Bouwel (2010) distinguish between three kinds of explanation in the medical sciences: (non-reductive) macro-explanations, non-reductive micro-explanations, and reductive micro-explanations. A macro-explanation refers to a property that all members of a group have in common, but which is external to the individuals in the group, in the sense that if an individual leaves the group and joins the other group, he will no longer possess the property. A micro-explanation, on the other hand, refers to differences among individuals *within* a group. While a *non-reductive* micro-explanation refers to the behavior, lifestyle, habits, etc. of these individuals, a *reductive* micro-explanation refers to properties that require parsing an individual in terms of his or her biological make-up rather than externally observable characteristics and behaviors (De Vreese, Weber & Van Bouwel 2010, pp. 373-374).

De Vreese, Weber & Van Bouwel illustrate these three kinds of explanation on the basis of the example of skin cancer. An example of a macro-explanation is then the following. In a group of Belgians spending their summer holidays each year in the Mediterranean, the incidence of skin cancer is higher than in a group of Belgians staying in Belgium during the summer holidays, because the exposure to sun rays is higher in the former group. If, on the other hand, we explain differences in the development of skin cancer among people exposed to a high amount of sun rays each year, by the fact that some have the habit of using suntan lotion while others do not, we offer a non-reductive micro-explanation. If we explain the fact that some of the Belgians spending their holidays in the Mediterranean develop skin cancer despite using suntan lotion, by referring to genes which make them more susceptible than others to risk factors such as excessive exposure to sun rays, then we offer a reductive micro-explanation (De Vreese, Weber & Van Bouwel 2010, pp. 373-374).

Which kind of explanation is best to explain a medical condition depends on the interests that the explanation ought to serve. As De Vreese, Weber & Van Bouwel (2010, p. 388) show, different interests lead to different explanation-seeking questions, and these require different kinds of explanation. Interests that motivate the request for non-reductive explanations are the following: the desire to have long-term rather than short-term solutions to health issues, the desire to have more efficient strategies for the prevention of disease at the population/policy level, the desire to provide whole societies with help rather than only those individuals that are able to afford healthcare, the desire to improve people's health in general rather than tackling each single health problem separately when it arises, and the desire to reduce the society's costs for healthcare. All of these interests are non-epistemic interests for which non-reductive explanations are most useful. There is also an epistemic interest for which non-reductive explanations are useful: the interest in a more complete understanding of the development of human disease (De Vreese, Weber & Van Bouwel 2010, p. 381).

To have a more complete understanding of the development of human disease, we also need reductive explanations. Hence, reductive explanations are useful for the latter epistemic interest as well. Other epistemic interests that reductive explanations serve, are the desire to diagnose disease and the desire to predict disease development and disease progress at the individual level. Knowing from which disease a person suffers and how it will develop and progress is useful for timely intervention. This means that reductive explanations also serve the non-epistemic interest to intervene in time. Another non-epistemic interest that reductive explanations serve is the desire to have easy and efficient treatments for diseased individuals (De Vreese, Weber & Van Bouwel 2010, p. 381). Finally, reductive explanations can help one develop a lucrative solution to a health problem (e.g., a patentable pill) (De Vreese, Weber & Van Bouwel 2010, p. 383).[3] Hence, they serve the financial interests of medical innovators. All of these epistemic and non-epistemic interests can motivate medical innovators to pursue a reductive explanation of a medical condition.

What does this say about the roles of non-epistemic interests in the medical sciences? We have seen that there is a variety of health-related non-epistemic interests that are served by explanations in this field: the desire to have long-term solutions to health issues, the desire to have efficient strategies for the prevention of disease at the population/policy level, the desire to have easy and efficient treatments for certain health conditions at the individual level, etc. These interests can play in research contexts in the medical sciences. A medical scientist can for instance regard contributing to one of these non-epistemic interests as his personal goal, which guides his research. Non-epistemic interests can also be present in a more structural way, e.g., when a research project in medical science is supported on the ground that it is expected to contribute to a certain non-epistemic aim, and when continued support depends on how well the project meets this expectation.

The following question arises: Where such non-epistemic interests are present in research contexts in the medical sciences, should they be allowed to directly influence decisions on the acceptance of explanations? I think they should. For instance, if a medical scientist aims to contribute to the development of an easy and efficient

[3] De Vreese, Weber & Van Bouwel (2010, p. 383) refer to patentability in the context of the question to what extent *unjustified* interests guide research and practice towards a reductive approach. They do, however, not clarify why the interest in a patentable explanation and remedy is unjustified, while other interests *are* justified. Why are the health interests of (potential) patients allowed to play a role in explanatory contexts, but an innovator's interest in a patentable explanation and remedy not? Since I do not see how the former interests are justified, contrary to the latter, I drop the justified/unjustified distinction.

treatment for a certain medical condition at the individual level, and a certain reductive explanation of the condition is most useful for this non-epistemic aim, then he should be allowed to use this as a stand-alone reason to accept the explanation (instead of, e.g., searching for other factors to explain the condition). I consider such behavior acceptable because it contributes to an optimal realization of the relevant non-epistemic aim. Accordingly, direct influences of non-epistemic interests on the acceptance of explanations should be permitted in the medical sciences.

I would even go a step further and state that reasons which are based on certain non-epistemic interests should in principle be allowed to overrule reasons concerning the accuracy of the explanation. If an explanation rests on an assumption which is shown to be inaccurate, then this can be a reason to not accept the explanation. But if the inaccuracy is irrelevant for the non-epistemic aim that the explanation is supposed to serve, and if a more accurate explanation would only be more complex and therefore less useful for this aim, then the explainer should be allowed to accept the explanation in its current form, despite the inaccuracy. The argument that the explanation should not be accepted because it is inaccurate is then overruled by the argument that it should be accepted because it is most useful for a certain non-epistemic aim. Allowing such behavior facilitates the realization of certain valuable non-epistemic goals, of which there are plenty for the medical sciences, as we have seen. Therefore, I think we should allow such behavior in the medical sciences.

5.4 Implications

The previous chapters show that direct influences of non-epistemic interests on the acceptance of explanations should be allowed in technological science. As indicated in the introduction of this chapter, this finding undermines the partial ideal of purity. More specifically, it challenges the theses that direct influences of non-epistemic interests on theory acceptance are unacceptable in science, and that such influences must be eliminated. Some may, however, cling to the ideal of purity by stating that the partial ideal of purity is only shown to be untenable as an ideal for technological science, but not as an ideal for other scientific disciplines. In the current chapter, I have attacked this view by arguing that direct influences of non-epistemic interests on the acceptance of explanations should also be allowed in the social sciences, in biology, and in the medical sciences. Therefore, I think we should reject the partial ideal of purity as an ideal for science *in general*.

Instead of further clinging to the ideal of purity by finding new ways to save parts of it, I think that it is better to fully abandon this ideal. Part 1 shows that its main assumptions are problematic, and once these assumptions are rejected, I see no reason to further hold on to it. In light of the many different kinds of interest influences considered in chapters 2-5, we should accept the fact that non-epistemic interests form an integral part of reasoning within scientific research. Their influence on scientific decisions is not something that we should necessarily be suspicious about.

This insight contains an important lesson for part 2 of this dissertation. Part 2 aims to contribute to the task of distinguishing *acceptable* from *unacceptable* interest influences in science. As indicated in the introduction, the influence of interests can sometimes be troublesome in science. To illustrate this, I offered the example of the Vioxx case, in which the influence of Merck's short-term financial interests on the research process is clearly problematic. Now, it may be tempting to explain this problem by stating that non-epistemic interests such as Merck's financial interests are external to scientific research, and that they are not allowed to play a role in it. Because they did play a role in the research on Vioxx, so the reasoning goes, their influence is problematic.

But this line of reasoning rests on a false presupposition, i.e. that non-epistemic interests should not play a role within scientific research. I have shown that non-epistemic interests form an integral part of scientific research. Hence, the fact that non-epistemic interests affected the research on Vioxx is in itself not sufficient to conclude that an unacceptable interest influence occurred in this case. We need new criteria to determine when exactly the influence of non-epistemic interests on the research process is problematic and when it isn't. The next part aims to develop one such criterion, namely, a criterion to determine when interest influences in science are *epistemically* problematic.

Part 2

EPISTEMIC INTEGRITY

Chapter 6
Developing a concept of epistemic integrity[1]

Perhaps the most obvious criterion to determine whether an interest influence in science is epistemically problematic, is to check whether or not it hinders the advancement of scientific knowledge; if it does then it is epistemically problematic, and if it doesn’t then it is not epistemically problematic. But this is not a very satisfying criterion. For instance, if a brilliant scientist decides to end his career in order to focus on his family life, then his family-related interests prevent him from making new valuable contributions to the body of scientific knowledge. In that way, these interests hinder the advancement of scientific knowledge. It seems, however, strange to conclude that therefore the influence of the scientist’s family-related interests is epistemically problematic. The decision to stop doing science does not seem to be an epistemic failing, and so it is not because certain interests cause such a decision, that their influence is epistemically problematic. With the notion of epistemically problematic interest influences, I aim to capture more disturbing interest influences; interest influences which hinder the advancement of knowledge in more disturbing ways. Hence, I need an alternative criterion to determine when exactly an interest influence is epistemically problematic. In this chapter, a new concept of epistemic integrity, which can function as such a criterion, is developed.

[1] In this chapter, I pave the way for and present a new concept of epistemic integrity. This concept differs from the concept of epistemic integrity developed in De Winter & Kosolosky (2013a, 2013b). I distance myself from the latter concept because I learned that it has an important shortcoming. This shortcoming is discussed in section 6.4.2. In section 6.7.2, I show that the new concept of epistemic integrity does not have this shortcoming.

	Ethical	Epistemological
Agents	Moral integrity of scientists and their institutions	Epistemic integrity of scientists and their institutions
Behavior	Moral integrity of the research process	Epistemic integrity of the research process

Table 6.1 Four kinds of scientific integrity.

6.1 Four kinds of scientific integrity

Let me start by distinguishing four kinds of scientific integrity, in order to clarify which specific kind I will try to explicate (and which kinds I will not try to explicate). There are at least four categories of concepts of scientific integrity. These categories can be distinguished along two dimensions (see table 6.1). The first dimension is the distinction between ethical and epistemological concepts. Ethical concepts focus on the moral good. The moral good is here understood as the good in general, that is, the good in light of all considerations (epistemological considerations, sociocultural considerations, economic considerations, etc.). Ethical concepts of scientific integrity aim to capture what is required for science to be in accordance with this general good. Epistemological concepts, on the other hand, focus on the epistemic good. The epistemic good is what is good in light of epistemological considerations, that is, what is good for the purpose of generating knowledge. Epistemological concepts of scientific integrity try to indicate what is required for science to be adequate for this purpose.[2]

The second dimension is the distinction between concepts that focus on agents and those that focus on behavior. Concepts of scientific integrity that focus on agents try to capture the integrity of the agents that participate in the research process, such as scientists and scientific institutions. Concepts of scientific integrity that focus on behavior, on the other hand, aim to describe the integrity of the research process itself, rather than of the agents participating in it. A research process is then understood as a set of behaviors that constitute a research project, including the research activities that are part of this project as well as the communication of the results of these activities.[3]

[2] This distinction between ethical and epistemological concepts of scientific integrity is inspired by Resnik's (1996) discussion on the connection between research ethics and epistemology.

[3] For a list of different activities that could be part of a research process, see PSRCR (1992, pp. 17-18).

On the basis of these two dimensions, four kinds of scientific integrity can be distinguished: (1) the moral integrity of scientists and their institutions, (2) the moral integrity of the research process, (3) the epistemic integrity of scientists and their institutions, and (4) the epistemic integrity of the research process (see table 6.1). The focus here is on the latter kind: I aim to develop a concept of epistemic integrity of the research process. Such a concept will enable us to identify epistemically problematic interest influences. More specifically, the idea is that the influence of an interest on the research process is epistemically problematic if the interest compromises the epistemic integrity of the research process.

Three remarks are in place. Firstly, by distinguishing four kinds of scientific integrity, I do not mean to say that they are unrelated; there can be certain links between the different kinds of scientific integrity. One can, for instance, characterize epistemic integrity as a component of moral integrity. The idea could then be that having (the highest degree of) epistemic integrity is necessary but not sufficient for an agent or research process to have (the highest degree of) moral integrity, as moral integrity also has other components in addition to epistemic integrity. Furthermore, concepts that focus on agents can be linked to concepts that focus on behavior. One can, for example, define the (moral or epistemic) integrity of the research process in terms of the (moral or epistemic) integrity of the agents involved in it. This is in accordance with virtue theories in ethics and epistemology: in virtue ethics, the moral properties of a behavior are explained in terms of the properties of the agent performing the behavior, and in virtue epistemology, the normative properties of a cognitive performance are explained in terms of the properties of the cognizer (Greco & Turri 2013).[4] Here, I do not wish to make any presuppositions on how the different kinds of scientific integrity are exactly related to each other. I only want to note that they need not be strictly separated.

Secondly, I want to clarify why I characterize the research process as a process which does not only include certain research activities, but also the communication of the results of these activities. For the purpose of this dissertation, it is useful to think of research and the communication of its results as one whole, instead of regarding them as distinct behaviors which should be evaluated separately. This enables me to make abstraction of which part of the research process I regard as problematic in a certain case. Consider, for instance, a process in which (1) data are made up, and (2) these data are presented as if they resulted from genuine empirical research. This process is obviously problematic. There are two ways to account for this. The first is to state that if data were made up, then one should not present them as if they resulted from genuine

[4] A specific example of a virtue epistemological approach to the notion of epistemic integrity of the research process is discussed in section 6.6.

empirical research. The second is to state that if one will present data as if they resulted from genuine empirical research, then one should not make them up. So how one accounts for the view that the process consisting of (1) and (2) is problematic, depends on which part one takes for granted; (1) or (2). Since the statement 'the process consisting of (1) and (2) is problematic' is compatible with both accounts, it makes abstraction of which part of the process is problematic. This is useful for the purpose of this dissertation because I just want a concept which enables us to determine whether or not a certain interest influence on the research process is epistemically problematic; I do not need to specify which part of the research process is problematic due to this interest influence.

Thirdly, I should mention that I wish to characterize epistemic integrity as a property that comes in degrees: a research process can have a higher or lower degree of epistemic integrity. We can then say, for instance, that a research process in which both the recording of data *and* the reporting of the recorded data are sloppy, has lower epistemic integrity than a research process in which only the recording of data is sloppy, and that the latter research process has lower epistemic integrity than a research process in which neither the recording of data nor the reporting of the recorded data is sloppy. Accordingly, what I will try to define in the remainder of this chapter is the degree of epistemic integrity of the research process. When I say that an interest compromises the epistemic integrity of the research process, then I mean that it causes the degree of epistemic integrity of this process to decrease. By the ideal of epistemic integrity, I mean the ideal that scientific research processes have the highest degree of epistemic integrity.

6.2 Integrity as purity

6.2.1 Definition

In order to develop a concept of epistemic integrity of the research process, let us start from the etymological roots of the term 'integrity'. It stems from the Latin word 'integritās', which is related to the word 'integer', meaning an indivisible whole. The Latin word 'integer' is in turn probably related to 'in' and 'tegere', or 'in' and 'tangere', which stands for 'not' 'touch'. Accordingly, what has integrity is untouched, uncorrupted, pure (Kasulis 2002, p. 25). In other words, having integrity means "being able to stand alone, having a self-contained identity without dependence on, or infringement by, the outside" (Kasulis 2002, p. 53). Stéphanie Ruphy's definition of the epistemic integrity of science is in accordance with these etymological roots: she defines the epistemic integrity of science as “its capacity to screen out the influence of contextual values on its content” (Ruphy 2006, p. 190).[5]

Inspired by these characterizations of (epistemic) integrity, we arrive at the following definition of the degree of epistemic integrity of the research process:

EIR1 The degree of epistemic integrity of the research process $=_{df}$ the degree to which the research process is untouched by the outside.

The underlying idea of this definition is that there are certain factors which are external to the research process, such as the financial interests of those involved in it, the social goals it is supposed to serve, etc. It is assumed that these factors should not affect the research process; this process should be directed by internal factors (e.g., epistemic goals, empirical evidence) only. Whenever external factors do have an impact on the research process, the epistemic integrity of this process is damaged according to EIR1 (in the sense that the research process does not have the highest degree of epistemic integrity).

6.2.2 Criticism

This view corresponds with the full ideal of purity, which states that scientific research should be detached from the non-epistemic interests of certain stakeholders. In part 1 of this dissertation, we have seen that non-epistemic interests, including financial interests and social interests, form an integral part of scientific research. This means that the underlying idea of EIR1 is problematic. Non-epistemic interests such as financial interests and social interests are *not* external to the research process. They influence research in a variety of ways, and their influence should be allowed, as we have seen. Since EIR1 is based on a problematic inside/outside distinction, I reject EIR1.

[5] For more on Ruphy’s view, see section 2.1.

6.3 Honesty and verifiability

6.3.1 Definition

Perhaps a more satisfying definition can be found in the literature on research integrity and the responsible conduct of research. One of the most commonly used definitions of research integrity is the one proposed by the Panel on Scientific Responsibility and the Conduct of Research (PSRCR) in 1992. It defines the integrity of the research process as “the adherence by scientists and their institutions to honest and verifiable methods in proposing, performing, evaluating, and reporting research activities” (PSRCR 1992, p. 17). If we use this idea as the basis for a definition of the degree of epistemic integrity of the research process, we get:

EIR2 The degree of epistemic integrity of the research process $=_{df}$ the degree to which scientists and their institutions adhere to honest and verifiable methods in processing, performing, evaluating, and reporting research activities.

6.3.2 Criticism

A problem for this definition is that it cannot account for certain unintentional infringements of epistemic integrity. Take, for instance, the fact that pharmaceutical companies sometimes use inadequate doses of comparator drugs in active-controlled trials (Reiss 2010, p. 432). Of course, they can do this on purpose, to make their own drug appear relatively more effective. They are then acting dishonestly and so the epistemic integrity of the research process is damaged according to EIR2. But they can also do it unintentionally, and that is where EIR2 comes into trouble.

Suppose a company wants to test its new drug in an active-controlled trial. It determines the dose of the comparator drug on the basis of an old study. Because the company does not bother to look at the recent literature on the comparator drug, it does not know that the dose suggested in the old study is now considered inadequate by all experts on the matter. As a consequence, it uses a comparator drug dose that is generally considered inadequate.

The reason why the company is being so sloppy in determining the dose of the comparator drug (in the sense that it does not consult the recent literature), is that it might benefit from holding a false belief on this matter. If it administers the comparator drug in an inadequate dose, then this increases the likelihood that its own drug turns out to be the more effective one. This could promote the sales of the company’s drug, which would serve the company's financial interests. In short: the fact that the company has an interest in getting the dose of the comparator drug wrong, causes it to be sloppy in determining this dose.

We have the intuition that there is something epistemically worrying about such behavior,[6] and an account of epistemic integrity should enable us to explain *why* it is epistemically worrying. Let us see whether EIR2 is adequate for this purpose.

According to EIR2, epistemic integrity requires honesty and verifiability. So to find out whether epistemic integrity is damaged in the case at hand, we have to check whether the company is acting dishonestly or uses unverifiable methods. Firstly, the company is not acting dishonestly. Sure, the company could be said to deceive its audience by raising the false impression that the comparator drug is administered in the dose which is generally considered most adequate, but that does not entail dishonesty. The company truly believes that the dose used is generally considered most adequate. Therefore, the deceit is unintentional. Since dishonesty requires that there is an intent to deceive (Resnik 1998, p. 54), the company is not acting dishonestly. Secondly, there is also no reason to assume that the company's methods are unverifiable. So according to EIR2, the epistemic integrity of the research process is not damaged in the case at hand.

I consider this a problem for EIR2. The account of epistemic integrity I am looking for should entail that epistemic integrity *is* damaged in the case under consideration. The company should not have ignored the recent literature on the comparator drug, and the fact that it did is an epistemic failure. An account of epistemic integrity of the research process should enable us to explain *why* such negligence is an epistemic failure. Because EIR2 does not enable us to do so, I consider it inadequate.

6.4 Deception

6.4.1 Definition

One way to deal with the 'inadequate doses' case is to consider deception in itself sufficient for epistemic integrity to be damaged, regardless of whether the deception is intended. In the 'inadequate doses' case, the pharmaceutical company raises the false

[6] Also see Reiss (2010, p. 432): he characterizes the use of inadequate doses of comparator drugs in active-controlled trials as one of the epistemic failures of biomedical research in the United States.

impression that the comparator drug is administered in the dose which is generally considered most adequate. We can think of such deception as an epistemic problem, even if the company did not have the intention to deceive. This leads us to the following definition:

EIR3 The degree of epistemic integrity of the research process $=_{df}$ the degree to which the information provided through the research process is non-deceptive.

A term in this definition that needs further clarification is the term 'non-deceptive'. Non-deceptiveness is the inverse of deceptiveness. What, then, does it mean for information to be deceptive?

Let us start with a standard definition, offered by the Oxford English Dictionary:

D1 To deceive $=_{df}$ to cause to believe what is false.

This definition has the following problem. Suppose *x* tells *y*: "I have bought a $100,000 car", which is a true statement. To make the story more impressive, *y* tells a third person *z* that "*x* has bought a $200,000 car", and *z* believes *y*. Now, if *x* had not told *y* about his expensive car, then *y* would not have mentioned the car to *z*, and so *z* would not have falsely believed that *x* has bought a $200,000 car. So *x* has indirectly caused *z* to hold a false belief. This means that *x* has deceived *z* according to D1. This is a counterintuitive implication of D1: it seems that *z* is deceived by *y*, but not by *x*.

One way to deal with this problem is to define deception in terms of the deceiver's intentions, as some others have done (see Mahon 2008). But intentions are exactly what I am trying to avoid by focusing on non-deceptiveness instead of honesty (see above). Moreover, it seems that appearances or observations, which cannot have intentions, can deceive as well. Hence, I need an alternative for D1 that does not refer to intentions. I propose the following:

D2 To deceive $=_{df}$ to provide information from which the audience can legitimately infer a false statement.

This definition avoids the aforementioned problem of D1. In the example, *z* falsely believes that *x* has bought a $200,000 car. This statement can be legitimately inferred from the information provided by *y* ("*x* has bought a $200,000 car"), but not from the information provided by *x* ("I have bought a $100,000 car"). So D2 implies that *z* is deceived by *y*, but not by *x*.

D2 allows me to specify EIR3:

EIR3’ The degree of epistemic integrity of the research process $=_{df}$ the inverse of the degree to which the audience can legitimately infer false statements from the information provided through the research process.[7]

If we apply this definition to the ‘inadequate doses’ case, we get the following. If a pharmaceutical company reports that the comparator drug in an active-controlled trial is administered in dose d, then it is legitimate for the audience to infer that d is the dose which is generally considered most adequate. The reason is that this corresponds with the normal course of affairs: the doses used in clinical trials are usually those which are generally considered most adequate. In the ‘inadequate doses’ case, there is a deviation from this normal course of affairs: the company administers the comparator drug in a dose d which is generally considered *in*adequate. So the statement that d is generally considered most adequate, is false in this particular case. Because the audience can legitimately infer a false statement from the company’s report, the epistemic integrity of the research process is damaged according to EIR3’.

6.4.2 Criticism

Unfortunately, EIR3’ is too demanding. Consider a fictional (but realistic) example. A scientist runs a large clinical trial in which part of the study population is given (the most adequate doses of) new test drug d_1, and the other part is given (the most adequate doses of) competing product d_2. The percentage of patients recovering from health condition C is significantly higher in the test group. The study leader concludes that d_1 is more effective for treating C than d_2. His audience legitimately infers from the study leader’s report that d_1 is more effective for treating C than d_2. Later, this claim is rejected; new tests suggest that the initial research results were due to coincidence, not to d_1 actually being more effective than d_2. It turns out that d_1 is in fact *less* effective for treating C than d_2. If EIR3’ is then applied to the initial research process, then the conclusion is that the epistemic integrity of that process was damaged, since the audience has legitimately inferred a false statement (i.e. that d_1 is *more* effective for treating C than d_2) from the information provided through that process.

This is an undesirable implication of EIR3’. The history of science teaches us that statements which are at a certain time accepted as true, often turn out to be false afterwards. Think, for instance, of Isaac Newton and his followers: some of their conclusions are now taken to be false. We do, however, not consider this a reason to

[7] This definition is a variant of the account of epistemic integrity presented in De Winter & Kosolosky (2013a, 2013b).

regard their research as epistemically problematic, or to question the epistemic integrity of their work. Rather, fallibility seems a normal, inevitable aspect of the scientific pursuit of knowledge. We cannot expect from research that it never leads to false conclusions. But EIR3' fosters such an expectation: it entails that the epistemic integrity of the research process is damaged whenever the audience can legitimately draw false conclusions from the information provided through that process. Therefore, I consider EIR3' too demanding.[8]

6.5 Adherence to standards

6.5.1 Definition

Maybe we should think of epistemic integrity in terms of adhering to certain standards. A definition of research integrity that is in accordance with this line of thought, has been proposed by Nicholas Steneck. He defines research integrity as "possessing and steadfastly adhering to professional standards, as outlined by professional organizations, research institutions and, when relevant, the government and public" (Steneck 2006, p. 56). Elsewhere, he defines integrity as "a measure of the degree to which researchers adhere to the rules or laws, regulations, guidelines, and commonly accepted professional codes and norms of their respective research areas" (Steneck 2002, p. 2).

Before this is translated into a definition of the degree of epistemic integrity of the research process, I should make a distinction between two different kinds of professional standards in science: epistemic standards and non-epistemic standards. Epistemic standards are standards that are justified on the basis of epistemic interests. Scientific research aims at satisfying certain epistemic interests, and it has to meet certain standards in order to accomplish that purpose. Consider for example an active-controlled trial which aims to reveal which of two drugs is most effective to treat a certain disease (= epistemic interest). In such a trial, we want the dose of the comparator drug to be determined in light of a decent literature study (rather than, say, on the basis of one old study), because we believe that research that meets this standard is more likely to provide genuine knowledge on which drug is most effective. Our interest in such knowledge justifies the standard.

[8] This criticism on EIR3' is inspired by comments from Hans Radder, Julian Reiss, and Maarten Van Dyck.

There are also standards for science that are *not* justified on the basis of epistemic interests, but only on other, non-epistemic grounds. Think for instance of standards relating to informed consent from research subjects, or standards concerning the ethical treatment of animals in experiments. Our reasons for endorsing such standards have nothing to do with the epistemic interests we want science to serve. Because such standards are not justified on epistemic grounds, I call them non-epistemic.

Violations of non-epistemic standards do not conflict with the epistemic good (i.e. what is good for the purpose of generating knowledge). Therefore, an account of epistemic integrity of the research process should not require the adherence to non-epistemic standards. So where Steneck's definition refers to professional standards in general, a definition of the degree of epistemic integrity of the research process should only refer to epistemic standards. We get:

EIR4 The degree of epistemic integrity of the research process $=_{df}$ the degree to which the research process adheres to professional epistemic standards, as outlined by professional organizations, research institutions and, when relevant, the government and public.

6.5.2 Criticism

A problem for EIR4 is that deviating from established standards can sometimes be justified. Suppose a number of pharmaceutical companies join forces to create a professional organization with the task of outlining epistemic standards for biomedical research. This organization then sets extremely demanding standards for showing that a drug which is already on the market, has a serious adverse effect. So in order to accept that such a drug has a serious adverse effect, extremely strong evidence is needed. In that way, the organization prevents lucrative drugs from being taken off the market. The fact that keeping potentially dangerous drugs on the market may cause severe suffering among a lot of patients, does not bother the organization. It justifies the extremely demanding standards by stating that research that meets these standards is more likely to deliver genuine knowledge on whether or not the drug has the adverse effect under consideration. So the relevant standards are epistemic standards.

Suppose, now, that a scientist has very good reasons to believe that drug *d* has serious adverse effect E. Although his evidence is very strong, it is, according to the standards established by the professional organization, not sufficiently strong to accept that *d* has adverse effect E. According to these standards, he should conclude that it is still insufficiently clear whether or not *d* has adverse effect E. But the scientist considers the established standards too demanding; they make it nearly impossible to show that a drug on the market has a serious adverse effect. He thinks that it is better for public health and general well-being if weaker evidence, such as the evidence he gathered, is considered sufficient to conclude that *d* has adverse effect E. Accordingly, he concludes, on the basis of his evidence, that *d* has adverse effect E.

As the scientist violates an epistemic standard that is established by a professional organization, EIR4 implies that the epistemic integrity of his research process is damaged. This is strange: if anything is questionable, then it is the professional organization's standards, rather than the scientist's research process. EIR4 cannot account for this; it takes established standards for granted. It characterizes any violation of established standards as damaging to epistemic integrity. But scientists should be free to question existing standards, and to violate them when they think they should, as the example shows. So if the concept of epistemic integrity of the research process is to function as an ideal for science (as I want it to), then it should not require strict adherence to established standards. Because EIR4 does require that, I reject it.

6.6 Intelligent adherence to one's convictions

6.6.1 Definition

An alternative way to develop an epistemological concept of the integrity of the research process, is to start from an epistemological concept of the integrity of agents. This is a virtue epistemological approach. In virtue epistemology, agents are the primary source of epistemic value and the primary focus of epistemic evaluation. Accordingly, a cognitive performance is evaluated on the basis of the properties of the cognizer (Greco & Turri 2013). Recently, Greg Scherkoske (2012) has discussed integrity as an epistemic virtue for persons.[9] Let me explore whether that discussion can form the basis for an adequate definition of the degree of epistemic integrity of the research process.

Scherkoske begins with the intuition that integrity requires that one sticks to one's convictions in the face of disagreement and challenge, and that one holds one's line in face of temptation to capitulate, 'sell-out' or backslide. In order to avoid that integrity amounts to obstinate, dogmatic adherence to whatever one happens to believe,

[9] For other accounts of integrity as an intellectual/epistemic virtue for persons, see, e.g., Baehr (2011, p. 20), Cooper (1994, p. 464), and Zagzebski (1996, p. 162). These accounts can all be situated in the top right corner of table 6.1. The reason why I focus on Scherkoske (2012) is that I consider his account of integrity the most advanced virtue epistemological account of integrity, and most promising for the purpose of this dissertation.

Scherkoske states that "integrity requires that one's convictions be responsive to relevant reasons bearing on the justifiability of those convictions" (Scherkoske 2012, p. 186). So when an agent has a good reason to change one of his convictions, then he is allowed (and maybe even required) to do so; this does not compromise the agent's integrity.

This leads Scherkoske to propose three components of integrity:

(a) The complex disposition to maintain and revise one's convictions in epistemically responsible ways. The person of integrity is committed to holding and revising her convictions on the basis of good reasons.
(b) Integrity involves a reflexive awareness of the quality of one's judgment – one has a warranted appropriate regard for one's competence to judge on a matter. In short, the person of integrity seeks to exhibit good judgment in her convictions: she would be disposed to hedge her convictions to the extent they failed by her lights to enjoy the support of good reasons; she would be disposed to strongly adhere to her convictions when they do enjoy the support of good reasons.
(c) Integrity finds full expression in its practical and social dimensions: the convictions for which persons of integrity stand will (barring forms of irrationality) be expressed in their actions and deliberation. The reasons which support a person's convictions are, at the same time, reasons for acting on those convictions – and reasons suited to offer to others when challenged, asked for advice or in conversation. (Scherkoske 2012, p. 199)

These three components can be used as the basis for a definition of the degree of epistemic integrity of the research process:

EIR5 The degree of epistemic integrity of the research process $=_{df}$ the degree to which the agents involved in the research process manifest components (a), (b), and (c) within the research process.

6.6.2 Criticism

A problem for EIR5 is that it is not very exact. To illustrate this, let me analyze the 'inadequate doses' case on the basis of component (a). In the 'inadequate doses' case, which dose of the comparator drug is most adequate is determined on the basis of one old study. The fact that the study suggests that a certain dose d is most adequate, is a good reason to believe that d is most adequate. But, on the other hand, the recent literature provides even better reasons *against* the belief that d is most adequate, and this literature is ignored by the people at the company. Does this mean that these people are not committed to holding and revising their convictions on the basis of good reasons, as component (a) requires?

EIR5 requires that the agents involved in the research process manifest a commitment to hold and revise their convictions on the basis of good reasons (see component (a)), but it is not clear what such a commitment exactly consists of. Does is it require that one is not ignorant of certain reasons against one's convictions, or is it sufficient that one's convictions are responsive to the reasons one knows of? The answer to this question is not clear, and depending on how we answer it, the verdict for the 'inadequate doses' case is different. If we assume that the agents involved in the research process should only consider the reasons they know of, then there is no integrity failure in the 'inadequate doses' case, since the people at the pharmaceutical company do not know of the reasons provided by the recent literature, and so they do not have to take them into account. If, on the other hand, we assume that the commitment to hold and revise one's convictions on the basis of good reasons also requires that one knows of certain reasons which are available in the relevant literature, then there *is* an integrity failure in the 'inadequate doses' case: because the people at the company are ignorant of the recently published reasons against their conviction that d is the most adequate dose of the comparator drug, they do not manifest the commitment to hold and revise their convictions on the basis of good reasons, and so the epistemic integrity of the research process is damaged according to EIR5.

Because the account of epistemic integrity I am looking for should entail that epistemic integrity is damaged in the 'inadequate doses' case (see section 6.3.2), the latter approach is preferable over the former. But the latter approach is problematic as well. More specifically, the problem is that we cannot require from scientists that they always know of all reasons for and against their convictions presented in the relevant literature. Suppose there are a thousand writings on a certain topic. A scientist interested in the topic faces a dilemma: either he reads all of these writings, but then he does not have the time left to perform his own research, or he only reads part of them. He takes the latter option, after which he performs his own research. As the scientist does not read all relevant writings, he remains ignorant of certain reasons against his convictions that are available in the relevant literature. But here, the ignorance seems acceptable, and requiring that the scientist avoids it would not only be too demanding, but also inappropriate, since he would then not be able to perform his own research.

In both the 'inadequate doses' case and the 'thousand writings' case, there is ignorance of certain reasons against one's convictions. But in the one case it is problematic and in the other it is not. The question is then: How much ignorance is epistemically acceptable? Or, to put it in terms of epistemic integrity: When exactly does ignorance become a threat to the epistemic integrity of the research process? EIR5 does not enable us to answer this question. This means that it does not enable us to make a clear distinction between what is epistemically problematic and what is not. Hence, EIR5 is insufficiently exact for the purpose of this dissertation.

6.7 Adherence to standards revisited

6.7.1 Definition

As all of the definitions of the degree of epistemic integrity of the research process considered so far have certain problems, let me propose a new definition, which I think avoids the disadvantages of the other accounts:

EIR6 The degree of epistemic integrity of the research process $=_{df}$ the degree to which the research process lives up to the epistemic standards that the audience can legitimately assume to be met in the research process.

This definition is, just as EIR4, based on the idea that research processes must adhere to certain epistemic standards in order to have the highest degree of epistemic integrity. But contrary to EIR4, the epistemic standards to be adhered to are not *established* epistemic standards. Instead, research processes should adhere to the epistemic standards that the audience can legitimately assume to be met. Let me explain what I mean by this.

When the results of research are reported to an audience (i.e. a group or individual to whom the research is communicated), this causes the audience to make certain assumptions about the epistemic standards that the research meets. These assumptions could be very vague or more specific. For instance, if a scientist reports that research shows that drug d_1 is more effective than drug d_2 for treating a certain health condition C, then the audience assumes that scientific research has been performed which delivers good, scientific reasons to believe that d_1 is more effective than d_2 for treating C. So the following epistemic standards are assumed to be met in the research process: (1) scientific research has been performed, and (2) this research delivers good, scientific reasons to believe that d_1 is more effective than d_2 for treating C. This is an example of a very vague assumption about the epistemic standards adhered to. But that does not imply that it could mean anything. (1) and (2) are for instance incompatible with flipping a coin to determine which of the two drugs is most effective. So if that would be what the scientist actually did, then he has violated the epistemic standards that the audience assumes to be met.

The audience can also have a more specific idea of the epistemic standards that are met. A specialized audience of biomedical scientists may for instance have a more precise view of what kind of empirical evidence would be needed to conclude that d_1 is more effective than d_2 for treating C, and it may assume that such empirical evidence has been gathered. A lay audience may, on the other hand, lack such a precise view. This indicates that one report could lead to different assumptions among different audiences.

Different assumptions about epistemic standards could also be due to the fact that the same research is often reported differently to different audiences. While only the results are communicated to laypeople, a detailed report of the different steps of the research process is usually made available to the expert community. This leads to experts making much more precise assumptions about the epistemic standards that are met in the research process.

Of all the assumptions that different audiences can make about the epistemic standards that a research process lives up to, I want to focus on those assumptions that can be made *legitimately*. Let me propose an account of what it means for such assumptions to be legitimate for a certain audience. I consider the assumption that research process *p* meets epistemic standard *s* legitimate for audience *a* if (1) those involved in the research process report or imply to *a* that *p* meets *s*, or (2) *s* is a common epistemic standard in the relevant field and those involved in the research process do not (clearly) report to *a* that *p* does not meet *s*. If one of these two conditions is fulfilled, then *s* is an epistemic standard which *a* can legitimately assume to be met in *p*.

The simplest case is when those involved in the research process *explicitly state* that the research process meets a certain epistemic standard. It is then legitimate for the audience to whom this statement is communicated to assume that the relevant epistemic standard is met, because condition (1) is fulfilled. But it is also possible that those involved in the research process do not explicitly state that the research process lives up to a certain epistemic standard, but that they *imply* it. Suppose a scientist argues in a research report that the relevant research should by all means meet a certain epistemic standard *s*. He does not add the explicit statement that *s* is actually met in the reported research since that is obvious in light of his argument, and since the report does not include any reason to think otherwise. So in this case, the scientist does not directly report that his research adheres to *s*, but he does say it in an indirect way. By endorsing *s* as an epistemic standard that his research must definitely adhere to (and by not mentioning a failure to meet *s*), he implies that *s* is met in the reported research. It is then, on the basis of condition (1), legitimate for the reader of the report to assume this.

Now consider condition (2), which concerns common epistemic standards. Common epistemic standards are here understood as epistemic standards that are generally accepted as standards which should (normally) be respected in the relevant research areas. This includes standards such as those that EIR4 refers to: professional epistemic standards, as outlined by professional organizations, research institutions and, when relevant, the government and public. It also includes common reasoning standards (which are not necessarily made explicit by a professional organization or some other relevant actor). An example is the standard that if 75% of the respondents in a survey have answered "yes" to question *q*, and 25% have answered "no" to *q*, then the conclusion should not be that the majority of the respondents have answered "no" to *q*. This is a common epistemic standard in science (regardless of whether or not it is made explicit by a professional organization or some other relevant actor).

The audience can legitimately assume that such common epistemic standards are met in a certain research process unless those involved in this process clearly communicate that this is not the case. That is the point of condition (2). Condition (2) implies that for common epistemic standards to be standards which the audience can legitimately assume to be met in a certain research process, it is not required that those involved in this process actively report or imply that these standards are met; it is sufficient that those involved in the research process do not report that these standards are violated.

As indicated above, research is often reported differently to different audiences. Consequently, it is possible that there are epistemic standards that one audience can legitimately assume to be met in the research process, while another audience cannot legitimately make that assumption. In other words, different assumptions about epistemic standards may be legitimate for different audiences. In such cases, it is sufficient that there is one audience that can legitimately assume that a certain epistemic standard is met, for this to be a standard that the research process should live up to (in order to have the highest degree of epistemic integrity). Researchers should be straight about their epistemic standards to all audiences to which they report their results. So if one audience can, on the basis of what is reported to this audience, legitimately make a false assumption about the epistemic standards that are met in the research process, then this is sufficient for the epistemic integrity of the research process to be damaged.

Finally, I should say something about the fact that epistemic integrity comes in degrees. A research process can have a higher or lower degree of epistemic integrity. This depends on the degree to which the research process lives up to the epistemic standards that the audience can legitimately assume to be met in it. If one such standard is violated, then epistemic integrity is damaged, in the sense that it is not optimal and that in principle it could be higher. But that does not mean that the research process has zero epistemic integrity. If the research process still meets plenty of other epistemic standards that the audience can legitimately assume to be met, then its epistemic integrity could still be very high.

A concrete measure of degrees of epistemic integrity is not yet available. It should be noted however that this does not mean that EIR6 is in its current form useless. In fact, I think that a concrete measure of degrees of epistemic integrity is not needed for EIR6 to be highly useful, as I hope to show in the remainder of this dissertation. More specifically, chapters 8-10 will reveal how EIR6 can lead to plenty of new insights into cases from scientific practice.

6.7.2 Dealing with the criticisms

In order to further clarify EIR6, I will now indicate how it deals with each of the criticisms on the other definitions of epistemic integrity of the research process discussed above. My criticism on EIR1 was that it starts from a problematic distinction between factors that are *external* to the research process (e.g., financial interests) and factors that are *internal* to the research process (e.g., epistemic goals, empirical evidence). EIR6 avoids this criticism because it does not presuppose such a distinction.

EIR2 was rejected because it cannot adequately deal with the ‘inadequate doses’ case. Let me analyze that case on the basis of EIR6. In the ‘inadequate doses’ case, the following epistemic standard is violated: in determining the dose of the comparator drug in an active-controlled trial, one ought to check the relevant literature, that is, not just one old study but also recent publications on the comparator drug (if available). This epistemic standard, let us call it s_1, is violated because the pharmaceutical company completely ignores the recent literature on the comparator drug. Whether this means that the epistemic integrity of the research process is damaged, depends, according to EIR6, on whether or not the audience can legitimately assume that s_1 is met; if it can, then the research process does not live up to an epistemic standard that the audience can legitimately assume to be met, and so its epistemic integrity is damaged.

In order to find out whether the audience can legitimately assume that s_1 is met, we need to examine whether one of the two conditions of legitimacy (see section 6.7.1) is fulfilled. Since s_1 is a common epistemic standard in biomedical science, the second condition of legitimacy would be fulfilled if the company would not clearly report that s_1 is violated. Let me show then that in the ‘inadequate doses’ case as I described it in section 6.3.2, the company does not report that s_1 is violated.

I described the violation of s_1 as a strategy to promote the sales of a drug. Being sloppy in determining the dose of the comparator drug increases the likelihood that the comparator drug is outperformed by the test drug, and this result could help the company in getting the test drug sold. But this strategy only works if the audience does not know that the research is sloppy. If the audience would know that the company has violated a basic epistemic standard such as s_1, then the favorable result would lose much of its persuasiveness. Accordingly, the result would lose much of its usefulness for getting the test drug sold. Therefore, part of the strategy is to keep the violation of s_1 secret.

So the company does not mention that s_1, which is a common epistemic standard in biomedical science, is violated. This means that the second condition of legitimacy (see section 6.7.1) is fulfilled. Accordingly, it is legitimate for the audience to assume that s_1 is met in the research process. Because s_1 is in fact not met, the research process does not live up to an epistemic standard that the audience can legitimately assume to be met, and so its epistemic integrity is compromised according to EIR6.

Now consider my criticism on EIR3'. I objected that EIR3' is too demanding because it implies that research which accidentally leads to false conclusions (e.g., Newton's research) is epistemically problematic. EIR6 does not have this implication. If research meets the epistemic standards that the audience can legitimately assume to be met, then its epistemic integrity is not damaged according to EIR6, even if the research leads to false conclusions.

My objection against EIR4 was that it does not respect the freedom of scientists to question established standards, and to violate these standards where they think this is appropriate. EIR6 does respect that freedom. If a scientist disagrees with an established epistemic standard, then he is allowed to violate that standard. It is then important that he clearly reports that his research does not adhere to the standard under consideration. In that way, he ensures that neither the first nor the second condition of legitimacy (see section 6.7.1) is fulfilled, so that the audience cannot legitimately assume that the established standard is met.[10] The violation of that standard does then not involve the violation of an epistemic standard that the audience can legitimately assume to be met, and so it does not involve damage to the epistemic integrity of the research process (as defined by EIR6). In a nutshell: the violation of an established standard does not damage the epistemic integrity of the research process as long as one is clear and open about this violation.[11]

---

[10] This does not imply that there are no other epistemic standards which the audience can legitimately assume to be met. The scientist may propose an alternative epistemic standard and claim that his research adheres to this standard. The audience can then legitimately assume that this alternative standard is met in the research process. Furthermore, it is possible that besides the established epistemic standard that the scientist disagrees with, there are plenty of other established epistemic standards, and it may be legitimate for the audience to assume that these other standards are still respected in the research process. In case there are no epistemic standards which the audience can legitimately assume to be met (which is very unlikely in actual scientific practice), the epistemic integrity of the research process is by definition not damaged, since damage to epistemic integrity requires the violation of an epistemic standard which the audience can legitimately assume to be met in the research process.

[11] Note that the issue here is not whether the research process is worthy of the term 'scientific'. If the research process does not live up to certain established standards, then some may claim that it is not science. I do not

Some may object that this makes my concept of epistemic integrity too liberal. Suppose a researcher believes that the most reliable way to answer yes-no questions is by flipping a coin, with 'heads' meaning 'yes' and 'tails' meaning 'no'. To him, the standard that one should answer yes-no questions in this way is an epistemic standard that research should adhere to. Suppose that he performs 'research' that meets this strange standard, and that he is clear and open about this and about the fact that his research does not meet common epistemic standards. EIR6 implies then that there is no damage to epistemic integrity. In light of this example, it seems that all possible epistemic standards must be permitted under EIR6, and so EIR6 seems to be much too liberal.

My response to this objection is twofold. Firstly, not *all* possible standards are compatible with my ideal of epistemic integrity. Take, for instance, the standard that one should fabricate data and then present these data as if they result from decent empirical research. Call this standard s_2. s_2 conflicts with the common epistemic standard s_3 that one should *not* fabricate data and then present these data as if they result from decent empirical research. If one clearly reports that the research process does not meet s_3, then it is clear that certain data are fabricated. Thus, the data are then not presented as if they result from decent empirical research, and so s_2 is violated. This shows that adhering to s_2 requires keeping the violation of s_3 secret. If the violation of s_3 is kept secret, then the audience can legitimately assume that s_3 is met, since s_3 is a common epistemic standard (which means that the second condition of legitimacy (see section 6.7.1) is fulfilled). So adhering to s_2 implies (1) that the research process does not meet the common epistemic standard s_3, and (2) that the audience can legitimately assume that the research process meets s_3. This shows that adhering to s_2 implies that the epistemic integrity of the research process is damaged. Therefore, s_2 is incompatible with my ideal of epistemic integrity.

Secondly, I admit that there are still a lot of possible standards that are compatible with this ideal, but this is a consequence of the fact that I want my concept of epistemic integrity to be independent of my personal beliefs on which standards have to be met in order to generate good, reliable knowledge. Different people have different beliefs on these matters. Even within the field of science, there are disagreements on which approaches lead to the best, most reliable knowledge (think, for instance, of disagreements on which values should be used in theory choice, and on how these values should be interpreted and weighed; see section 2.4). I want my concept of epistemic integrity to transcend such disagreements, so that arguments in which this

contest such claims. My discussion is about the epistemic integrity of the research process, not about the demarcation of science from non-science. The latter is a different issue.

concept is used are compelling to people holding different beliefs on these matters. I want my concept of epistemic integrity to be acceptable even to people who disagree with me on which particular epistemic standards research should adhere to. The fact that this concept is very liberal with respect to epistemic standards should therefore be regarded as an advantage rather than a disadvantage of this concept. The concept captures the idea that, regardless of which particular epistemic standards a research process adheres to, it is important that there is clarity about these standards to the audience, so that each audience member can decide on its own how good or reliable it considers the relevant research process and the knowledge it delivers.

Finally, consider my criticism on EIR5, which was that EIR5 is not sufficiently exact. In the next chapter, I will say something more about the exactness of EIR6 (see section 7.3). Let me here briefly show that EIR6 enables us to answer the question 'When exactly does ignorance become a threat to the epistemic integrity of the research process?', a question which cannot be answered on the basis of EIR5, as we have seen.

EIR6 suggests the following answer: ignorance is a threat to the epistemic integrity of the research process whenever the audience can legitimately assume that this ignorance has been avoided. This answer enables us to distinguish the ignorance in the 'inadequate doses' case from the ignorance in the 'thousand writings' case (a distinction that cannot be made on the basis of EIR5, see section 6.6.2). In the 'inadequate doses' case, the audience can legitimately assume that the researchers have checked the relevant literature on the comparator drug, including some recent publications, in that way avoiding complete ignorance of what has been recently published. Because the researchers have in fact not checked any recent publications, the epistemic integrity of the research process is damaged according to EIR6. Things are different in the 'thousand writings' case. The standard that one should read all writings on a research topic even if it means that one does not have any time left to perform one's own research on the topic, is not a common epistemic standard. So the second condition of legitimacy (see section 6.7.1) is not fulfilled. Consequently, if the scientist does not state or imply that he has read all writings on the relevant topic, then neither condition of legitimacy is fulfilled. The audience can then not legitimately assume that the scientist has checked all writings on the relevant topic. Accordingly, the fact that the scientist has not checked all writings on the relevant topic does not involve the violation of an epistemic standard that the audience can legitimately assume to be met. This explains why the scientist's ignorance in the 'thousand writings' case does not threaten the epistemic integrity of the research process.

Chapter 7
The four requirements of explication

Now that I have shown that my criticisms on EIR1-EIR5 do not apply to EIR6, it is time to offer a more systematic, positive defense of EIR6. More specifically, I will argue that EIR6 provides a good explication of 'the degree of epistemic integrity of the research process'. In order to do that, I start from Rudolf Carnap's characterization of the task of explication:

> If a concept is given as explicandum, the task consists in finding another concept as its explicatum which fulfils the following requirements to a sufficient degree.
> 1. The explicatum is to be similar to the explicandum in such a way that, in most cases in which the explicandum has so far been used, the explicatum can be used; however, close similarity is not required, and considerable differences are permitted.
> 2. The characterization of the explicatum, that is, the rules of its use (for instance, in the form of a definition), is to be given in an exact form, so as to introduce the explicatum into a well-connected system of scientific concepts.
> 3. The explicatum is to be a fruitful concept, that is, useful for the formulation of many universal statements (empirical laws in the case of a nonlogical concept, logical theorems in the case of a logical concept).
> 4. The explicatum should be as simple as possible; this means as simple as the more important requirements (1), (2), and (3) permit. (Carnap 1950, p. 7)

Carnap states that philosophers, scientists, and mathematicians make explications very frequently (Carnap 1950, p. 7). Following Weber, Van Bouwel & De Vreese (2013, Chap. 2), I wish to note however that the fruitfulness of explications (see third requirement) in the field of philosophy of science may not lie in their usefulness for formulating empirical generalizations (like scientific explications) and/or mathematical/logical theorems (like mathematical/logical explications). Rather, it lies in their usefulness for formulating clear guidelines for science, and for analyzing cases from scientific practice. So in order to show that the concept defined by EIR6 is a good explicatum for 'the degree of epistemic integrity of the research process', I will argue that it meets the following conditions: (1) it is similar to our common sense notion of epistemic integrity of the research process, (2) it is exact, (3) it is fruitful, that is, useful for the development and justification of norms for science and for the analysis of cases from scientific practice, and (4) it is as simple as requirements (1), (2), and (3) permit. But first, let me explain for each of these conditions why it is important that my concept of epistemic integrity meets it.

7.1 A justification of the four requirements

The first requirement, similarity to the explicandum, is important because I want my concept of epistemic integrity to be adequate for elucidating common intuitions on epistemic integrity. In cases in which it is common sense that the epistemic integrity of the research process is damaged, threatened, promoted, etc., my concept of epistemic integrity should enable us to explain why we think this is the case. For instance, while it is common sense that short-term commercial interests compromised the epistemic integrity of the research process in the Vioxx case, a decent clarification of this common intuition is absent (see section 1.3). My concept of epistemic integrity should deliver the basis for such a clarification. It can only do that if it is more or less similar to our common sense notion of epistemic integrity of the research process; if the two would be totally different, then the concept would not be helpful to deal with common intuitions on epistemic integrity. Complete similarity is, however, not necessary, as we may be prepared to give up some of our common intuitions in light of new analyses. So my concept of epistemic integrity should be similar, but not necessarily identical to our common sense notion of epistemic integrity of the research process. This corresponds with the first requirement of explication.

The second requirement of explication, exactness, is important because the application of inexact concepts is undetermined. I introduced the concept of epistemic integrity because it would enable us to identify epistemically problematic interest influences in the sciences. But if this concept would be inexact, then there would be cases in which it is not clear whether or not epistemic integrity is compromised. Accordingly, it would not always be clear whether or not an interest influence is epistemically problematic. Since my concept of epistemic integrity should provide a clear-cut criterion to determine whether or not an interest influence in science is epistemically problematic,[1] it should be as exact as possible.

The concept should also meet the third requirement of explication: fruitfulness. Two kinds of fruitfulness can be distinguished: (1) general fruitfulness, i.e. usefulness for the development and justification of general norms for science, and (2) applied fruitfulness, i.e. usefulness for the analysis of particular cases from scientific practice. It is important that general norms for science, which indicate how scientific research must be organized and carried out, are adequately formulated and justified. It is also important that we analyze and evaluate particular cases from scientific practice in such a way that we draw the right lessons from these cases. Therefore, we need ideals for science which are adequate for these purposes. My concept of epistemic integrity is supposed to function as such an ideal. Accordingly, I want this concept to have both kinds of fruitfulness.

The fourth requirement of explication, simplicity, is important mainly for practical reasons. The simpler the concept, the easier it is to explain this concept, to fully understand it, and to apply it to specific cases. As a concept becomes more complex, more effort is required to clarify its meaning and to analyze particular cases on the basis of this concept. This also makes the concept less attractive to others who might want to use it. Therefore, I prefer a concept that is as simple as the other three requirements of explication permit.

7.2 Similarity to the explicandum

In order to show that my concept of epistemic integrity fulfils the first requirement of explanation, I should argue that is (more or less) in accordance with common intuitions on epistemic integrity. In section 6.7.2, I have already analyzed several cases on the basis of this concept, showing that my concept of epistemic integrity does not have counterintuitive implications with respect to these cases. Let me here consider some additional cases. I start with two of the most obvious threats to epistemic integrity: fabrication and falsification.

[1] I characterized epistemic integrity as a property that comes in degrees. That does not mean that it cannot be a clear-cut criterion to determine whether or not an interest influence in science is epistemically problematic. The influence of an interest on the research process is epistemically problematic if the interest causes the research process to have lower epistemic integrity. If the interest does not cause the epistemic integrity of the research process to be lower, then its influence on this process is *not* epistemically problematic.

Fabrication is "making up data or results and recording or reporting them" (Office of Science and Technology Policy 2000, p. 76262). When made up data are presented as if they were obtained through empirical research, then it is legitimate for the audience to assume that the research process adheres to the following epistemic standard s_1: empirical research was performed and the data obtained through this research are reported. As the reported data were actually made up in the case of fabrication, s_1 is *not* adhered to. The epistemic integrity of the research process is then damaged according to EIR6.

But fabrication does not *necessarily* imply damage to epistemic integrity. Suppose a scientist makes up certain data in order to establish certain statistical or conceptual problems, and that he openly communicates that the data were not obtained through empirical research, but that they were made up. It is then *not* legitimate for the audience to assume that the research process adheres to s_1, as it is clear from the scientist's report that this is not the case. So here, the fact that the data were made up does *not* involve a violation of an epistemic standard which the audience can legitimately assume to be met, and so there is no reason to believe that the epistemic integrity of the research process is damaged. This analysis corresponds with common sense: fabrication only damages epistemic integrity if it is kept secret.

Now consider falsification. Falsification is "manipulating research materials, equipment, or processes, or changing or omitting data or results such that the research is not accurately represented in the research record" (Office of Science and Technology Policy 2000, p. 76262). So falsification occurs when (1) research materials, equipment, or processes are manipulated, or data or results are changed or omitted, and (2) the research report raises the impression that the research did not involve such manipulations, adjustments, or omissions. The epistemic standard that no such manipulations, adjustments, or omissions occurred is then an epistemic standard which the audience can legitimately assume to be met in the research process. Because this epistemic standard was actually *not* met, the epistemic integrity of the research process is damaged according to EIR6. This conclusion is in accordance with common sense.

In order to further support the idea that my concept of epistemic integrity is similar to our common sense notion of epistemic integrity of the research process, let me consider some earlier uses of this notion, and see whether my concept does not have counterintuitive implications in those contexts. Unfortunately, the term 'epistemic integrity' is not a term that is frequently used. It seems to be used almost exclusively by philosophers, and even in philosophy it is not ubiquitous. Moreover, where the term *is* used, it is often assumed to be a property of agents (e.g., Loder 2002; van Fraassen 1995). But with my concept I have not tried to capture the epistemic integrity of *agents*. Instead, I have developed a concept of the epistemic integrity of *behavior* (see section 6.1). The use of such a concept is relatively rare.

There are, however, some cases in which epistemic integrity is taken to be a property of behavior (rather than of agents). In McMullin (1983), we find the following passage:

> Of course, it is not pragmatic values that pose the main challenge to the epistemic integrity of the appraisal process. If values are needed in order to close the gap between underdetermined theory and the evidence brought in its support, presumably all sorts of values can slip in: political, moral, social, religious. (McMullin 1983, pp. 18-19)

This passage explicitly presents epistemic integrity as a property of the appraisal process. It suggests that political, moral, social, and religious values pose a more serious challenge to the epistemic integrity of the appraisal process than pragmatic values. McMullin does, however, not specify what he exactly means by 'epistemic integrity of the appraisal process'. What I will argue then, is that my concept of epistemic integrity of the research process is in accordance with this passage from McMullin.

The problem with political, moral, social, and religious values in research is that they are often linked to a certain research outcome, in the sense that the value leads to a preference for this outcome over other potential outcomes of research. For instance, if the impact of capitalism on world poverty is investigated, then a capitalist investigator may hope that the conclusion is that capitalism substantially reduces world poverty, while a communist investigator may hope for the opposite outcome. Anti-racist researchers may hope that their research does not reveal a correlation between skin color and the suitability for certain jobs, and religious researchers may hope that their findings do not contradict their religious beliefs.

When researchers have such a preference for a certain outcome, this can stimulate them to deviate from common epistemic standards in order to make the preferred outcome more likely. They could, for instance, systematically ignore unfavorable data. Such violations are then kept secret because exposing them would undermine the audience's trust in the results, and this would render the results useless for the relevant political, moral, social, or religious goals. By keeping the violations secret, it becomes legitimate for the audience to assume that the research adheres to common epistemic standards. Because this assumption is false, the research process does not live up to certain epistemic standards that the audience can legitimately assume to be met. Thus, according to EIR6, its epistemic integrity is damaged.

Such a danger does not arise with respect to *pragmatic* values. Pragmatic values concern the scarcity of the time and resources that are available for research (McMullin 1983, p. 18). This scarcity could in certain contexts cause researchers to lower epistemic standards. For instance, if the time and/or resources needed to collect the usual amount of evidence are not available, researchers may consider a lower amount of evidence sufficient to draw a certain conclusion. But there is an important difference between this situation and the above situations. In the above situations, a common epistemic standard is violated in order to make a preferred research outcome more likely, and this violation is kept secret in order not to undermine the audience's trust in the outcome. When, on the other hand, one deviates from a common epistemic standard for pragmatic reasons, this is not to make a preferred research outcome more likely, and there is no reason why one should cause the audience to overestimate the trustworthiness of the outcome. Hence, if researchers lower epistemic standards for pragmatic reasons, they can be clear and open about this, in that way protecting the epistemic integrity of the research process (as defined by EIR6). (Epistemic integrity is protected because if the researchers are clear and open about the fact that a certain common epistemic standard is not met, then the audience cannot legitimately assume that it is, which means that deviating from this standard does not involve the violation of an epistemic standard that the audience can legitimately assume to be met.)

This brief analysis suggests that political, moral, social, and religious values pose a more serious challenge to the epistemic integrity of the research process than pragmatic values, and this corresponds with the above passage from McMullin. The conclusion is that my concept of epistemic integrity is in accordance with McMullin's intuitions on epistemic integrity.

My concept of epistemic integrity should also be in accordance with Biddle's (2007) use of the term; it should enable us to clarify why short-term commercial interests compromised epistemic integrity in the Vioxx case, as Biddle claims in his abstract (Biddle 2007, p. 21). Let me show that my concept of epistemic integrity is adequate for this purpose.

Biddle mentions two reasons for believing that Merck's research on Vioxx was epistemically irresponsible:

> The first is that, between 2000 (or even earlier) and 2004, Merck consistently mischaracterized the current state of knowledge regarding the possible cardiovascular side effects of Vioxx. The hypothesis that Vioxx did not possess dangerous cardiovascular side effects was, until 2004, uncertain, and Merck *knew* that it was uncertain. Despite this, the company consistently and publicly asserted that Vioxx was safe. Secondly, there are serious questions about whether Merck honestly reported data that was unfavorable to its financial interests. Since the withdrawal of the drug, information has come to light suggesting that Merck scientists did not report all of the cardiovascular events that they should have, thus calling into question at least two of the studies that reflected favorably upon Vioxx. (Biddle 2007, p. 27)

My concept of epistemic integrity can be used to explain why these two practices are epistemically problematic.

Firstly, consider the mischaracterization of the contemporary state of knowledge regarding the possible cardiovascular side effects of Vioxx. For instance, on May 22, 2001, Merck published a press release entitled 'Merck confirms favorable cardiovascular safety profile of Vioxx®'[2] (Biddle 2007, p. 26). The press release mentioned that in the VIGOR study (Vioxx Gastrointestinal Outcomes Research), "significantly fewer heart attacks were observed in patients taking naproxen (0.1 percent) compared to the group taking Vioxx 50 mg (0.5 percent) in this study." This finding suggests that Vioxx has dangerous cardiovascular side effects. Merck scientists, however, explained the finding by "naproxen's ability to block platelet aggregation by inhibiting COX-1 like aspirin, which is used to prevent second cardiac events in patients with a history of heart attack, stroke or other cardiac events."[3] So according to Merck scientists, the VIGOR finding was not due to a harmful effect of Vioxx, but to a cardioprotective effect of naproxen.

It was then legitimate for the audience to assume that Merck scientists had decently investigated this explanation, and that there was evidence which clearly demonstrated that the higher incidence of heart attacks among patients taking Vioxx in the VIGOR study was due to a cardioprotective effect of naproxen, and not to a harmful effect of Vioxx. In the absence of such evidence, it would have been premature for Merck to publicly state that it confirmed the favorable cardiovascular safety profile of Vioxx®; first, Merck should have collected the evidence necessary to rule out the possibility that the VIGOR finding was due to a harmful effect of Vioxx.

It seems, however, that this standard was not met; Merck scientists had not decently investigated whether a cardioprotective effect of naproxen could (fully) explain the VIGOR finding. They should at least have checked whether naproxen actually has a cardioprotective effect, and whether this effect is large enough to account for the VIGOR finding. If they would have checked that, they would have discovered that the size of the cardioprotective effect of naproxen is in fact insufficient to fully account for the difference in observed heart attacks in the VIGOR study (see Ray et al. 2002).

Because Merck scientists had not put their speculative explanation of this difference to the test (while at the same time presenting their conclusions as confirmed by scientific evidence), they violated an epistemic standard that the audience could legitimately assume to be met. This means that, according to EIR6, the epistemic integrity of Merck's research process was compromised.

Now consider the second epistemically problematic practice identified by Biddle: inadequate reporting of unfavorable data. He mentions a paper that was published in the *New England Journal of Medicine* on the VIGOR study. The paper reports that, in the

[2] See http://dida.library.ucsf.edu/pdf/oxx12o10.

[3] See press release, http://dida.library.ucsf.edu/pdf/oxx12o10.

VIGOR trial, 17 heart attacks occurred in the Vioxx group. But actually, there were 20 heart attacks in this group, and at least two of the authors were aware of this before the paper was published (Biddle 2007, p. 28). In other words, if common standards for identifying heart attack would have been used in the study under consideration, then 20 heart attacks would have been identified in the Vioxx group.

A similar problem occurred in another study, called ‘Advantage’. In a paper published in *Annals of Medicine*, Merck scientists reported that in the Advantage trial, five patients from the Vioxx group suffered heart attack or sudden cardiac death. But if common standards for identifying heart attack and sudden cardiac death would have been used in this study, then six or more cases would have been identified in the Vioxx group (Biddle 2007, p. 29).

So in both studies, common standards for identifying cardiovascular problems were not respected. Since the use of alternative standards was in neither case communicated to the audience, the audience could legitimately assume that common standards for identifying cardiovascular problems *were* respected in the two studies. Thus, epistemic standards that the audience could legitimately assume to be met were in fact violated in both the VIGOR study and the Advantage study. This means that the epistemic integrity of the research process (as defined by EIR6) was compromised in both cases.

The different ways in which the epistemic integrity of the research process was compromised in the Vioxx case, can be explained by Merck’s short-term commercial interests. By presenting the conclusion that Vioxx did not have dangerous cardiovascular side effects as confirmed by scientific evidence (while in fact the necessary evidence was not yet available), and by secretly violating common standards for identifying cardiovascular problems (in order to reduce the number of heart attacks or sudden cardiac deaths identified among patients taking Vioxx), Merck made Vioxx appear safer than it actually was. The latter promoted the sale of Vioxx in the short term. This may be the reason why the epistemically problematic practices occurred; Merck’s interest to promote the sale of Vioxx in the short term may have caused these practices. If we accept this explanation, as I think Biddle would, then the conclusion is that Merck’s short-term commercial interests compromised the epistemic integrity of the research process (as defined by EIR6) in the Vioxx case. This corresponds with Biddle’s claim that “the research was organized in such a way as to allow short-term commercial interests to compromise epistemic integrity” (Biddle 2007, p. 21). So the concept of epistemic integrity defined by EIR6 enables us to account for Biddle’s intuitions on epistemic integrity.

Since my concept of epistemic integrity of the research process is in accordance with many common intuitions on epistemic integrity, I conclude that it is more or less similar to our common sense notion of epistemic integrity of the research process. Thus, my concept of epistemic integrity fulfils the first requirement of explication.

7.3 Exactness

In the previous chapter, I have tried to make my concept of epistemic integrity of the research process as exact as possible. Ambiguities may remain, but I did my best to reduce such ambiguities as much as I could in this dissertation. I defined the degree of epistemic integrity of the research process as the degree to which the research process lives up to the epistemic standards that the audience can legitimately assume to be met in it. I have described what I mean by a research process in section 6.1. My notion of epistemic standards is clarified in section 6.5.1. Two sufficient conditions for an epistemic standard to be an epistemic standard that the audience can legitimately assume to be met, are presented and explained in section 6.7.1. Section 6.7.1 also indicates what is meant by an audience, and how the definition should be applied in case there are multiple audiences, for whom different assumptions about epistemic standards are legitimate. I hope that these clarifications suffice for my concept of epistemic integrity to have an acceptable level of exactness.

7.4 Fruitfulness

As indicated in section 7.1, two kinds of fruitfulness can be distinguished: (1) general fruitfulness, i.e. usefulness for the development and justification of general norms for science, and (2) applied fruitfulness, i.e. usefulness for the analysis of particular cases from scientific practice. In chapters 8-10, my concept of epistemic integrity will be applied to particular cases from scientific practice, showing that this concept has applied fruitfulness. In the current section, the focus is on general fruitfulness.

For my concept of epistemic integrity to be useful for the development and justification of general norms for science, having the highest degree of epistemic integrity should be something that we can regard as ideal for scientific research processes. So let me start by arguing that having the highest degree of epistemic integrity, as I defined it, is an acceptable ideal for science.

If a research process does not live up to an epistemic standard that the audience can legitimately assume to be met in the research process, then a discrepancy arises between the epistemic standards that are actually met in the research process and those that the audience believes to be met in the research process. This implies that the audience has a distorted view on the justification of the research conclusions; it believes that these conclusions are supported by research meeting certain specific epistemic standards, while in fact it is not supported by such research (though it may be supported by another kind of research, in which other epistemic standards are used). Because it is undesirable that scientific agents give the audience a distorted view on the justification of their research conclusions, scientific agents should try to make sure that there are no epistemic standards that (1) the audience can legitimately assume to be met in the research process, and (2) are actually violated in the research process. In other words, scientific agents should aim for the highest degree of epistemic integrity (as defined by EIR6).

Furthermore, when a research process does not live up to the epistemic standards that the audience can legitimately assume to be met, it is always possible that someone discovers this and makes it public. This can undermine public trust in science. If people find out that certain epistemic standards, which the audience could legitimately assume to be met in the research process, are actually violated in the research process, then this can damage these people's trust in the relevant research, and possibly even in science as a whole. This is an additional reason for scientific agents to ensure that the epistemic integrity of the research process remains untouched.

These arguments indicate that having the highest degree of epistemic integrity (as defined by EIR6) can function as an ideal for science. Now, I can turn to the question whether it is a *generally fruitful* ideal; Does it enable us to formulate and justify general norms for science? As such norms have already been developed in earlier work (e.g., Committee on Assessing Integrity in Research Environments 2002; Resnik 1998; Shamoo & Resnik 2009; Shrader-Frechette 1994), I will not start from scratch here. I start from the principles for science developed in Resnik (1998). David Resnik proposes the following principles:

> *Honesty:* Scientists should not fabricate, falsify, or misrepresent data or results. They should be objective, unbiased, and truthful in all aspects of the research process.
> *Carefulness:* Scientists should avoid errors in research, especially in presenting results. They should minimize experimental, methodological, and human errors and avoid self-deception, bias, and conflicts of interest.
> *Openness:* Scientists should share data, results, methods, ideas, techniques, and tools. They should allow other scientists to review their work and be open to criticism and new ideas.
> *Freedom:* Scientists should be free to conduct research on any problem or hypothesis. They should be allowed to pursue new ideas and criticize old ones.
> *Credit:* Credit should be given where credit is due but not where it is not due.
> *Education:* Scientists should educate prospective scientists and insure that they learn how to conduct good science. Scientists should educate and inform the public about science.
> *Social responsibility:* Scientists should avoid causing harms to society and they should attempt to produce social benefits. Scientists should be responsible for the

> consequences of their research and they should inform the public about those consequences.
> *Legality:* In the process of research, scientists should obey the laws pertaining to their work.
> *Opportunity:* Scientists should not be unfairly denied the opportunity to use scientific resources or advance in the scientific profession.
> *Mutual respect:* Scientists should treat colleagues with respect.
> *Efficiency:* Scientists should use resources efficiently.
> *Respect for subjects:* Scientists should not violate rights or dignity when using human subjects in experiments. Scientists should treat non-human, animal subjects with appropriate respect and care when using them in experiments. (Resnik 1998, Chap. 4)

What I will try to show then is that the concept of epistemic integrity is useful to justify and refine some of these principles.

Resnik uses different grounds to justify his principles. A first is that violating the principle hinders the achievement of the goals of science. Resnik distinguishes two kinds of goals: epistemic and practical. The epistemic goals are related to the advancement of human knowledge, and the practical goals concern solving problems in engineering, medicine, economics, agriculture, and other areas of applied research (Resnik 1998, p. 39). A second kind of justification is that adhering to the principle promotes cooperation and trust in science, and a third is that it promotes the public's trust in science. Fourthly, most principles are also justified on moral grounds. I will here only focus on Resnik's epistemological justification of principles, that is, the justification of a principle on the ground that adhering to it is important for the advancement of human knowledge.

The problem with this kind of justification is that scientists should not always do what is best for the advancement of human knowledge. The research that is most useful for this purpose can have a very high cost, either in terms of the resources needed for the research or in terms of the harmful effects on the research subjects. Due to this cost, it may be better for scientists to pursue an alternative line of research, which makes a smaller contribution to the advancement of human knowledge, but which has a lower cost. So even though the former research line could be regarded as important for the advancement of human knowledge, it should not be pursued. This indicates that it is not because a certain behavior or practice is important for the advancement of human knowledge, that we should accept it as something that scientists should engage in. Accordingly, the assumption that adhering to a certain principle is important for the advancement of human knowledge, is not a very compelling reason to endorse this principle.[4]

On the basis of the ideal of epistemic integrity, principles for science can be supported in a more compelling way. This ideal is much less demanding than the ideal of contributing as much as possible to the advancement of human knowledge. The ideal of epistemic integrity allows scientists to pursue research that makes a relatively small contribution to the advancement of human knowledge. The only thing that is required is that the research lives up to those epistemic standards that the audience can legitimately assume to be met. This is a very minimal requirement, and I cannot think of any cases in science in which it is not desirable that this minimal requirement is fulfilled. Ideally, the requirement is always fulfilled in science, so that all research processes in science have the highest degree of epistemic integrity. Accordingly, if adhering to a certain principle can be shown to be important for epistemic integrity, in the sense that violating the principle threatens the epistemic integrity of the research process, then this is a very compelling reason to endorse the principle.

Let me now show how some of Resnik's principles can be justified and refined on the basis of the concept of epistemic integrity. Three kinds of principles can be distinguished: those that can be completely justified on the basis of epistemic integrity, those that can only partially be justified on the basis of epistemic integrity, and those that cannot be justified on the basis of epistemic integrity at all. The principles belonging to the latter category are social responsibility, legality, efficiency, and respect for subjects. Respecting these principles is important for moral reasons, not because respecting them promotes the epistemic integrity of scientific research processes.

The principles that can be completely justified on the basis of the concept of epistemic integrity are honesty, carefulness, openness, and freedom. Firstly, take the principle of honesty: scientists should not fabricate, falsify, or misrepresent data or results, and they should be objective, unbiased, and truthful in all aspects of the research process. When scientists present certain data or results as obtained through scientific research, then it is common for the audience to assume that these data or results are not fabricated or falsified, and that these data or results are correctly represented, that is, in accordance with common standards for representing data and results. If the scientists do not explicitly state that this assumption is false, then this is a legitimate assumption for the audience to make. So in such contexts, the standards 'data and results should not be fabricated', 'data and results should not be falsified', and 'data and results should be represented correctly' are standards that the audience can

[4] I am not contesting that this assumption can be a good reason to endorse a principle; I only claim that it is not a very compelling reason.

legitimately assume to be met in the research process. They are epistemic standards because adhering to them is thought to be crucial for the research to serve the epistemic interests that the research is supposed to serve. This means that if data or results are fabricated, falsified, or misrepresented in such contexts, then the research process does not live up to the epistemic standards that the audience can legitimately assume to be met, and so its epistemic integrity is damaged. Because we want research processes to have the highest degree of epistemic integrity, data and results should not be fabricated, falsified, or misrepresented in such contexts.

There are, however, also contexts in which fabrication does not damage epistemic integrity, as we have seen. If a scientist fabricates data in order to establish certain statistical or conceptual problems, and he is clear about the fact that the data are fabricated, then the epistemic integrity of the research process is not damaged (see section 7.2). Note that the idea that fabrication is allowed in such contexts, conflicts with the principle of honesty as described by Resnik, which states that scientists should not fabricate data or results. Therefore, that principle needs to be refined. I suggest the following adjustment: scientists should not fabricate data or results *and then present these data or results as the outcome of (ordinary) scientific research.*

The concept of epistemic integrity can also be used to explain why it is important that scientists are objective and unbiased in different aspects of the research process. To illustrate this, consider decisions on whether or not to publish certain results. In biomedical science, such decisions are sometimes biased; results that are unfavorable to a certain drug (e.g., results which indicate that the drug is ineffective or not safe) are not published, while favorable results *are* published, sometimes even more than once (Melander et al. 2003; Schott et al. 2010, pp. 296-298). This damages the epistemic integrity of the research process. Let me explain.

Publication bias implies that the following standard is violated in the body of research on the relevant drug: when a clinical trial is finished, its results should be published, even if they are unfavorable to the drug. Call this standard s_2. s_2 is an epistemic standard because we believe that general adherence to s_2 serves our epistemic interests; we want to know whether the drug is effective or safe (= epistemic interest), and the publication of all clinical trial results concerning the effectiveness/safety of the drug (and not just the favorable ones) is considered helpful for satisfying that epistemic interest. More specifically, s_2 is an epistemic standard *that the audience can legitimately assume to be met in the relevant body of research*. Let me show this on the basis of the second condition of legitimacy (see section 6.7.1).

Violations of s_2 are generally considered scientific malpractice (see, e.g., Chalmers 1990; Schott et al. 2010). Today, there are even laws which require that the results of clinical trials are made publicly available.[5] So s_2 is a common epistemic standard in biomedical science. Accordingly, when s_2 is violated in a certain research process, those involved will not flaunt this violation, since they do not want to risk being accused of scientific malpractice. Moreover, if a pharmaceutical company publishes favorable results and suppresses unfavorable results in order to make one of its products appear more effective and/or safer than it actually is, then it is crucial that the audience does not know about this strategy. If the audience would know that unfavorable results are suppressed, then it would know that the published results say little about the actual effectiveness/safety of the relevant product, and so the company's strategy would fail. This is a second reason why violations of s_2 are kept secret by those involved in the research process.

So in the case of publication bias, s_2 is a standard which the audience can legitimately assume to be met in the body of research on the relevant drug, as the second condition of legitimacy is fulfilled: s_2 is a common epistemic standard in the relevant field and violations of s_2 are not clearly reported to the audience by those involved in the research process. Because publication bias implies that s_2 is violated, it involves the violation of an epistemic standard that the audience can legitimately assume to be met in the research process. This means that publication bias compromises the epistemic integrity of the research process. Therefore, publication bias ought to be avoided in biomedical science.

Now consider the principle of carefulness: scientists should avoid errors in research, especially in presenting results, they should minimize experimental, methodological, and human errors, and avoid self-deception, bias, and conflicts of interest. Some errors are in themselves sufficient for the epistemic integrity of the research process to be compromised. Take the accidental use of a statistical method which is generally considered inappropriate for the relevant area of research (= methodological error). A common standard in science is that appropriate statistical methods should be used, that is, statistical methods which are generally considered appropriate for the relevant area of research. This standard, let us call it s_3, is an epistemic standard because we believe that adherence to s_3 serves our epistemic interests. When scientists decide to violate s_3 (e.g., because they disagree with generally accepted standards on which statistical methods should be used where), they should be clear about this. But if scientists *erroneously* violate s_3, then they are not aware of it, and so they cannot openly communicate this violation to the audience. Thus, in case of an erroneous violation of s_3, it is legitimate for the audience to assume that s_3 is met, since s_3 is a common epistemic

[5] For Europe, see Official Journal of the European Union, 2012/C 302/03. For the U.S., see Section 801 of the Food and Drug Administration Amendments Act.

standard, and those involved in the research process do not clearly report that s_3 is violated (which means that the second condition of legitimacy (see section 6.7.1) is fulfilled). Because s_3 is in fact violated, the epistemic integrity of the research process is damaged.

There are also errors that do not imply that epistemic integrity is damaged. For instance, when a great amount of data are recorded by human observers, then a certain amount of erroneous recordings (= human error) is acceptable. The reason is that the audience cannot legitimately assume that not one error was made. In fact, it is common sense that human observers do make errors. What is required in such cases is that the research is characterized by a certain level of carefulness. This need not be the highest level of carefulness; e.g., each recording should not be double-checked by ten different observers. Research should be characterized by the level of carefulness *that the audience can legitimately assume to be manifested in the research*. For instance, it is usually legitimate for the audience to assume that scientific observers have made a well-considered judgment on how to record the data, instead of, say, a hasty judgment on the basis of rough estimations. It is then important that the research adheres to this epistemic standard. Otherwise, an epistemic standard that the audience can legitimately assume to be met in the research process is violated, which means that the epistemic integrity of the research process is damaged.

There may, however, also be contexts in which observers are bound to make hasty judgments, e.g., because of stringent time constraints. It is then mandatory that this is clearly communicated to the audience, so that the audience cannot legitimately assume that the observers' judgments are as well-considered as they are in other contexts. In that way, the epistemic integrity of the research process is protected.

The concept of epistemic integrity can also be used to explain why it is desirable that scientists avoid conflicts of interest. When scientists have conflicts of interest, this may have a negative impact on the quality of their research, and it becomes more likely that their research does not manifest the minimal level of carefulness that people expect to be manifested in scientific research. The 'inadequate doses' case (see section 6.3.2) can serve as an example here. In the 'inadequate doses' case, a conflict of interest leads to sloppiness in determining the dose of the comparator drug in an active-controlled trial. I have argued that this sloppiness involves the violation of an epistemic standard that the audience can legitimately assume to be met, and so the epistemic integrity of the research process is damaged (see section 6.7.2). This example shows that a conflict of interest can lead to compromised epistemic integrity. Accordingly, avoiding conflicts of interest is useful to protect the epistemic integrity of the research process.

The third principle that can be completely justified on the basis of the concept of epistemic integrity is openness: scientists should share data, results, methods, ideas, techniques, and tools, and they should allow other scientists to review their work and be open to criticism and new ideas. Openness is important because it helps the audience obtain a correct picture of the research process. If all aspects of the research process are shared with the audience, then it becomes less likely that the assumptions that audience members make about the epistemic standards that the research process lives up to, and which they can *legitimately* make on the basis of what is reported to them, are false. Put differently: if scientists are open about all aspects of the research process, then the research process is more likely to live up to the epistemic standards that the audience can legitimately assume to be met in it. This means that openness promotes the epistemic integrity of the research process.

Furthermore, where an epistemic standard that the audience can legitimately assume to be met in the research process is *not* adhered to, openness makes it more likely that this violation gets detected by an outsider. Suppose for example that in determining the dose of the comparator drug in an active-controlled clinical trial, one ignores the recent publications on the drug, resulting in the use of a dose which is considered inadequate by all experts on the matter. This violation is more likely to get discovered by an outsider if the doses used in the trial are shared with outsiders.

Making it easier for outsiders to discover cases of compromised epistemic integrity is valuable for two reasons. Firstly, such discoveries can help those involved to restore epistemic integrity (e.g., by explicitly communicating that the research does not meet the relevant epistemic standard, or by adjusting the research so that it does meet this standard). Secondly, the more likely it is that compromised epistemic integrity will be detected by outsiders, the more important it is for scientists to make sure that the epistemic integrity of their research is not compromised, since such detections will damage their reputation. So by helping outsiders to detect damage to epistemic integrity, openness promotes the epistemic integrity of the research process.

The concept of epistemic integrity can also be used to *refine* the principle of openness. As a specification to this principle, we can add that in articles in academic journals that require disclosure of conflicts of interest, the authors should disclose their financial interests, such as those that are due to industry ties (also see Schachman 2006, p. 6901); certain other interests, such as the interest to publish in high-impact journals in order to advance one's career, should not be mentioned. The reason is that by not disclosing financial ties to industry in such articles, the authors imply that such ties were absent. The audience can then legitimately assume that the reported conclusions result from industry-independent research, so that financial ties to industry could not have distorted the research. Consequently, they may put more trust in the research, considering it more appropriate for serving their epistemic interests than when there would be financial ties to industry. Because actually the research is not industry-independent, an epistemic standard that the audience can legitimately assume to be met is violated in the research process. This reveals that disclosing financial ties to industry is in such cases crucial to protect the epistemic integrity of the research process.

Things are different with respect to interests such as the interest to publish in high-impact journals in order to advance one's career. By not disclosing such interests, the authors do not imply that they were absent; on the contrary, the presence of such interests is taken for granted. The interest to publish in high-impact journals could distort research just as much as financial ties to industry, but the difference is that by not disclosing the former interest, one does not make it legitimate for the audience to assume that it wasn't there. So even if interests such as the interest to publish in high-impact journals are not disclosed in a declaration of conflicts of interest, the assumption that such interests were absent is not an assumption that the audience can legitimately make. Therefore, including taken-for-granted career interests in a declaration of conflicts of interest is not required to protect the epistemic integrity of the research process.

Furthermore, the concept of epistemic integrity enables us to refine the norm that scientists should be open to new ideas (which is part of the principle of openness). Scientists should not be open to *all* ideas, but only to those ideas *which the audience can legitimately assume to be taken seriously*. The latter depends on the values that the audience can legitimately assume to be used in the relevant research process. Let me give an example.

Simplicity is a common value in physics (Maxwell 1974, 2010). Therefore, physicists do not have to take into account overly complex theories or hypotheses (unless they explicitly mention that they do). Suppose, for instance, that physicists have developed a model *m* that describes the motion patterns of different planets, and that this model is consistent with a huge amount of empirical data (e.g., on the positions of these planets at different points in time). We can now easily come up with endless aberrant versions of *m* that are at least as empirically adequate as *m*; we just have to add absurd qualifications that have not been refuted yet. We could, for instance, say that planets move as described by *m*, but when an Asian boy with the Statue of Liberty tattooed on his forehead yells "God bless Europe" on top of Table Mountain, the planets stop moving and explode after 5 seconds. This is an aberrant version of *m*. The point is that physicists are allowed to ignore such aberrant hypotheses because, in light of common standards in physics, the audience cannot legitimately assume that they were taken into account. Therefore, ignoring such hypotheses does not compromise the epistemic integrity of the research process.

If, on the other hand, a certain new hypothesis *h* on a phenomenon *p* in physics has the potential to score better on common values in physics (e.g., empirical adequacy, internal consistency, simplicity) than competing hypotheses on *p*, then physicists should either be open to *h*, or they should clearly report that they are not (e.g., because alternative values are used in the particular research process). The reason is that a commitment to common values in physics, which the audience can legitimately assume to be present in a research context in physics unless mentioned otherwise, involves being open to new hypotheses which have the potential to score better on these values. Put differently: if those involved in research on *p* in the domain of physics do not make clear to the audience that hypotheses like *h* are ignored in this research, then the epistemic standard that they should be open to *h* is an epistemic standard which the audience can legitimately assume to be met in the relevant research process. So in such cases, the ideal of epistemic integrity requires the scientists involved to be open to *h*.

Fourthly, consider the principle of freedom: scientists should be free to conduct research on any problem or hypothesis, and they should be allowed to pursue new ideas and criticize old ones. Strict adherence to this principle is, I think, not required in science. For instance, if one can show that a certain line of research is likely to have very harmful effects on society, then this can be a good reason to prohibit such research. There are, however, cases in which the suppression of certain lines of research or ideas compromises the epistemic integrity of the research process, and it is such cases that ought to be avoided.

Suppose there are a couple of rival theories on a certain phenomenon. It would serve the government's interests if the public would believe that scientific evidence unambiguously shows that one particular theory on this phenomenon is the best one. In order to produce that belief among the public, the government forbids scientists (1) to criticize the favored theory, and (2) to support one of its rivals. Scientists are only allowed to support the favored theory. The result is that all the scientists' arguments point in the direction of the favored theory.

Part of the government's strategy is to make sure that the public does not know about the restrictions on the scientists' freedom. Because it wants the audience to trust the scientists' judgment, it gives the audience the impression that the scientists were free to develop criticisms on the favored theory, as well as to develop arguments in favor of its rivals. It is then legitimate for the audience to assume that the research process meets the following epistemic standard: during the process, scientists should be allowed to criticize the favored theory as well as to support its rivals. Because this standard is not met, the epistemic integrity of the research process is compromised. The ideal of epistemic integrity implies that cases such as this one ought to be avoided.

There are also some principles that can only partially be justified on the basis of the concept of epistemic integrity: credit, education, opportunity, and mutual respect. Respecting these principles is important regardless of whether or not violating them compromises epistemic integrity. However, some behaviors that violate them compromise epistemic integrity, and this is a reason not to practice such behaviors.

Take the principle of credit: credit should be given where credit is due but not where it is not due. The main reason why this standard should be respected is that all people, including scientists, should be given just rewards for their contributions and efforts (Resnik 1998, p. 62). But some aspects of the principle of credit can also be justified on the basis of the concept of epistemic integrity. An example is the use of guest authors and ghostwriters in biomedical science. In biomedical science, persons who did not make a significant contribution to a paper – typically academically affiliated investigators that are not employed by the company that financed the research – are listed as authors (guest authors) (Ross et al. 2008), while company employees who did make a significant contribution are not (ghostwriters) (Gøtzsche et al. 2007). The audience can then legitimately assume that the research was performed more or less independently of the relevant company, so that possible biases linked to the company's financial interests were avoided in the research. Because this epistemic standard was actually violated – the research was not performed independently of the relevant company – the epistemic integrity of the research process is compromised. Therefore, this practice ought to be avoided.

A second principle that can be partly justified on the basis of the concept of epistemic integrity is the principle of education: scientists should educate prospective scientists and ensure that they learn how to conduct good science, and they should educate and inform the public about science. If we want scientific research to maintain a high degree of epistemic integrity in the future, it is crucial that prospective scientists learn how to adhere to epistemic standards that the audience can legitimately assume to be met in the research process. For instance, prospective scientists in fields in which statistics are used, should learn which statistical methods are appropriate for which purposes. In that way, they will be able to avoid methodological errors, which could, as we have seen, damage epistemic integrity. If scientists would not educate prospective scientists in this way, the epistemic integrity of future scientific research is threatened. Scientists should also educate and inform the public about how science works. In that way, it becomes less likely that the public can legitimately make false assumptions on the epistemic standards used in scientific research processes. Accordingly, scientific research processes will be more likely to live up to the epistemic standards that the audience can legitimately assume to be met in these processes, and so their epistemic integrity is promoted. This justification of the principle of education is, however, only partial, since scientists should teach prospective scientists more than just how to respect the epistemic standards that the audience can legitimately assume to be met.

Thirdly, the ideal of epistemic integrity partly justifies the principle that scientists should not be unfairly denied the opportunity to use scientific resources or advance in the scientific profession (principle of opportunity). More specifically, the following violation of this principle compromises epistemic integrity. Suppose adherents of a certain theory are systematically denied the opportunity to advance in the scientific profession (e.g., because the government doesn't like that theory), and that this is kept secret to the public. Consequently, all scientists reject the theory, and the audience legitimately assumes that this unanimous rejection is due to a fair debate between proponents and opponents of the theory. Because this epistemic standard is actually not met (there has not been a fair debate), the epistemic integrity of the research process is compromised. Therefore, such violations of the principle of opportunity should be avoided. This is only a *partial* justification of the principle of opportunity because, even in cases in which unfairly denying someone the opportunity to use scientific resources or advance in the scientific profession (e.g., on the basis of his skin color) can be shown not to threaten the epistemic integrity of the research process in any way, it is still unacceptable (on moral grounds).

A final principle is the principle of mutual respect, which implies that "scientists should not harm one another, either physically or psychologically, that they should respect personal privacy, that they should not tamper with each other's experiments or results, and so forth" (Resnik 1998, p. 67). If there would not be mutual respect between scientists, scientists may be less inclined to honestly and openly communicate the epistemic standards that their research meets to other scientists. They would probably care less about the benefits of their research to other scientists, and they may put less effort in making sure that their research reports are as clear as possible to colleagues who might benefit from it. This could lead to more discrepancies between (1) the epistemic standards which are met in research, and (2) the epistemic standards which other scientists can legitimately assume to be met on the basis of what is reported to them. Because such discrepancies imply that the epistemic integrity of the research process is damaged, we can conclude that a lack of mutual respect between scientists could have a negative effect on the epistemic integrity of scientific research processes. Of course, even in cases in which one scientist can harm another without threatening epistemic integrity, scientists should not do so (on moral grounds). Therefore, this justification of the principle of mutual respect is only partial.

7.5 Simplicity

In order to show that my concept of epistemic integrity is as simple as the three other requirements of explication permit, I have to argue that no parts of EIR6 can be omitted without violating the other requirements. More specifically, I will argue that if certain parts are omitted, then the explicatum will be less exact or less similar to the explicandum.

Suppose we omit 'that the audience can legitimately assume to be met in the research process' from EIR6. We then have:

EIR7 The degree of epistemic integrity of the research process $=_{df}$ the degree to which the research process lives up to epistemic standards.

But there are plenty of possible epistemic standards, and not *all* possible epistemic standards have to be met in *each* research process. Different epistemic standards apply to different fields of science and to different research contexts. Therefore, it has to be specified *which* epistemic standards the research process has to live up to in order to have the highest degree of epistemic integrity. Since EIR7 does not do that, it is not sufficiently exact.

EIR6 does specify which epistemic standards the research process has to live up to in order to have the highest degree of epistemic integrity: those epistemic standards *that the audience can legitimately assume to be met in the research process*. But what if we would only omit 'can legitimately' from EIR6? We then have:

EIR8 The degree of epistemic integrity of the research process $=_{df}$ the degree to which the research process lives up to the epistemic standards that the audience assumes to be met in the research process.

This concept is less similar to our common sense notion of epistemic integrity of the research process than the concept defined by EIR6. This can be argued on the basis of the following scenario. Suppose a scientist thinks that a certain common epistemic standard is inappropriate for *his* research. Consequently, he deviates from this traditional standard. In order to avoid misunderstandings, he reports this to his audience. He clearly states that his research does not meet the traditional standard, and that he used an alternative standard instead. But the audience ignores this clear message and sticks to the assumption that the traditional standard is met in the scientist's research. EIR8 implies that the epistemic integrity of the research process is then compromised.

But this conflicts with common sense. It seems that in this case the problem does not lie with the research process, but with the audience; the audience should not have neglected the scientist's clear statement that he deviated from the traditional standard. As the scientist is clear and open about the fact that his research does not meet the traditional standard, the epistemic integrity of the research process does not seem to be threatened.

EIR6 can account for this common intuition. Because the scientist has clearly communicated to the audience that his research does not meet the traditional standard, it is not legitimate for the audience to make this assumption. Thus, the traditional standard is not a standard that the audience can legitimately assume to be met in the scientist's research process. This implies that the violation of this standard does not threaten the epistemic integrity of the research process as defined by EIR6.

This brief analysis shows that the concept defined by EIR6 is more similar to our common sense notion of epistemic integrity of the research process than the concept defined by EIR8. Hence, the first requirement of explication does not permit the omission of ‘can legitimately’ from EIR6.

Chapter 8
The Space Shuttle Program[1]

The concept of epistemic integrity of the research process can be used to identify epistemically problematic interest influences in the sciences. This will be illustrated in chapters 8-10. In this chapter, the focus is on a case from aerospace science: the Space Shuttle Program. The Space Shuttle Program was the U.S. government's program for the development and exploitation of reusable spacecrafts from 1981 to 2011, conducted by National Aeronautics and Space Administration (NASA). Five Space Shuttle orbiters were developed as part of this program – Columbia, Challenger, Discovery, Atlantis, and Endeavour – and in total there were 135 Space Shuttle missions – Columbia had 28 flights, Challenger 10, Discovery 39, Atlantis 33, and Endeavour 25. What I will argue here is that the financial and career interests of those involved in the Space Shuttle Program compromised the epistemic integrity of different kinds of research processes. In that way, the influence of these interests is shown to be epistemically problematic.

I proceed as follows. First, three different kinds of research processes in the Space Shuttle Program are distinguished (section 8.1). For each of these research processes, I offer some specific examples of compromised epistemic integrity (sections 8.2-8.4). Next, the fact that the epistemic integrity of the relevant research processes was damaged is explained on the basis of the financial and career interests of those involved (section 8.5). Finally, some important advantages of my analysis over earlier analyses of the Space Shuttle Program are discussed (section 8.6).

[1] The examples used in this chapter also appear in De Winter & Kosolosky (2013a). But the analyses of these examples offered here are entirely new, since they are based on a concept of epistemic integrity that differs from the concept used in De Winter & Kosolosky (2013a). (Laszlo Kosolosky is not mentioned as a second author of this chapter because I have not only carried out the new analyses, but I have also collected the original examples.)

8.1 Three different kinds of research processes

In the Space Shuttle Program, there were several different kinds of research activities and different channels to report the results of these activities. Together, a research activity and the communication of its results constitute a research process. At least three different kinds of research processes can be discerned in the Space Shuttle Program.

A first kind of research process consists of research performed by working engineers and the communication of the results to other working engineers and to middle-level managers. This includes tests on shuttle components before launch, and post-flight analysis of shuttle components (checking whether it was damaged during flight, whether it had worked properly, etc.). Involved in research on a shuttle component were: (1) working engineers from the NASA contractor that developed this component, and (2) working engineers from the NASA center that was responsible for the management of this component. They reported their findings to each other and to middle-level managers. A formal channel for such communication were the Level IV and Level III meetings of Flight Readiness Review (FRR). FRR was a set of meetings (proceeding from Level IV to Level I) that preceded each launch, and at which the readiness of the shuttle to fly and to fly safely was determined. At Level IV meetings, contractor engineers presented their data analyses and conclusions to their own managers and to personnel from the relevant NASA center, and at Level III meetings, they informed NASA Project Managers of data and findings on risk acceptability (Vaughan 1996, p. 84). There was also a lot of communication between contractor engineers, NASA center engineers, and their managers outside FRR. For instance, during flight, there was daily communication between contractor engineers and NASA center engineers (Vaughan 1996, p. 85).

A second kind of research process is the review of working engineers' research results by middle-level managers (from NASA contractors and NASA centers). These middle-level managers checked whether data were consistent, whether conclusions were adequately supported by data, whether arguments met NASA standards of quantification, etc. They communicated their findings to the working engineers, and after the required adjustments were made in the engineering analysis (by themselves and/or the working engineers), they passed the results up the hierarchy, to top-level managers. The former kind of communication (feedback to working engineers) occurred at Level IV and Level III meetings, and the latter kind of communication (passing the results up the hierarchy) occurred at Level II and Level I meetings (Vaughan 1996, p. 84).

A third kind of research process consists of the research and/or discussions that are the basis of NASA's communication with the public. This communication includes the many press releases from NASA. For instance, in 1991, there were 203 press releases from NASA. Of the 195 press releases from 1991 that I could consult on the website of NASA, 68 contained the word 'shuttle' (which means that they related to the Space Shuttle Program). But this is not the only way in which NASA informed the public. It also published technical reports on the Space Shuttle, FRR reports, messages between NASA employees, etc. Several of these documents were released as a response to requests from the Freedom of Information Act.[2]

8.2 Research by working engineers

In order to demonstrate that the epistemic integrity of working engineers' research processes was sometimes damaged, I should show that certain epistemic standards which the audience could legitimately assume to be met, were violated in these research processes. Let me give an example from the eve of the Challenger disaster, i.e. the breakdown of Space Shuttle Challenger on January 28, 1986, which was caused by failure in the O-rings that were supposed to seal a joint on one of the shuttle's solid rocket boosters.

On January 27, 1986, the day before Challenger's fatal mission, two teleconferences were held, in which engineers and managers associated with the Solid Rocket Booster (SRB) Project located at Morton Thiokol Corporation (NASA contractor that built the SRBs), Marshall Space Flight Center (NASA center that bore managerial responsibility for the SRBs), and Kennedy Space Center (NASA center that assembled and tested the shuttle components and conducted launches) discussed the effect of the expected cold temperatures on the O-rings that were supposed to seal certain joints on the SRBs. Before the teleconferences, launch time was set at 9:38 A.M., and at that time, temperature was predicted to be 26°F. During the first teleconference, Thiokol suggested to delay launch until noon or after, when temperatures would be higher. A second teleconference was arranged in which Thiokol would take an official position and in which it would support this position with engineering data (Vaughan 1996, Chap. 8).

[2] See http://www.nasa.gov/.

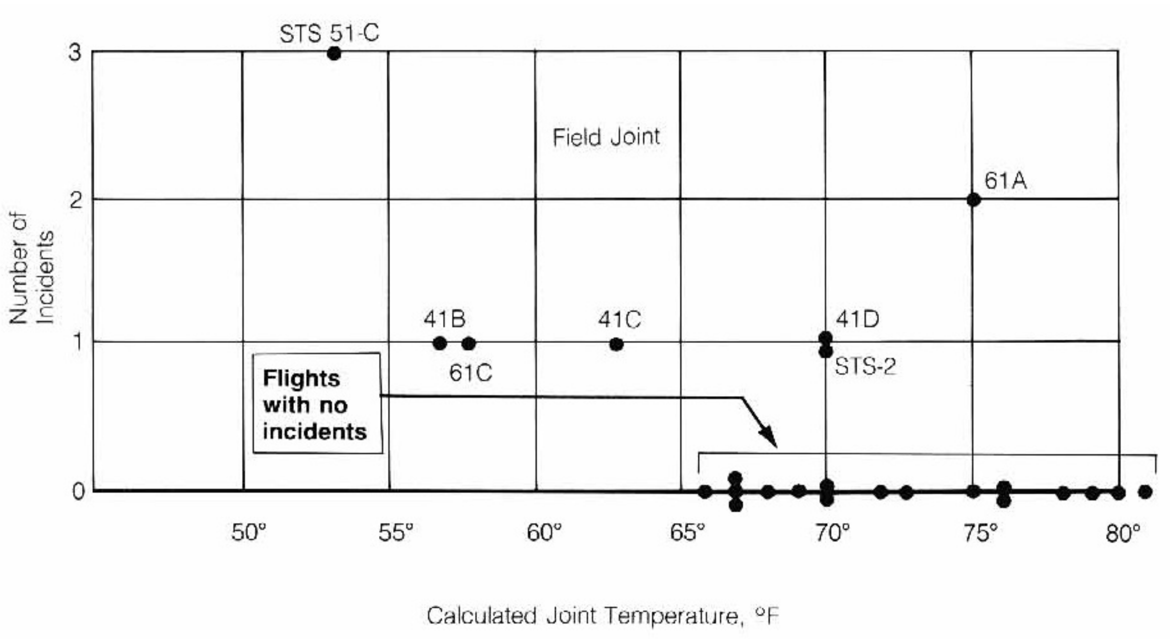

Figure 8.1 Chart on the relation between incidents of O-ring thermal distress and temperature, created by Alton Keel and Randy Kehrli. (Reproduced from Presidential Commission 1986, Vol. 1, p. 146.)

At the second teleconference, Thiokol initially recommended not to launch unless O-ring temperature was at least 53°F. Several charts were presented to support that position. The charts revealed that at the Space Shuttle's fifteenth mission, when the calculated O-ring temperature was 53°F, there was O-ring damage. Thiokol engineer Roger Boisjoly was asked what evidence Thiokol had that O-ring erosion on that mission was due to the cold. After all, there also was field joint blow-by on the Space Shuttle's twenty-second mission, when calculated O-ring temperature was not that low. Boisjoly answered that he did not have the data to quantify the temperature concerns (Vaughan 1996, pp. 303-304).

But in fact, Thiokol did have such data. This was shown by two members of the investigative staff of the Presidential Commission that U.S. President Ronald Reagan established after the Challenger disaster to investigate the circumstances leading up to the accident. Alton Keel, the executive director of this investigative staff, and Randy Kehrli, a Department of Justice attorney, created a chart that included all missions, indicating for each mission the number of O-ring anomalies and calculated O-ring temperature (see figure 8.1). This chart showed a clear correlation between O-ring anomalies and temperature: while only 17.6 percent of the flights with O-ring temperature above 65°F had O-ring anomalies, 100 percent of the flights with O-ring temperature below 65°F had O-ring anomalies (Vaughan 1996, p. 382).

Boisjoly could have found this correlation as well; he only had to apply basic statistical methods to the data available to Thiokol. But he did not do that, which led him to the conclusion that the temperature concerns could not be quantified on the basis of the available data. The fact that basic statistical methods had not been applied to the data available to Thiokol, entails that an epistemic standard which the audience could legitimately assume to be met was in fact not met in the research process. The audience could legitimately expect that Boisjoly had at least applied basic statistical methods to the data available to Thiokol before he concluded that he did not have the data to quantify the temperature concerns. Because this epistemic standard was not met, the epistemic integrity of the research process was compromised.

It should be noted that not only Thiokol's Boisjoly was involved in this research process; engineers from Marshall were involved as well. Take for instance Marshall S&E engineer (engineer from Marshall's Science and Engineering Directorate) Ben Powers. The job of S&E engineers was to keep the contractor honest (Vaughan 1996, p. 86). Hence, if someone from Thiokol said that he/she did not have the data to quantify certain temperature concerns, then Powers should have checked whether this was indeed the case. He should have checked whether the temperature concerns could not be quantified by applying basic statistical methods to the available data, and if they could, he should have protested. This was his job. But Powers did not protest, not even when George Hardy, Marshall's Deputy Director of Science and Engineering, asked at the end of the teleconference "Has anybody got anything to add to this?" (Vaughan 1996, p. 368-369). Thus, not only Thiokol's Boisjoly, but Marshall's Powers too could be held responsible for the failure to apply basic statistical methods to the available data. This shows that the problem did not lie with Thiokol alone; Marshall too was partly responsible for the fact that the epistemic integrity of the research process was compromised.

8.3 Review by middle-level managers

The epistemic integrity of review by middle-level managers was sometimes damaged as well. Consider some examples from Lawrence Mulloy, SRB Project Manager at Marshall from 1982 to 1986. On June 25, 1985, it was found that both the primary and secondary O-ring of one of the SRB joints had eroded on the seventeenth mission of the Space Shuttle (Winsor 1988, p. 104). It was the first time that a primary O-ring had burned all the way through, allowing hot gases to erode a secondary O-ring. This finding resulted in a Launch Constraint, which meant that flight could only proceed if either the problem was resolved, or engineering analysis showed that the risk was acceptable (Vaughan 1996, pp. 163-164).

But the Launch Constraint only included the SRB nozzle joints, and not the SRB field joints, as it was a nozzle joint that failed on the Space Shuttle's seventeenth mission. The reason was that the nozzle joints were tested for leaks at 100 psi, while the other SRB joints were leak-tested at 200 psi, and Mulloy assumed that the damage in the nozzle joint was due to the less rigorous leak test for the nozzle joints. He assumed that a defective O-ring escaped notice in the 100-psi leak test, and that this caused the damage. As leak check pressure for the nozzle joints was raised to 200 psi for the subsequent flights, Mulloy waived the Launch Constraint (Winsor 1988, p. 104). This was the result of Level IV and Level III meetings, and Mulloy reported a summary of the problem and its resolution to Levels II and I (Vaughan 1996, p. 169).

It was legitimate for the audiences at Level II and Level I meetings to assume that Mulloy's explanation of the damage found in the nozzle joint (i.e. the less rigorous leak test for the nozzle joints) was adequately supported by decent scientific reasoning. But this was actually not the case; the assumption that the damage was caused by a defective O-ring that escaped notice due to the 100-psi leak test, was an inadequately supported conjecture rather than an explanation which was backed by proper scientific arguments.[3] Let me clarify this.

An alternative explanation for the damage in the nozzle joint was that it was caused by defective design. This hypothesis was, however, dismissed. Wiley Bunn, Marshall Director of Reliability and Quality Assurance, justified the rejection of this hypothesis on the following grounds:

> We had six joints on that vehicle. If the design is that darn bad, all six of them should have leaked. We only had one leak. Therefore, if we only had one leak, it had to be a Quality escape. And so we just renewed our vigor to find that Quality escape. (Bunn, interview transcript, April 17, 1986, pp. 61-62, cited in Vaughan 1996, pp. 164-165)

This is obviously a very weak argument: the fact that five out of six joints did not leak does not imply that the design was fully adequate – even a design that is not fully adequate can work most of the time. Hence, Mulloy's assumption that the damage found was not due to defective design, but to the less rigorous leak test for the nozzle joints, was not backed by proper scientific arguments.

This means that an epistemic standard which the audiences at Levels II and I could legitimately assume to be met in the research process, was violated. These audiences could legitimately assume that Mulloy's review was based on proper scientific arguments rather than on an inadequately supported conjecture. Because Mulloy's review did not meet this standard, its epistemic integrity was damaged.

This was not the first time that the epistemic integrity of Mulloy's work was compromised. When the Space Shuttle's SRBs were examined after its fifteenth flight,

[3] Thiokol's Boisjoly seemed to be aware of this. In an internal memo he sent to Robert Lund on July 31, 1985, he stated that the same scenario that resulted in the failure of the nozzle joint, could also occur in a field joint (Presidential Commission 1986, Vol. 1, pp. 249-250).

Thiokol engineers found blow-by on two SRB field joints, which was 'jet black' and intermixed with the grease between the O-rings. This alarmed them, because it indicated that there were two destructive forces at work on the O-rings, impingement erosion and blow-by erosion. In explaining this, Thiokol engineers referred to the cold: the fifteenth mission was preceded by the three coldest days in Florida history and O-ring temperature was 53°F, which was the lowest up to that time, and such low temperatures enhanced the probability of blow-by. This figured in their rationale for accepting risk: as Thiokol engineers did not expect to have the three coldest days in Florida history again, they accepted the risk for the next flight (Vaughan 1996, pp. 154-163).

However, as the rationale for accepting risk was reported up the hierarchy, temperature considerations were omitted. When the risk associated with O-ring erosion was assessed for the sixteenth shuttle mission, the temperature concerns of Thiokol engineers were considered at lower levels of FRR, but when Mulloy presented the rationale for accepting risk to Level I, he did not mention these concerns. He stated that the risk was acceptable because (1) the amount of primary erosion (0.038") was within the experience base (the 0.053" erosion found after the Space Shuttle's second mission), (2) the amount of erosion was within the 0.090" safety margin, and (3) the effect of hot gas flow on the O-rings – impingement erosion and blow-by – was self-limiting. No reference was made to the temperature concerns (Vaughan 1996, p. 161).

The Level I audience could legitimately assume that Mulloy's review met the following epistemic standard: all reasons for accepting the risk associated with O-ring erosion should be summed up, and reported to Level I. This standard was, however, violated. The expectation that the cold preceding the fifteenth mission would not recur was one of the reasons for accepting this risk, but Mulloy did not include it in his report. This implies that the epistemic integrity of his review was compromised.

Besides the work of Marshall's Mulloy, other reviews by middle-level managers were epistemically troublesome as well. Let me give an example from management at Thiokol. Richard Feynman, who was a member of the Presidential Commission on the Space Shuttle Challenger Accident, has exposed a strange use of the term 'safety factor'. When the risk associated with O-ring erosion was assessed for the Space Shuttle's twenty-fifth mission, it was asserted that there was a safety factor of three to one. The idea was that on the Space Shuttle's fifteenth mission, erosion depth was one-third of the radius, while it had to be at least one radius before the O-ring failed (Feynman 1986, pp. F1-F2). This use of the term 'safety factor' was adopted by Thiokol management, as the following quote from Allan McDonald, Director of the Solid Rocket Motor Project at Thiokol, indicates:

> If you took our worst measured erosion on the O-ring relative to what it took to really fail it, it was nearly a factor of three to one. Recognizing the fidelity of the math

> model is not real good, we did not feel it was that bad at three to one, and as long as we could retain the secondary seal during a good portion of the erosion time period, we felt good. (Presidential Commission 1986, Vol. 5, p. 1591)

But this use of the term ‘safety factor’ did not correspond with how this term was used in other engineering contexts, as Feynman showed on the basis of the following example:

> If a bridge is built to withstand a certain load without the beams permanently deforming, cracking, or breaking, it may be designed for the materials used to actually stand up under three times the load. This ‘safety factor’ is to allow for uncertain excesses of load, or unknown extra loads, or weaknesses in the material that might have unexpected flaws, etc. If now the expected load comes on to the new bridge and a crack appears in a beam, this is a failure of the design. There was no safety factor at all; even though the bridge did not actually collapse because the crack went only one-third of the way through the beam. The O-rings of the Solid Rocket Boosters were not designed to erode. Erosion was a clue that something was wrong. Erosion was not something from which safety can be inferred. (Feynman 1986, p. F2)

This example demonstrates that the safety analysis of the O-rings by Thiokol management did not meet common standards for assessing safety. It was however legitimate for the audience to assume that common standards for calculating safety factors *were* met in this analysis, as the deviation from these standards was never clearly communicated. Since Thiokol management violated certain epistemic standards which the audience could legitimately assume to be met, the epistemic integrity of its work was compromised.

8.4 Communication with the public

The third kind of research process I identified is the research and/or discussions that are the basis of NASA’s communication with the public. An example is NASA’s decision to declare the Space Shuttle Program ‘operational’, which was publicly announced on July 4, 1982 (Vaughan 1996, p. 125). This announcement suggested that the Space Shuttle had attained an airline-like degree of routine operation (Presidential Commission 1986, Vol. 1, p. 5). It was then legitimate for the audience to assume that research had shown that the Space Shuttle was sufficiently safe for routine use. Furthermore, the audience could legitimately assume that this research met common epistemic standards for judging a vehicle to be sufficiently safe for routine use. The latter was, however, not the case; NASA used much lower epistemic standards.

When NASA publicly declared the Space Shuttle Program 'operational', the Space Shuttle was still a developmental craft with constantly changing technology and mysterious problems that were not predicted from design (Hall 2003, p. 240). In other areas of aviation, it would in such a situation definitely be too early to proceed to the operational phase. So in other areas of aviation, much higher standards were used to consider a vehicle sufficiently safe to proceed to the operational phase. This indicates that NASA's research process did not live up to common epistemic standards for judging a vehicle to be sufficiently safe for routine use. Because the audience could legitimately assume that common standards *were* lived up to, the epistemic integrity of the research process was compromised.

Another example is the estimation that the probability of mission failure was 1 in 100,000, a figure that was published in 'Space Shuttle Data for Planetary Mission Radioisotope Thermoelectric Generator (RTG) Safety Analysis' on February 15, 1985 (Salmon 2005, p. 127). Let me show that the research process leading to this figure did not live up to the epistemic standards which the audience could legitimately assume to be met in it.

An estimate of the reliability of the SRBs was made by the range safety officer on the basis of past performance. As 121 out of 2,900 flights failed, the probability of failure was approximately 1 in 25. But this included rockets that were flown for the first few times, and for the mature rockets, a figure of 1 in 50 might have been more reasonable. Furthermore, if parts would be carefully selected and inspected, even a figure of 1 in 100 might have been achievable. A figure of 1 in 1,000 was probably not achievable at the time. But this conflicted with NASA management's opinion on the matter. NASA officials estimated that the probability of failure with loss of vehicle and of human life was much lower than 1 in 100 – the lowest estimate being 1 in 100,000. They argued that the high figures were based on past performance of unmanned rockets, and that a distinction had to be made between manned space flight programs and unmanned programs. The argument was that because manned programs had an extremely high degree of mission success, standard statistical methods could not be used to determine the probability of mission failure, and therefore they had to rely on engineering judgment, and not on numerical probability usage (Feynman 1986, p. F1).

As NASA claimed to rely on engineering judgment, it was legitimate for the audience to assume that the research process which led to the figure of 1 in 100,000 met the following epistemic standards: (1) working engineers judged the probability of mission failure, and (2) the figure of 1 in 100,000 was supported by their judgments. But the latter epistemic standard was *not* met; the figure of 1 in 100,000 was in fact *not* supported by working engineers' judgments, as they estimated the probability of mission failure much higher (Feynman 1986, p. F1). Therefore, the epistemic integrity of the research process was compromised.

8.5 The role of non-epistemic interests

So the epistemic integrity of different kinds of research processes in the Space Shuttle Program was compromised. In this section, I offer an explanation for this fact. More specifically, I will explain this fact on the basis of the financial and career interests of the different agents involved in the Space Shuttle Program.

Let me start with a very brief sketch of the historical context in which this program was situated. It was a context of competition for scarce resources. The Space Shuttle Program was born in the aftermath of the Vietnam War, when spaceflight was no longer national priority and when NASA lost the budgetary certainty it had before. NASA personnel was reduced by 1,000 employees per year, and of the three projects that NASA planned – a mission to Mars, a space station in earth orbit, and a space shuttle to transport people and materials in space – it could only execute the space shuttle project, which aimed at the development of a reusable space shuttle that should reduce the cost of putting objects into orbit (Vaughan 1996, pp. 18-19).

While the Apollo Program was justified by the desire to respond to the Soviet launch of Sputnik, the Space Shuttle Program became justified on the basis of cost-effectiveness. The Space Shuttle Program gained approval on the basis of a study by Mathematica, Inc., a think tank that NASA called in in 1971 to assess the program's cost-effectiveness. Mathematica reported that, given the Space Shuttle's payload capacity, it would pay for itself provided that there would be more than 30 flights each year. In light of this economic justification for the Space Shuttle Program, NASA had to maintain before Congress and the general public that the program was a good investment on economic grounds. But the estimates by Mathematica were overly optimistic – among others because it was based on data furnished by contractors hoping to receive shuttle contracts – and given workforce reductions at NASA, it became increasingly difficult for NASA to meet performance expectations. The gap between what NASA could do and what it was expected to do was widening, and this resulted in production pressure, i.e. the pressure to launch a certain amount of flights each year, more than was possible given NASA's means (Vaughan 1996, pp. 19-32). NASA had to launch as much flights as possible in order to maintain government funding. Of course, this pressure also had an impact on NASA contractors, since they had to meet NASA requirements in order to maintain contracts.

These circumstances could explain why epistemic integrity was compromised in the Space Shuttle Program. Take for instance the three-factor rationale for accepting risk associated with O-ring erosion which Mulloy presented to Level I (see section 8.3). As we have seen, this rationale did not include Thiokol engineers' temperature concerns, which means that the following epistemic standard was violated: all reasons for accepting the risk associated with O-ring erosion should be summed up, and reported to Level I. But before the temperature concerns could be communicated to Level I, systematic data on the association between temperature and O-ring erosion had to be collected, since only 'solid engineering data' were admissible in FRR presentation (Vaughan 1996, pp. 161-162). This would, however, have taken time, and a more time-saving option was to simply omit these concerns. By omitting these concerns, the process of getting a shuttle ready for launch was accelerated. The latter served the financial interests of NASA and its centers, since the more shuttles would be launched per year, the more likely it was that the government would continue to support the Space Shuttle Program. Accordingly, it also served the financial and career interests of the people involved in this program. So these non-epistemic interests could explain why Mulloy violated the relevant epistemic standard: by violating the standard that all reasons for accepting the risk associated with O-ring erosion should be reported to Level I, he protected the financial and career interests of himself and his coworkers.

It should be noted that these interests were only served if the Level I audience did not know that the relevant epistemic standard was violated. The Level I audience had to believe that this standard was met. Otherwise, Level I would have protested, and it would oblige Mulloy to present the *complete* rationale, which would mean that systematic data on the association between temperature and O-ring erosion would have to be collected after all. Hence, Mulloy's violation would then have slowed down rather than accelerated the process of getting a shuttle ready for launch. So for Mulloy's strategy to work it was crucial that the violation of the relevant epistemic standard was kept secret, so that the Level I audience could legitimately assume that it was met. In summary: to protect the financial and career interests of himself and his coworkers, Mulloy had to (1) violate an epistemic standard, and (2) let his audience assume that it was met. We can then conclude that these interests compromised the epistemic integrity of the research process.

Financial and career interests could also explain why epistemic integrity was damaged in the other examples discussed above (sections 8.2-8.4). Consider, for instance, NASA's public statement that the probability of mission failure was 1 in 100,000, a figure which the audience could legitimately assume to be supported by the judgments of working engineers (see section 8.4). This statement gave the audience the impression that shuttle launches were extremely safe, which contributed to the public acceptability of the Space Shuttle Program, and of government funding of this program.

So indirectly it served the financial and career interests of those involved in the program.

The figure of 1 in 100,000 could, however, not be obtained on the basis of the judgments of working engineers, since they estimated the probability of mission failure much higher. This explains why NASA violated the standard of relying on working engineers' judgments: because violating this standard was necessary to arrive at a favorable figure, that is, a figure which would serve the financial and career interests of those involved in the Space Shuttle Program. So an epistemic standard which the audience could legitimately assume to be met was violated because this served the financial and career interests of those involved. These interests can then be said to have compromised the epistemic integrity of the research process.

Note that it was important for NASA that the audience did not know that the relevant epistemic standard was violated. If the audience had known that the figure of 1 in 100,000 was in fact not supported by working engineers' judgments, then this figure would probably not have had the same impact on the audience's opinion. The audience would then probably have been more skeptical towards the Space Shuttle Program and towards government funding of this program, so that the financial and career interests of those involved in the program would not have been served.

Similar explanations can be constructed for the other examples of compromised epistemic integrity discussed above (sections 8.2-8.4). In each example, the financial and career interests of those involved in the research process were served by violating an epistemic standard which the audience could legitimately assume to be met. These interests could therefore explain why the epistemic integrity of the research process was compromised. This is not to say that in each case, an epistemic standard which the audience could legitimately assume to be met was *intentionally* violated, that is, with the explicit intention of serving one's non-epistemic interests. It is also possible that such interests compromised the epistemic integrity of the research process without those involved being aware of this. To illustrate this, consider Roger Boisjoly's failure to apply basic statistical methods to the data available to Thiokol on the eve of the Challenger disaster (see section 8.2). It seems that this was an inadvertent mistake on behalf of Boisjoly; he did not seem to be aware of his failure. Still, this failure could be explained on the basis of non-epistemic interests. Let me clarify how.

Firstly, there was a certain time pressure at NASA and its contractors, since working too slowly could jeopardize future funding or contracts. Accordingly, employees had to proceed as quickly as possible in order to satisfy their financial and career interests. This made failures such as Boisjoly's more likely. In the absence of the relevant financial and career interests, Boisjoly may have seen how applying basic statistical methods to the data available to Thiokol would deliver a clear quantitative argument for his temperature concerns. The epistemic integrity of the research process would then not have been damaged.

There is also a second way in which non-epistemic interests may have affected Boisjoly's work. Due to the pressure to launch, NASA contractors and their employees did not want to be the organization/person that was responsible for postponing launch. This applied especially to Thiokol on the eve of the Challenger launch, since a teleconference to discuss Thiokol's one-billion-dollar contract was scheduled for January 28, after the Challenger lift-off (Charles 1989, p. 118). So if Thiokol would be responsible for the Challenger launch to be postponed, the financial consequences for Thiokol could be terrible.

Now, if Thiokol's Boisjoly had applied basic statistical methods to the available data, then this would have resulted in a clear quantitative signal that cold temperatures could cause O-ring erosion, and launch would have probably been postponed. Boisjoly and Thiokol would then be responsible for launch delay, which would, as we have seen, threaten Thiokol's financial interests, and thus also Boisjoly's financial and career interests. Of course, being responsible for launch delay was still preferable to being responsible for mission failure. But this was not what was expected; even Boisjoly seemed to expect that the mission would return (even though he was aware of the risk involved) (Vaughan 1996, p. 380). Moreover, if the mission would fail, as it did, Boisjoly would not be seen as the person who was responsible for this, but rather as the hero who tried to prevent the tragedy.[4]

So the failure to develop a clear quantitative argument for his temperature concerns was expected to serve Boisjoly's non-epistemic interests. This may have inhibited his creativity in developing such an argument. If his interests were not expected to be served by the failure to develop a quantitative argument for his temperature concerns, then Boisjoly may have been more inventive in developing such an argument, and he would probably have at least checked whether such an argument could be produced by applying basic statistical methods to the available data. The epistemic integrity of the research process would then not have been damaged. So by demotivating him to develop a clear quantitative argument for his temperature concerns, Boisjoly's non-epistemic interests may have caused the epistemic integrity of the research process to be damaged in a way that Boisjoly himself was not aware of.

I should remark that it is difficult to prove that the explanations developed in this section are correct. This is a question which requires more research. The main virtue of my analysis is then not that it provides evidence for the existence of certain interest

---

[4] Also see the report of the U.S. House Committee on Science and Technology:

> [T]he present system permits [contractors] to 'express concern' without actually saying, 'stop the flight, it is unsafe'. If the odds favor a successful flight they do not have to be responsible for cancelling, yet if the mission fails they are on record as having warned about potential dangers. (House Committee 1986, p. 152)

influences in the Space Shuttle Program. Its main virtue is that if we assume that research processes in the Space Shuttle Program were actually influenced by non-epistemic interests in the ways described in this section, then my analysis shows why these influences were epistemically problematic.

8.6 Advantages of my analysis

In this last section of chapter 8, I want to sum up some advantages of my analysis over earlier analyses of the Space Shuttle Program. The practices that were characterized as problematic in this chapter have already been recognized as problematic in earlier work. But the justification I offered is new. Moreover, I think that my justification in terms of epistemic integrity outperforms earlier justifications. Take, for instance, Feynman's criticism on NASA's use of the term 'safety factor'. He argues that the term 'safety factor' was used in a strange way. But it is not because something is strange that it is necessarily problematic. I was able to show why the strange use of the term 'safety factor' *was* problematic: because it implied that the epistemic integrity of the research process was compromised (see section 8.3).

More generally, it seems that earlier justifications often start from certain traditional or common standards, which were violated in the relevant cases. They do, however, not clarify why adherence to these standards is crucial. In itself, deviating from traditional or common standards is not necessarily problematic (see section 6.5.2). My analysis reveals why it *was* problematic in the relevant cases: because it implied that the epistemic integrity of the research process was compromised.

It should also be noted that most earlier analyses focus on a certain accident (the Challenger disaster or the Columbia disaster) and search for factors that contributed to this accident. These factors are then regarded as problematic partly on that ground. But things might have been different. If some environmental factors (e.g., temperature) were a bit different, the Challenger or Columbia disaster might not have occurred, even if NASA and its contractors would have proceeded exactly the same as they currently have. Perhaps NASA just had bad luck. The problem with accounts that evaluate practices in light of events such as the Challenger or Columbia disaster is then that the outcomes of such studies are heavily dependent upon contingent circumstances. If things had been different, and no accidents had occurred, such studies might have very different outcomes; they might conclude that NASA should be praised for its efficiency at getting shuttles in space and back on earth safely. What I did, is show that certain research processes in the Space Shuttle Program were problematic – their epistemic integrity was damaged – in a way that is independent of whether or not these research processes resulted in certain accidents.

Chapter 9
Controversies concerning the IPCC[1]

(Co-author: Laszlo Kosolosky)

The concept of epistemic integrity can also be used to analyze cases from other fields than aerospace science. This chapter concerns climate science. The focus is on the work of the Intergovernmental Panel on Climate Change (IPCC). The IPCC is the leading international body for assessing climate change. It provides the world with a clear scientific view on the current state of knowledge on climate change and its potential environmental and socio-economic impacts. What it does exactly is review and assess the most recent scientific, technical, and socio-economic information produced worldwide relevant to the understanding of climate change.[2] It publishes its analyses in what it calls Assessment Reports.

There has been a lot of controversy around the IPCC's Assessment Reports and how they were produced. It has been insinuated that political interests have corrupted the production of these reports. In this chapter, two particular examples of such controversy are presented (section 9.2) and analyzed in terms of epistemic integrity (section 9.3). We assess the epistemic integrity of the work by the IPCC which was the target of controversy, in order to see whether this work may have been influenced by political interests in epistemically unacceptable ways. But first, let us give some background information on the IPCC and its procedures.

---

[1] Laszlo Kosolosky has collected the examples that are used in this chapter, and I have analyzed them on the basis of the new concept of epistemic integrity (as defined by EIR6; see section 6.7.1).
[2] See http://www.ipcc.ch/organization/organization.shtml#.UtZ8xO0VGM8.

9.1 Inner workings of the IPCC

The IPCC is a scientific body under the auspices of the United Nations (UN). It was established in 1988 by the United Nations Environment (UNEP) and the World Meteorological Organization (WMO) with the purpose of providing a clear scientific view on the current state of knowledge on climate change and its potential environmental and socio-economic impacts. Its task is to review and assess the most recent scientific, technical, and socio-economic information produced worldwide relevant to the understanding of climate change. The results are bundled in Assessment Reports. The IPCC tries to ensure that these reports offer an objective and complete assessment of current information, which reflects a wide range of views and expertise. Thousands of scientists from all over the world contribute to the work of the IPCC on a voluntary basis. The IPCC does not conduct its own research nor does it monitor climate related data or parameters.[3]

The IPCC is an intergovernmental body which is open to all member countries of the UN and the WMO. Currently 195 countries are members of the IPCC. Governments participate in the review process and the plenary Sessions, where main decisions about the IPCC work program are taken and reports are accepted, adopted, and approved. By endorsing the IPCC Assessment Reports, governments acknowledge the authority of their scientific content.[4]

To divide the work of writing an Assessment Report, the organization divided itself into three Working Groups:

(1) Working Group I assesses the physical scientific aspects of the climate system and climate change, including attribution of past change and projections of future change.
(2) Working Group II assesses the vulnerability of socioeconomic and natural systems to climate change, negative and positive consequences of climate change, and options for adapting to it.
(3) Working Group III assesses policy and technology options for mitigating climate change through, for example, limiting and preventing greenhouse gas emissions and enhancing activities that remove them from the atmosphere. (InterAcademy Council 2010, p. 6)

---

[3] See http://www.ipcc.ch/organization/organization.shtml#.UtZ8xO0VGM8.
[4] See http://www.ipcc.ch/organization/organization.shtml#.UtZ8xO0VGM8.

To ensure proper preparation and analysis, the IPCC devised a thorough peer review procedure that contains the following chronological steps:

(1) Compilation of lists of potential Coordinating Lead Authors, Lead Authors, Contributing Authors, Expert Reviewers, Review Editors and Government Focal Points.
(2) Selection of Coordinating Lead Authors, Lead Authors, and Review Editors.
(3) Preparation of draft report.
(4) Review:
    (a) First review (by experts).
    (b) Second review (by governments and experts).
(5) Preparation of final draft report.
(6) Acceptance of report at a Session of the Working Group(s). (IPCC 2013, pp. 3-9)[5]

In step 1, potential Coordinating Lead Authors, Lead Authors, Contributing Authors, Expert Reviewers, Review Editors and Government Focal Points are identified. In step 2, a selection of Coordinating Lead Authors, Lead Authors, and Review Editors is made. These groups should reflect different scientific, technical, and socio-economic views, and different geographical backgrounds. In step 3, the Coordinating Lead Authors and Lead Authors make up a draft report. This procedure is governed by three principles. First, the best available scientific and technical advice has to be included in the report to assure that the report is up-to-date and as comprehensive as possible. Second, the procedures have to assure that participating countries are represented as much as possible. Third, the review process has to be objective, open, and transparent. In step 4, experts and governments analyze the draft report by investigating how well the authors have acted in accordance with the three principles spelled out above. There are two review circulations: the first by experts, and the second by experts and governments. The group of expert reviewers should represent, insofar as possible, a wide range of scientific, technical, and socio-economic views, expertise, and geographical backgrounds. In step 5, Coordinating Lead Authors and Lead Authors prepare the final version of the report in consultation with the Review Editors, taking into account possible criticisms resulting from step 4. The last step is the acceptance of the report at a Session of the Working Group(s) (IPCC 2013, pp. 3-9).

Given the relevance of the IPCC's work for policy, political interests may come into play in the review process. For instance, the governments involved in this process may

[5] For a detailed description of the tasks and responsibilities for Lead Authors, Coordinating Lead Authors, Contributing Authors, Expert Reviewers, Review Editors and Government Focal Points, see IPCC (2013, pp. 14-16).

take such interests into account in reviewing the report. When such interests affect the IPCC's work, this is not necessarily a problem, but it can be. The concept of epistemic integrity enables us to determine whether or not a certain interest influence on the IPCC's work is epistemically problematic (as will be illustrated in section 9.3).

9.2 Two controversies

Not everyone accepts the authority of the IPCC and its Assessment Reports. This section considers two controversies in which skeptics have tried to discredit the IPCC. The first relates to the IPCC's Second Assessment Report (IPCC 1995), and the second to the IPCC's Fourth Assessment Report (IPCC 2007).

9.2.1 Second Assessment Report

The focus of the first controversy was on chapter 8 of Working Group I's contribution to the Second Assessment Report. That chapter summarizes the evidence concerning human influences on global climate, which suggests that climate change is partly caused by humans. Chapter 8 also indicates that there were still some large uncertainties remaining (Santer et al. 1995).

The chapter under consideration, and more specifically Benjamin Santer, one of its Lead Authors, has been criticized by a group of physicists tied to the George C. Marshall Institute, a conservative think tank in Washington, D.C. They accused Santer of 'scientific cleansing', in the sense that he had removed unfavorable views from the final Assessment Report. Some of them linked this misconduct to political purposes. Such accusations were spread through reports in newspapers, and through letters to congressmen, to officials of the Department of Energy, and to editors of scientific journals. Santer responded that he had not done anything wrong, which plenty of scientists confirmed (Oreskes & Conway 2010, pp. 3-5).

A question that arises is the following: Did political interests compromise the epistemic integrity of the review by the IPCC, and by Santer in particular? We address this question in section 9.3.1, arguing that the epistemic integrity of the IPCC's work was not damaged.

9.2.2 Fourth Assessment Report

More recent skeptic attacks were targeted against the IPCC's Fourth Assessment Report. These critical voices reached a peak when the IPCC-board had to acknowledge that the report contained a false prognosis of the disappearance of the Himalayan glaciers (IPCC 2010; Kosolosky forthcoming). The disputed paragraph addressing the melting rates goes as follows:

> Glaciers in the Himalaya are receding faster than in any other part of the world (see Table 10.9) and, if the present rate continues, the likelihood of them disappearing by the year 2035 and perhaps sooner is very high if the Earth keeps warming at the current rate. Its total area will likely shrink from the present 500,000 to 100,000 km^2 by the year 2035 (WWF, 2005). (Cruz et al. 2007, p. 493)

This paragraph is based on a source issued by the World Wide Fund for Nature (WWF). In that source, the prediction that the Himalayan glaciers will vanish is supported by a 1999 report by the Working Group on Himalayan Glaciology of the International Commission for Snow and Ice, and a 1999 article in popular science magazine New Scientist (WWF 2005, p. 29). Both are based on non-refereed findings by glaciologist Syed Hasnain, who later admitted that these findings were speculative (Schiermeier 2010, p. 276).

Moreover, "its total area will likely shrink from the present 500,000 to 100,000 km^2 by the year 2035" in the above paragraph cannot refer to the Himalayan glaciers, since the Himalayan glacier area was about 33,000 km^2 (Cogley et al. 2010). According to Cogley et al. (2010), that statement probably originates from a report published by UNESCO, which states that "its total area [of all glaciers worldwide] will shrink from 500,000 to 100,000 km^2 by the year 2350" (Kotlyakov 1996, p. 66).

Skeptics have used the Himalayan glacier flaw to cast doubt on the IPCC's conclusions and on human-caused climate change. An example is Mark Steyn's opinion piece 'Credibility is what's really melting', published in the Canadian news magazine Maclean's. In that opinion piece, Steyn uses the fact that the above quote from the Fourth Assessment Report is ultimately based on idle speculation by Hasnain, to show that 'climate change' is not a story of climate change, but of the corruption of science. He also claims that the quote was included in the IPCC report for political reasons (Steyn 2010).

In light of Steyn's opinion piece, it seems that political interests have compromised the integrity of the IPCC process. But is this correct? In section 9.3.2, we discuss the epistemic integrity of the IPCC process resulting in the Himalayan glacier flaw. We argue that epistemic integrity was indeed compromised (possibly by non-epistemic interests, including political interests), but that it was restored after a while.

9.3 Analysis in terms of epistemic integrity

9.3.1 Second Assessment Report

To recap: a group of physicists accused Benjamin Santer of 'scientific cleansing' (that is, expunging unfavorable views) in producing chapter 8 of Working Group I's contribution to the Second Assessment Report. Perhaps the most famous of these attacks is Frederick Seitz's op-ed piece in the Wall Street Journal, entitled 'A major deception on 'global warming''. In that piece, Seitz claims that IPCC rules were violated in the process leading to the chapter under consideration. More specifically, certain passages were removed from this chapter after it was approved by the contributing scientists, which was according to Seitz not permitted by IPCC rules. Seitz maintains that the changes to the report "worked to remove hints of the skepticism with which many scientists regard claims that human activities are having a major impact on climate in general and on global warming in particular" (Seitz 1996). The effect was "to deceive policy makers and the public into believing that the scientific evidence shows human activities are causing global warming" (Seitz 1996). Although Seitz admits that he does not know who exactly made the relevant changes, he states that "the report's lead author, Benjamin D. Santer, must presumably take the major responsibility" (Seitz 1996).

While Seitz does not explicitly state that the alleged misconduct by the IPCC is linked to political interests, Fred Singer does make such a link in an op-ed piece that appears a month later in the Wall Street Journal. Singer refers to Seitz as revealing "that a UN-sponsored scientific report promoting global warming has been tampered with for political purposes", and he insinuates that IPCC officials are on a "crusade to provide a scientific cover for political action" (Singer 1996).

If Seitz's and Singer's op-ed pieces are correct, then it seems that the influence of political interests on the IPCC process was epistemically problematic. If political interests caused a violation of IPCC rules of procedure, which are epistemic standards which the audience can legitimately assume to be met, then they have compromised the epistemic integrity of the IPCC process. But a closer look at the actual IPCC process reveals that this is not the case. What Santer actually did, was make adjustments to the report in response to comments from reviewers. Later, the adjusted report was accepted by the full IPCC Plenary (Edwards & Schneider 1997). This was completely in accordance with IPCC rules of procedure (cf. section 9.1 above; also see Edwards & Schneider 1997; Oreskes & Conway 2010, p. 4; Santer 1996).

It should also be stressed that, contrary to what Seitz and Singer insinuate, remaining uncertainties or doubts about human-caused climate change were not suppressed in the published version of the IPCC report. Consider, for instance, the following passage in the last section of chapter 8:

> Finally, we come to the difficult question of when the detection and attribution of human-induced climate change is likely to occur. The answer to this question must be subjective, particularly in the light of the large signal and noise uncertainties discussed in this chapter. Some scientists maintain that these uncertainties currently preclude any answer to the question posed above. (Santer et al. 1995, p. 439)

This passage shows that Santer has not covered up uncertainties or doubts with respect to human-caused climate change.

It seems then that Santer has not violated any epistemic standards which the audience can legitimately assume to be met in the IPCC process; neither IPCC rules of procedure, nor the common epistemic standard that remaining uncertainties should be honestly reported. Thus, we have no reason to believe that the epistemic integrity of his work was compromised. Accordingly, we have no reason to believe that political interests influenced the IPCC process in an epistemically problematic way.

Note that this does not imply that political interests have not affected the IPCC process in any way. It is possible that some of the reviewers' comments were partly politically motivated, and that the authors' political interests had some influence on the content of the report, in the sense that some nuances would have been different if their political interests would have been different. The point is that as long as such interest influences do not involve violations of the epistemic standards which the audience can legitimately assume to be met in the process, they are not epistemically problematic.

9.3.2 Fourth Assessment Report

Now consider the IPCC process leading to the Himalayan glacier flaw in the Fourth Assessment Report. In this process, at least two different epistemic standards which the audience could legitimately assume to be met were violated. The first is the following: IPCC authors should critically assess any source that they wish to use, and if they find that the results of the source are not supported by genuine scientific evidence, then they should not include these results in the IPCC report. This standard, let us call it s_1, was violated with respect to the prediction that the Himalayan glaciers will vanish by 2035. This prediction was not based on genuine scientific evidence, but on speculation by Syed Hasnain (see section 9.2.2). If the IPCC authors had critically assessed the source of the prediction, and checked whether it was supported by genuine scientific evidence, then they would have discovered that it was not, and they should have decided not to

Comment number	Comment	Writing team notes
E10-466	100,000? You just said it will disappear. (David Saltz, Desert Research Institute, Ben Gurion University)	Missed to clarify this one
E10-468	I am not sure that this is true for the very large Karakoram glaciers in the western Himalaya. Hewitt (2005) suggests from measurements that these are expanding – and this would certainly be explained by climatic change in precipitation and temperature trends seen in the Karakoram region (Fowler and Archer, J Climate in press; Archer and Fowler, 2004) You need to quote Barnett et al.'s 2005 Nature paper here—this seems very similar to what they said. (Hayley Fowler, Newcastle University)	Was unable to get hold of the suggested references will consider in the final version (sic)

Table 9.1 Reviewer comments concerning the erroneous estimation of the Himalayan glaciers' melting rate in the IPCC's Fourth Assessment Report. (Reproduced from InterAcademy Council 2010, p. 22.)

include the prediction in the IPCC report. That is what s_1 required them to do. Because the Himalayan glacier prediction was included in the IPCC report, we can conclude that s_1 was violated.

The audience could, however, legitimately assume that s_1 was met. A document issued by the IPCC on its procedures states that "[a]uthors who wish to include information from a non-published/non-peer-reviewed source are requested to [...] [c]ritically assess any source that they wish to include" (IPCC 2008). The principle that they should not include the source's results if they find that these results are not supported by genuine scientific evidence, corresponds with the IPCC's aim "to provide the world with a clear *scientific* view on the current state of knowledge in climate change and its potential environmental and socio-economic impacts".[6] In light of these objectives of the IPCC, it was legitimate for the audience to assume that s_1 was met in the IPCC process leading to the Fourth Assessment Report. Since s_1 was actually not met in this process, the epistemic integrity of this process was damaged.

There is also a second way in which epistemic integrity was compromised. The problem lay with the review of an earlier draft of the Fourth Assessment Report. This review consisted of two review circulations. Two of the comments in the second review concerned the erroneous estimation of the Himalayan glaciers' melting rate (see table 9.1). The first pointed at the contradiction between the statement that the Himalayan

[6] http://www.ipcc.ch/organization/organization.shtml#.Utz6kO0VGM8, italics added.

glaciers are likely to disappear by 2035 and the statement that they are likely to shrink to 100,000 km^2 by 2035. The second reviewer doubted whether the Himalayan glaciers were likely to disappear and referred to some sources which supported his doubt (InterAcademy Council 2010, p. 23).

The authors have not adequately dealt with these comments (Kosolosky forthcoming). They have not removed the contradiction revealed by comment E10-466 (see table 9.1), and they have not followed comment E10-468's (see table 9.1) suggestion to take into account Hewitt (2005). Hewitt (2005) mentions evidence against the claim that the Himalayan glaciers' are likely to disappear:

> In the late 1990s widespread evidence of glacier expansion was found in the central Karakoram, in contrast to a worldwide decline of mountain glaciers. The expansions were almost exclusively in glacier basins from the highest parts of the range and developed quickly after decades of decline. Exceptional numbers of glacier surges were also reported. (Hewitt 2005, p. 332)

The IPCC authors should not have neglected such evidence. Moreover, in light of this evidence, they should have deleted the prediction that the Himalayan glaciers are likely to disappear by 2035, since it was not supported by the available scientific evidence.

So the following epistemic standard s_2 was violated: IPCC authors should adequately deal with the comments of reviewers, in the sense that (1) if a reviewer points at a flaw (e.g., a contradiction) in the draft report, then this flaw should be removed from the report, and (2) if a reviewer points at evidence against a certain claim in the report, then the authors should take this evidence into account, and remove the claim if it turns out that this claim is not in accordance with the existing scientific evidence. The audience could legitimately assume that this standard was met because s_2 is a common epistemic standard in the relevant field, and no violations of s_2 were reported. Hence, the fact that s_2 was violated in the process resulting in the Fourth Assessment Report implies that an epistemic standard which the audience could legitimately assume to be met was violated in this process. This is the second way in which the epistemic integrity of this process was compromised.

It is possible that the violations of s_1 and s_2 were (perhaps partly) caused by certain non-epistemic interests. Consider two different non-epistemic interests that could have contributed to these violations. The first is the interest to save time. Most scientists work under time pressure, and this gives them an interest in saving time. This interest is served by skipping certain parts of the research process. This could explain why IPCC authors have not checked whether the prediction that the Himalayan glaciers will disappear by 2035 was supported by genuine scientific evidence, and why they simply adopted it from WWF (2005) without critically assessing this source. The interest to save time could also explain why the IPCC authors have not adequately dealt with the reviewers' comments. Because adhering to the epistemic standards of critically assessing one's sources and adequately dealing with reviewers' comments takes time, and because IPCC authors have an interest in saving time, these epistemic standards were not adhered to.

But there is also a second non-epistemic interest that may have played in the violation of these standards. IPCC authors are, in light of their expertise, very well aware of climate change and its disastrous effects. At the same time, they are confronted with the lack of effective political action to mitigate climate change. As a consequence, they may have the non-epistemic interest to do something about this, that is, to stimulate such political action. This political interest is served by clear signals that climate change will have disastrous consequences in the near future. The claim that the Himalayan glaciers are likely to disappear by 2035 can function as such a signal.

It is possible that this contributed to the fact that the IPCC authors were not eager to critically assess the source of the Himalayan glacier prediction, nor to look into evidence against it (as one of the reviewers suggested; see table 9.1). Such actions involved a risk, namely, the risk of finding that this prediction should actually not be in the report. But including it would serve the interest to promote effective political action, as we have just seen. This may be part of the reason why the epistemic standards of critically assessing one's sources and adequately dealing with reviewers' comments were not adhered to.

In summary, we propose the following possible explanation: the IPCC process resulting in the Himalayan glacier flaw violated certain epistemic standards which the audience could legitimately assume to be met because these violations served the IPCC authors' interests to save time and to stimulate effective political action. If this explanation is correct, then certain non-epistemic interests of the IPCC authors compromised the epistemic integrity of the research process. Such an influence of non-epistemic interests on the research process is epistemically problematic.

It should be mentioned that the IPCC restored the epistemic integrity of the process under consideration after a while. In January 2010, almost two and a half years after the publication of the Himalayan glacier flaw in Cruz et al. (2007), the IPCC publicly admitted that "the clear and well-established standards of evidence, required by the IPCC procedures, were not applied properly" (IPCC 2010) in the process leading to the Himalayan glacier flaw. This official apology made clear to the audience that standards like s_1 and s_2 were violated in the process under consideration. Accordingly, it was no longer legitimate for the audience to assume that s_1 and s_2 were met in this process. This means that the process that resulted in the Himalayan glacier flaw no longer violated an epistemic standard which the audience could legitimately assume to be met in it. Hence, the epistemic integrity of this process was restored after the IPCC's public apology.[7]

9.4 Advantages of our analysis

To conclude this chapter, let us point out some advantages of our analysis. What is new about this analysis is that existing controversies concerning the IPCC have been critically examined on the basis of the clearly explicated concept of epistemic integrity. This examination shows how the concept of epistemic integrity can be used to elaborate charges against the IPCC. Our discussion of the controversy about the Himalayan glacier flaw illustrates how critics of the IPCC can use this concept to work out arguments for the view that the IPCC process has been corrupted by non-epistemic interests. But the concept of epistemic integrity is also useful for countering the critics' attacks. We have seen how it can be used to argue (1) that the IPCC process underlying the Second Assessment Report was not influenced by political interests in a way that is epistemically unacceptable, and (2) that the epistemic integrity of the process resulting in the Himalayan glacier flaw in the Fourth Assessment Report was restored after a while.

The main virtue of using our concept of epistemic integrity in such discussions is that it enables us to make a clear distinction between interest influences which are epistemically acceptable and interest influences which are epistemically *un*acceptable. Given the political relevance of the IPCC's work, it seems naïve to believe that the political interests of IPCC authors and reviewers do not affect this work in any way. The expected political consequences of different possible formulations of the same idea may for instance be considered in deciding which specific formulation to put in the IPCC report. This is not necessarily problematic. What is needed then, are criteria that enable us to draw the line between acceptable and unacceptable interest influences. The concept of epistemic integrity can function as such a criterion, as has been illustrated in this chapter. Earlier discussions of the IPCC processes considered in this chapter[8] do not

[7] This does not imply that the process under consideration suddenly became genuinely scientific. One could still argue that the process was unscientific; we only claim that its epistemic integrity was restored. The question whether or not the process is worthy of the term 'scientific' is a different issue.
[8] See, e.g., Oreskes & Conway (2010) for a discussion of the controversy concerning the Second Assessment Report. For discussions of the process resulting in the Himalayan glacier flaw, see, e.g., Cogley et al. (2010), InterAcademy Council (2010), and Ravindranath (2010).

provide such a clear-cut criterion to distinguish epistemically acceptable interest influences from epistemically unacceptable interest influences.

Chapter 10
Conclusion

Let me now return to the two challenges for philosophy of science that were presented at the beginning of this dissertation: (1) to describe and analyze the different roles that different kinds of interests play in science, and (2) to distinguish acceptable interest influences in science from unacceptable interest influences in science (see section 1.3). The purpose of this dissertation was to address these challenges. In this concluding chapter, I recapitulate how I have exactly contributed to tackling the two challenges, and some lines of further research are proposed.

10.1 The first challenge

The first challenge was to study the various influences of interests in the scientific process. My contribution consisted of a discussion of the influences of different kinds of interests in reasoning processes within scientific research (see part 1). More specifically, I discussed interest influences on hypothesis acceptance, theory choice, and the scientific explanation of phenomena. The following interest influences were identified. Firstly, non-epistemic interests affect whether or not a certain hypothesis is accepted by influencing whether or not the evidence in favor of the hypothesis is considered sufficiently strong to accept the hypothesis. Secondly, non-epistemic interests can affect theory choice by influencing (1) which values are used to assess theories, (2) how these values are interpreted, and (3) how they are weighed against each other. Thirdly, epistemic and non-epistemic interests affect the construction and evaluation of scientific explanations. We have seen that the adequacy of an explanation depends on which particular interests (including non-epistemic interests) it has to serve. Because scientists want to end up with the most adequate explanation, they take these interests into account in developing and evaluating explanations.

On the basis of these interest influences, I attacked a view which many non-philosophers seem to hold, and which, I think, still underlies the approach of many philosophers today (even though they may reject it in its original form): the ideal of purity. I distinguished two versions of this ideal: the full ideal of purity and the partial ideal of purity. The former states that scientific decisions such as hypothesis acceptance and theory choice should not be affected by non-epistemic interests. I attacked this view by arguing (1) that non-epistemic interests should be considered in decisions on whether or not to accept a hypothesis, and (2) that the requirement that theory choice is unaffected by non-epistemic interests may be too demanding.

My discussion of interest influences in the construction and evaluation of scientific explanations served to undermine the *partial* ideal of purity, which consists of the following theses:

(1) An interest *indirectly* influences a decision if it is used to determine whether the available evidence is sufficient to make that decision.
(2) Indirect influences of non-epistemic interests on scientific decisions are harmless (and even desirable).
(3) An interest *directly* influences a decision if it is used as a stand-alone reason to motivate the decision.
(4) Direct influences of non-epistemic interests on decisions concerning the characterization of data, the interpretation of evidence, or theory acceptance are unacceptable in science.
(5) Such influences can and must be eliminated.

By arguing that direct influences (as characterized by (3)) of non-epistemic interests on the acceptance of explanations should be allowed in science, I challenged theses (4) and (5) of the partial ideal of purity.

One of the main purposes of part 1 was to convince people to fully abandon the ideal of purity (rather than to try to save parts of it), and to accept that non-epistemic interests form an integral part of scientific reasoning. I hope that part 1 was sufficient for that, but if not, then further research on interest influences in science may be helpful. I wish to encourage researchers to find more examples of the different kinds of interest influences identified in this dissertation, as well as to spell out and document other kinds of interest influences in scientific research processes. This is not only useful to get more people to understand the ubiquity and importance of non-epistemic interests in scientific research, but also to acquire a more comprehensive view of interests in science.

10.2 The second challenge

The fact that I reject the ideal of purity does not imply that all possible influences of non-epistemic interests on scientific research processes are therefore acceptable in my view. Sometimes, such influences are *un*acceptable. Since the ideal of purity does not enable us to adequately account for such problematic interest influences, new criteria are needed for this. That was the second challenge for philosophy of science I presented: to develop criteria that enable us to (1) identify problematic interest influences in science, and (2) explain *why* these interest influences are problematic.

In part 2 of this dissertation, one such criterion was offered, namely, a criterion to account for *epistemically* problematic interest influences. More specifically, I introduced a new concept of epistemic integrity that can function as such a criterion. The underlying idea was that if an interest causes the epistemic integrity of the research process to be lower, then its particular influence on this research process is epistemically problematic.

The degree of epistemic integrity of the research process was defined as the degree to which the research process lives up to the epistemic standards that the audience can legitimately assume to be met in it. This definition enables us to identify epistemically problematic interest influences. More specifically, it is the basis for the following algorithm for finding such interest influences. First, one checks, on the basis of what is reported to the audience, which epistemic standards the audience can legitimately assume to be met in the research process. The second step is to investigate whether the research process actually adhered to these epistemic standards. Where one discovers that this is not the case, one searches for an explanation, which is the third step. If it turns out that non-adherence to the epistemic standards identified in the first step is due to the influence of certain interests, then the conclusion is that these interests had an epistemically problematic influence on the research process.

I have applied this algorithm to different research processes in aerospace science and climate science. For both areas, I was able to specify epistemically problematic interest influences on research processes. My analyses show that the concept of epistemic integrity is useful to (1) identify epistemically problematic interest influences in science, and (2) explain *why* these interest influences are epistemically problematic. Thus, the concept of epistemic integrity is a valuable tool for dealing with the second challenge for philosophy of science I presented.

It should be noted that the concept is not only useful for philosophers of science, but also for scientists themselves, and for science policy-makers. Scientists can use it as an ideal that their research processes should live up to, in order to make sure that their work is not epistemically problematic. My advice to scientists is to communicate most clearly about which epistemic standards their research processes meet, and which not.

In that way, they avoid that the audience can legitimately hold false beliefs about the epistemic standards that are met. By avoiding this, the epistemic integrity of the research process is protected. Science policy-makers can use the concept of epistemic integrity to outline and assess different policy options. They can use it to detect epistemic problems, and where such problems occur, strategies to deal with these problems can be developed. Different strategies can be implemented, and the concept of epistemic integrity makes it possible to examine how each affects the epistemic integrity of the relevant research processes. The more a strategy promotes epistemic integrity, the better it is.

A lot more research remains to be done from here on. Firstly, I wish to encourage research in the vein of chapters 8-9 of this dissertation. It is valuable to analyze research processes in a variety of scientific fields in terms of epistemic integrity. Such analyses help us to discover new epistemic problems, and to understand the exact nature of the epistemic problems that earlier analyses have hinted at. This is crucial to adequately deal with these problems.

Secondly, part 2 of this dissertation focused on interest influences that are epistemically problematic, but interest influences can also be problematic for non-epistemic reasons. An interesting article on this matter is De Winter & Kosolosky (forthcoming). In that article, different ethical theories (utilitarianism, John Rawls's theory of justice, Thomas Pogge's rights-based account of minimal justice, Philip Kitcher's ethical theory, and classical liberalism) are used to argue that the research agendas in the health sciences and the agricultural sciences are morally problematic, and these moral failures are explained by the influence of commercial interests on agenda-setting in these fields. Accordingly, the influence of these interests can be regarded as morally problematic. This shows that the concept of epistemic integrity is not the only criterion to identify problematic interest influences in science. Further research is needed to acquire a comprehensive picture of the different criteria that can be used to assess interest influences in science. Only if such research is carried out will we be able to develop a clear view on the distinction between acceptable and unacceptable interest influences in science (= the second challenge).

A final topic for further research I wish to mention is the relation between the different kinds of scientific integrity. In chapter 6, four kinds of scientific integrity were distinguished: (1) the moral integrity of scientists and their institutions, (2) the moral integrity of the research process, (3) the epistemic integrity of scientists and their institutions, and (4) the epistemic integrity of the research process (see section 6.1). Only the latter kind of scientific integrity was defined in this dissertation. But I think it is interesting to also search for adequate definitions with respect to the three other kinds, and to clarify the relations between the four different kinds of scientific integrity on the basis of these definitions. This will result in a comprehensive account of scientific integrity which is, I think, most welcome in light of (1) the widely-shared concern for scientific integrity, and (2) the remaining confusion surrounding this concept.

Bibliography

Baehr, J. (2011). *The inquiring mind: On intellectual virtues and virtue epistemology*. New York: Oxford University Press.

Bechtel, W., and Richardson, R. (1993). *Discovering complexity: Decomposition and localization as strategies in scientific research*. Princeton: Princeton University Press.

Bekelman, J. E., Li, Y., and Gross, C. P. (2003). Scope and impact of financial conflicts of interest in biomedical research: A systematic review. *Journal of the American Medical Association, 289*(4), 454-465.

Biddle, J. (2007). Lessons from the Vioxx debacle: What the privatization of science can teach us about social epistemology. *Social Epistemology, 21*(1), 21-39.

Brigandt, I. (2013). Explanation in biology: Reduction, pluralism, and explanatory aims. *Science & Education, 22*(1), 69-91.

Bush, V. (1945). *Science: The endless frontier*. A report to the President by Vannevar Bush, Director of the Office of Scientific Research and Development, July 1945. Washington, DC: United States Government Printing Office.
http://www.nsf.gov/about/history/vbush1945.htm (accessed November 15, 2012).

Carnap, R. (1950). *Logical foundations of probability*. London: Routledge and Keegan Paul.

Chalmers, I. (1990). Underreporting research is scientific misconduct. *Journal of the American Medical Association, 263*(10), 1405-1408.

Charles, M. T. (1989). The last flight of Space Shuttle Challenger. In U. Rosenthal, M. T. Charles, and P. T. Hart (Eds.), *Coping with crises: The management of disaster, riots and terrorism* (pp. 110-122). Springfield: Charles C. Thomas Publisher Ltd.

Cogley, J. G., Kargel, J. S., Kaser, G., and van der Veen, C. J. (2010). Tracking the source of glacier misinformation. *Science, 327*(5965), 522.

Committee on Assessing Integrity in Research Environments (2002). *Integrity in scientific research: Creating an environment that promotes responsible conduct*. Washington, DC: National Academies Press.

Cooper, N. (1994). The intellectual virtues. *Philosophy, 69*(270), 459-469.

Cruz, R.V., Harasawa, H., Lal, M., Wu, S., Anokhin, Y., Punsalmaa, B., Honda, Y., Jafari, M., Li, C., and Huu Ninh, N. (2007). Asia. In M. L. Parry, O. F. Canziani, J. P. Palutikof, P. J.

van der Linden, and C. E. Hanson (Eds.), *Change 2007: Impacts, adaptation and vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 459-506). Cambridge: Cambridge University Press.

Cummins, R. (1975). Functional analysis. *Journal of Philosophy, 72*(20), 741-765.

De Ridder, J. (2006). Mechanistic artefact explanation. *Studies in History and Philosophy of Science, 37*(1), 81-96.

De Ridder, J. (2007). *Reconstructing design, explaining artifacts*. Ph.D. Dissertation, Delft University of Technology.

De Schutter, O. (2011). *Agroecology and the right to food*. Report presented at the 16th Session of the United Nations Human Rights Council. http://www.srfood.org/images/stories/pdf/officialreports/20110308_a-hrc-16-49_agroecology_en.pdf (accessed October 5, 2012).

De Vreese, L., Weber, E., and Van Bouwel, J. (2010). Explanatory pluralism in the medical sciences: Theory and practice. *Theoretical Medicine and Bioethics, 31*(5), 371-390.

De Winter, J. (2010). Explanations in software engineering: The pragmatic point of view. *Minds and Machines, 20*(2), 277-289.

De Winter, J. (2011). A pragmatic account of mechanistic artifact explanation. *Studies in History and Philosophy of Science, 42*(4), 602-609.

De Winter, J., and Kosolosky, L. (2013a). The epistemic integrity of NASA practices in the Space Shuttle Program. *Accountability in Research: Policies and Quality Assurance, 20*(2), 72-92.

De Winter, J., and Kosolosky, L. (2013b). The epistemic integrity of scientific research. *Science and Engineering Ethics, 19*(3), 757-774.

De Winter, J., and Kosolosky, L. (forthcoming). Health, food, and science: An ethical assessment of research agendas. *Logique et Analyse*.

Douglas, H. (2009). *Science, policy, and the value-free ideal*. Pittsburgh: University of Pittsburgh Press.

Duhem, P. (1954). *The aim and structure of physical theory* (P. W. Wiener, Trans.). Princeton: Princeton University Press. (Original work published 1914).

Edwards, P. N., and Schneider, S. H. (1997). The 1995 IPCC report: Broad consensus or "scientific cleansing"? *Ecofable/Ecoscience, 1*(1), 3-9.

Feynman, R. (1986). Personal observations on the reliability of the Shuttle. In Presidential Commission, *Report of the Presidential Commission on the Space Shuttle Challenger accident* (Vol. 2, Appendix F). Washington, DC: United States Government Printing Office.

Gervais, R., and Weber, E. (2013). Inferential explanations in biology. *Studies in History and Philosophy of Biological and Biomedical Sciences, 44*(3), 356-364.

Glennan, S. (2002). Rethinking mechanistic explanation. *Philosophy of Science, 69*(S3), S342-S353.

Gøtzsche, P. C., Hróbjartsson, A., Johansen, H. K., Haahr, M. T., Altman, D. G., and Chan, A. (2007). Ghost authorship in industry-initiated randomised trials. *PloS Medicine, 4*(1), 47-52.

Greco, J., and Turri, J. (2013). Virtue Epistemology. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (winter 2013 Edition). http://plato.stanford.edu/entries/epistemology-virtue/ (accessed October 31, 2013).

Hall, J. L. (2003). Columbia and Challenger: Organizational failure at NASA. *Space Policy, 19*(4), 239-247.

Hansson, S. O. (2007). What is technological science? *Studies in History and Philosophy of Science, 38*(3), 523-527.

Hewitt, K. (2005). The Karakoram anomaly? Glacier expansion and the 'elevation effect', Karakoram Himalaya. *Mountain Research and Development, 25*(4), 332-340.

House Committee (1986). *Investigation of the Challenger accident: Report of the Committee on Science and Technology, House of Representatives, Ninety-ninth Congress, second session.* Washington, DC: United States Government Printing Office. http://www.gpo.gov/fdsys/pkg/GPO-CRPT-99hrpt1016/pdf/GPO-CRPT-99hrpt1016.pdf (accessed September 25, 2012).

InterAcademy Council (2010). *Climate change assessments: Review of the processes and procedures of the IPCC*. Alkmaar: Bejo Druk & Print. http://reviewipcc.interacademycouncil.net/report/Climate%20Change%20Assessments,%20Review%20of%20the%20Processes%20&%20Procedures%20of%20the%20IPCC.pdf (accessed January 15, 2014).

IPCC (1995). *IPCC Second Assessment Report: Climate change 1995*. Cambridge: Cambridge University Press.

IPCC (2007). *IPCC Fourth Assessment Report: Climate change 2007*. Cambridge: Cambridge University Press.

IPCC (2008). Procedures for the preparation, review, acceptance, adoption, approval, and publication of IPCC reports. http://www.ipcc.ch/pdf/ipcc-principles/ipcc-principles-appendix-a.pdf (accessed January 20, 2014).

IPCC (2010). IPCC statement on the melting of Himalayan glaciers. http://www.ipcc.ch/pdf/presentations/himalaya-statement-20january2010.pdf (accessed January 20, 2014).

IPCC (2013). Procedures for the preparation, review, acceptance, adoption, approval, and publication of IPCC reports. http://www.ipcc.ch/pdf/ipcc-principles/ipcc-principles-appendix-a-final.pdf (accessed January 15, 2014).

Jackson, F., and Pettit, P. (1992). In defense of explanatory ecumenism. *Economics and Philosophy, 8*(1), 1-21.

Kasulis, T. P. (2002). *Intimacy or integrity: Philosophy and cultural difference* (*The 1998 Gilbert Ryle Lectures*). Honolulu: University of Hawai'i Press.

Kitcher, P. (2001). *Science, truth, and democracy*. New York: Oxford University Press.

Kosolosky, L. (forthcoming). "Peer review is melting our glaciers": What led the Intergovernmental Panel on Climate Change (IPCC) to go astray? *Journal for General Philosophy of Science*.

Kotlyakov, V. M. (Ed.) (1996). *Variations of snow and ice in the past and at present on a global and regional scale*. Paris: Unesco.

Kourany, J. A. (2010). *Philosophy of science after feminism*. New York: Oxford University Press.

Krimsky, S. (2003). *Science in the private interest: Has the lure of profits corrupted biomedical research?* Lanham: Rowman & Littlefield Publishers.

Kuhn, T. S. (1977). *The essential tension: Selected studies in scientific tradition and change*. Chicago: The University of Chicago Press.

Laudan, L. (1984). *Science and values: The aims of science and their role in scientific debate*. London: University of California Press.

Laudan, L. (2004). The epistemic, the cognitive, and the social. In P. Machamer and G. Wolters (Eds.), *Science, values, and objectivity* (pp. 14-23). Pittsburgh: University of Pittsburgh Press.

Lichtenberg, F. R. (2003). Pharmaceutical innovation, mortality reduction, and economic growth. In K. M. Murphy and R. H. Topel (Eds.), *Measuring the gains from medical research: An economic approach* (pp. 41-73). London: The University of Chicago Press.

Loder, R. E. (2002). Integrity and epistemic passion. *Notre Dame Law Review, 77*(3), 841-888.

Longino, H. E. (1983). Beyond "bad science": Skeptical reflections on the value-freedom of scientific inquiry. *Science, Technology, & Human Values, 8*(1), 7-17.

Longino, H. E. (1990). *Science as social knowledge: Values and objectivity in scientific inquiry*. Princeton: Princeton University Press.

Longino, H. E. (1996). Cognitive and non-cognitive values in science: Rethinking the dichotomy. In L. H. Nelson and J. Nelson (Eds.), *Feminism, science, and the philosophy of science* (pp. 39-58). Dordrecht: Kluwer.

Mackie, J. (1974). *The cement of the universe: A study of causation*. Oxford: Clarendon Press.

Magnus, P. D. (2005). Hormone research as an exemplar of underdetermination. *Studies in History and Philosophy of Biological and Biomedical Sciences, 36*(3), 559-567.

Mahon, J. E. (2008). The definition of lying and deception. In E. N. Zelta (Ed.), *The Stanford encyclopedia of philosophy* (fall 2008 edition). http://plato.stanford.edu/archives/fall2008/entries/lying-definition (accessed November 7, 2013).

Mäki, U. (2002). Symposium on explanations and social ontology 2: Explanatory ecumenism and economics imperialism. *Economics and Philosophy, 18*(2), 235-257.

Maxwell, N. (1974). The rationality of scientific discovery part I: The traditional rationality problem. *Philosophy of Science, 41*(2), 123-153.

Maxwell, N. (2010). Review – Science, policy, and the value-free ideal (by Heather Douglas). *Metapsychology: online reviews, 14*(10). http://metapsychology.mentalhelp.net/poc/view_doc.php?type=book&id=5434&cn=394 (accessed December 2, 2011).

McMullin, E. (1983). Values in science. In P. Asquith and T. Nickles (Eds.), *Proceedings of the 1982 Biennial Meeting of the Philosophy of Science Association* (pp. 3-28). East Lansing: Philosophy of Science Association.

McMullin, E. (1984). A case for scientific realism. In J. Leplin (Ed.), *Scientific realism* (pp. 8-40). London: University of California Press.

Melander, H., Ahlqvist-Rastad, J., Meijer, G., and Beermann, B. (2003). Evidence b(i)ased medicine – selective reporting from studies sponsored by pharmaceutical industry: Review of studies in new drug applications. *British Medical Journal, 326*(7400), 1171-1173.

Mowery, D. C., and Sampat, B. N. (2005). The Bayh-Dole Act of 1980 and university-industry technology transfer: A model for other OECD governments? *Journal of Technology Transfer, 30*(1/2), 115-127.

Office of Science and Technology Policy (2000). Federal policy on research misconduct; Preamble for research misconduct policy. *Federal Register, 65*(235), 76260–76264.

Oreskes, N., and Conway, E. M. (2010). *Merchants of doubt: How a handful of scientists obscured the truth on issues from tobacco smoke to global warming*. New York: Bloomsbury Press.

Pettit, P. (1995). The virtual reality of 'Homo Economicus'. *The Monist, 78*(3), 308-329.

Pettit, P. (2000). Rational choice, functional selection and empty black boxes. Journal of *Economic Methodology, 7*(1), 33–57.

Pitt, J. C. (2009). Technological explanation. In A. W. M. Meijers, D. M. Gabbay, P. Thagard, and J. Woods (Eds.), *Philosophy of technology and engineering sciences* (pp. 861-879). Amsterdam: Elsevier.

Presidential Commission (1986). *Report of the Presidential Commission on the Space Shuttle Challenger accident*. Washington, DC: United States Government Printing Office. http://history.nasa.gov/rogersrep/genindex.htm (accessed September 25, 2012).

Proctor, R. N. (1991). *Value-free science? Purity and power of modern knowledge*. Cambridge: Harvard University Press.

PSRCR (1992). *Responsible science, volume I: Ensuring the integrity of the research process*. Washington, DC: National Academies Press.

Quine, W. V. O. (1951). Two dogmas of empiricism. *The Philosophical Review, 60*(1), 20-43.

Radu, S. (2008). *The Fisher-Yates shuffle algorithm*. Paper presented at the Mini Conference on Computing Algorithms, Bryn Mawr.

Ravindranath, N. H. (2010) IPCC: Accomplishments, controversies and challenges. *Current Science, 99*(1), 26-35.

Ray, W. A., Stein, C. M., Hall, K., Arbogast, P. G., Daugherty, J. R., and Griffin., M. R. (2002). Non-steroidal anti-inflammatory drugs and risk of serious coronary heart disease: An observational cohort study. *The Lancet, 359*(9301), 118-123.

Reiss, J. (2010). In favour of a Millian proposal to reform biomedical research. *Synthese, 177*(3), 427-447.

Resnik, D. B. (1996). Social epistemology and the ethics of research. *Studies in History and Philosophy of Science Part A, 27*(4), 565-586.

Resnik, D. B. (1998). *The ethics of science: An introduction*. London: Routledge.

Rooney, P. (1992). On values in science: Is the epistemic/non-epistemic distinction useful? In D. Hull, M. Forbes, & K. Okruhlik (Eds.), *Proceedings of the 1992 biennial meeting of the Philosophy of Science Association* (Vol. 2, pp. 13-22). East Lansing: Philosophy of Science Association.

Ross, J. S., Hill, K. P., Egilman, D. S., and Krumholz, H. M. (2008). Guest authorship and ghostwriting in publications related to rofecoxib. A case study of industry documents from rofecoxib litigation. *Journal of the American Medical Association, 299*(15), 1800-1812.

Rowe, S., Alexander, N., Clydesdale, F. M., Applebaum, R. S., Atkinson, S., Black, R. M., Dwyer, J. T., Hentges, E., Higley, N. A., Lefevre, M., Lupton, J. R., Miller, S. A., Tancredi, D. L., Weaver, C. M., Woteki, C. E., and Wedral, E. (2009). Funding food science and nutrition research: Financial conflicts and scientific integrity. *The American Journal of Clinical Nutrition, 89*(5), 1285-1291.

Rudner, R. (1953). The scientist *qua* scientist makes value judgments. *Philosophy of Science, 20*(1), 1-6.

Ruphy, S. (2006). "Empiricism all the way down": A defense of the value-neutrality of science in response to Helen Longino's contextual empiricism. *Perspectives on Science, 14*(2), 189-214.

Salmon, W. C. (2005). *Reality and rationality.* New York: Oxford University Press.

Santer, B. D. (1996). Letters to the editor: No deception in global warming report. *Wall Street Journal*, June 25.

Santer, B. D., Wigley, T. M. L., Barnett, T. P., and Anyamba, E. (1995). Detection of climate change and attribution of causes. In J. T. Houghton, L. G. Meira Filho, B. A. Callander, N. Harris, A. Kattenberg, and K. Maskell (Eds.), *Climate change 1995: The science of climate change. Contribution of Working Group I to the Second Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 407-443). Cambridge: Cambridge University Press.

Satz, D., and Ferejohn, J. (1994). Rational choice and social theory. *Journal of Philosophy, 91*(2), 71-87.

Schachman, H. K. (2006). From "publish or perish" to "patent and prosper". *The Journal of Biological Chemistry, 281*(11), 6889-6903.

Scherkoske, G. (2012). Could integrity be an epistemic virtue? *International Journal of Philosophical Studies,* 20(2), 185–215.

Schiermeier, Q. (2010). Glacier estimate is on thin ice: IPCC may modify its Himalayan melting forecasts. *Nature, 463*(7279), 276-277.

Schott, G., Pachl, H., Limbach, U., Gundert-Remy, U., Lieb, K., and Ludwig, W. (2010). The financing of drug trials by pharmaceutical companies and its consequences: Part 2: A qualitative, systematic review of the literature on possible influences on authorship, access to trial data, and trial registration and publication. *Deutsches Ärzteblatt International, 107*(17), 295-301.

Seitz, F. (1996). A major deception on 'global warming'. *Wall Street Journal*, June 12.

Shamoo, A. E., and Resnik, D. B. (2009). *Responsible conduct of research* (second edition). Oxford: Oxford University Press.

Shrader-Frechette, K. S. (1994). *Ethics of scientific research.* London: Rowman & Littlefield Publishers.

Siepmann, T. J. (2004). The global exportation of the U.S. Bayh-Dole Act. *University of Dayton Law Review, 30*(2), 209-244.

Stanford, K. (2009). Underdetermination of scientific theory. In E. N. Zelta (Ed.), *The Stanford encyclopedia of philosophy* (winter 2009 edition). http://plato.stanford.edu/archives/win2009/entries/scientific-underdetermination (accessed June 20, 2013).

Steel, D. (2004). Can a reductionist be a pluralist? *Biology and Philosophy, 19*(1), 55-73.

Steyn, M. (2010). Credibility is what's really melting. *Maclean's*, February 3. http://www2.macleans.ca/2010/02/03/credibility-is-what-is-really-melting/ (accessed January 16, 2014).

Steneck, N. H. (2002). Assessing the integrity of publicly funded research. In N. H. Steneck and M. D. Scheetz (Eds.), *Investigating research integrity: Proceedings of the First ORI Research Conference on Research Integrity* (pp. 1-16). http://ori.hhs.gov/documents/proceedings_rri.pdf (accessed June 10, 2014).

Steneck, N. H. (2006). Fostering integrity in research: Definitions, current knowledge, and future directions. *Science and Engineering Ethics, 12*(1), 53-74.

Tedre, M. (2011). Computing as a science: A survey of competing viewpoints. *Minds and Machines, 21*(3), 361-387.

Tiles, M. (1987). A science of Mars or of Venus? *Philosophy, 62*(241), 293-306.

Van Bouwel, J., and Weber, E. (2002). Remote causes, bad explanations. *Journal for the Theory of Social Behaviour, 32*(4), 437-449.

Van Bouwel, J., and Weber, E. (2008). A pragmatist defense of non-relativistic explanatory pluralism in history and social science. *History & Theory, 47*(2), 168-182.

Van Fraassen, B. C. (1980). *The scientific image*. Oxford: Clarendon Press.

Van Fraassen, B. C. (1995). Belief and the problem of Ulysses and the sirens. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition, 77*(1), 7-37.

Vaughan, D. (1996). *The Challenger launch decision: Risky technology, culture, and deviance at NASA*. Chicago: University of Chicago Press.

Weber, E., Van Bouwel, J., and De Vreese, L. (2013). *Scientific explanation*. Dordrecht: Springer.

Winsor, D. A. (1988). Communication failures contributing to the Challenger accident: An example for technical communicators. *IEEE Transactions on Professional Communication, 31*(3), 101-107.

WWF (2005). An overview of glaciers, glacier retreat, and subsequent impacts in Nepal, India and China. http://www.wwf.or.jp/activities/lib/pdf_climate/environment/Overview_of_Glaciers.pdf (accessed January 16, 2014).

Zagzebski, L. T. (1996). *Virtues of the mind: An inquiry into the nature of virtue and the ethical foundations of knowledge*. New York: Cambridge University Press.

Summary in English

Interests and epistemic integrity in science

A new framework to assess interest influences in scientific research processes

This dissertation starts from three assumptions. The first is that science serves a variety of interests: epistemic interests (i.e. interests in knowing something), social interests, financial interests, career interests, etc. The second assumption is that interests influence science. The idea is that serving interests is not just an ad hoc aspect of science, but that, from the beginning of the scientific process, serving certain interests is anticipated. For instance, a scientist pursues a certain line of research because he expects it to serve his epistemic interests, his career interests, and/or certain social interests. The third assumption is that some interest influences in science are unacceptable; e.g., it is unacceptable for a pharmaceutical company to falsify clinical trial data in order to serve the company's commercial interests.

These assumptions present two challenges for philosophy of science. The first is to describe and analyze the different roles that different kinds of interests play in science. The second challenge is to distinguish acceptable from unacceptable interest influences in science. Part 1 of this dissertation contributes to tackling the first challenge, and part 2 contributes to tackling the second challenge.

1. The roles of interests within scientific research

My contribution with respect to the first challenge consists of a discussion of the roles of different kinds of interests in reasoning processes within scientific research. More specifically, I discuss interest influences on hypothesis acceptance, theory choice, and the scientific explanation of phenomena. The following interest influences are identified. Firstly, non-epistemic interests affect whether or not a certain hypothesis is accepted by influencing whether or not the evidence in favor of the hypothesis is considered sufficiently strong to accept the hypothesis. Secondly, non-epistemic interests can affect theory choice by influencing (1) which values are used to assess theories, (2) how these values are interpreted, and (3) how they are weighed against each other. Thirdly, epistemic and non-epistemic interests affect the construction and evaluation of scientific explanations. Because scientists want to end up with the most adequate explanation, and because the adequacy of an explanation depends on which particular interests (including non-epistemic interests) this explanation has to serve, scientists take these interests into account in developing and evaluating explanations.

On the basis of these interest influences, I attack a view which many non-philosophers seem to hold, and which, I think, still underlies the approach of many philosophers today (even though they may reject it in its original form): the ideal of purity. I distinguish two versions of this ideal: (1) the full ideal of purity, which states that ideally, non-epistemic interests do not play a role in scientific decisions such as hypothesis acceptance and theory choice, and (2) the partial ideal of purity, which states that ideally, non-epistemic interests do not play a *direct* role in such scientific decisions. I attack the full ideal of purity by arguing (1) that non-epistemic interests should be considered in decisions on whether or not to accept a hypothesis, and (2) that the requirement that theory choice is unaffected by non-epistemic interests may be too demanding. My discussion of interest influences in the construction and evaluation of scientific explanations serves to undermine the partial ideal of purity. I criticize this ideal by arguing that direct influences of non-epistemic interests on the acceptance of explanations should be allowed in science. The general conclusion is that non-epistemic interests form an integral part of reasoning processes within scientific research.

2. Epistemic integrity

But when is an interest influence in science problematic, then? Since the ideal of purity is not a good basis for distinguishing acceptable from unacceptable interest influences in science, new criteria are needed for this. In part 2, one such criterion is offered, namely, a criterion to identify epistemically unacceptable interest influences. More specifically, I introduce a new concept of epistemic integrity that can function as such a criterion. The underlying idea is that if an interest compromises the epistemic integrity of the research process, then its particular influence on this research process is epistemically unacceptable.

The degree of epistemic integrity of the research process is defined as the degree to which the research process lives up to the epistemic standards that the audience can legitimately assume to be met in it. After explaining this definition, I show that it avoids the shortcomings of other definitions. Furthermore, it is argued that my concept of epistemic integrity meets Rudolf Carnap's four requirements of explication: (1) it is similar to our common sense notion of epistemic integrity of the research process, (2) it is exact, (3) it is fruitful, that is, useful for the development and justification of norms for science and for the analysis of cases from scientific practice, and (4) it is as simple as requirements (1), (2), and (3) permit.

To illustrate the usefulness of this concept for identifying epistemically unacceptable interest influences in science, it is applied to several cases from different areas of research. Firstly, I demonstrate that the epistemic integrity of different kinds of research processes in the U.S. Space Shuttle Program was damaged. This damage is explained by the financial and career interests of those involved in the research processes; these interests compromised epistemic integrity. Such an interest influence is epistemically unacceptable. Secondly, I discuss the epistemic integrity of research processes in climate science. The focus is on controversies concerning the Intergovernmental Panel on Climate Change (IPCC). The epistemic integrity of the work by the IPCC that has been the target of criticism is assessed. Where epistemic integrity was compromised, this is explained on the basis of the non-epistemic interests of the person(s) involved. This suggests that these interests influenced the relevant research processes in epistemically unacceptable ways.

Summary in Dutch

Belangen en epistemische integriteit in wetenschap

Een nieuw kader voor het beoordelen van belangeninvloeden in wetenschappelijke onderzoeksprocessen

Dit proefschrift vertrekt vanuit drie veronderstellingen. De eerste is dat wetenschap allerlei belangen dient: epistemische belangen (het verlangen om iets te weten), sociale belangen, financiële belangen, carrièrebelangen etc. De tweede veronderstelling is dat belangen wetenschap beïnvloeden. De onderliggende gedachte is dat het dienen van belangen geen ad hoc aspect van wetenschap is; het dienen van bepaalde belangen wordt geanticipeerd vanaf het begin van het wetenschappelijke proces. Een wetenschapper zal bijvoorbeeld voor een bepaalde onderzoekslijn kiezen omdat hij verwacht dat deze onderzoekslijn zijn epistemische belangen, zijn carrièrebelangen en/of bepaalde sociale belangen zal dienen. De derde veronderstelling is dat sommige belangeninvloeden in wetenschap onaanvaardbaar zijn. Het is bijvoorbeeld onaanvaardbaar voor een farmaceutisch bedrijf om data uit klinische studies te vervalsen om de commerciële belangen van het bedrijf te beschermen.

Deze veronderstellingen leiden tot twee uitdagingen voor de wetenschapsfilosofie. De eerste is om de verschillende rollen die verschillende soorten belangen spelen in wetenschap, te beschrijven en te analyseren. De tweede uitdaging is om aanvaardbare van onaanvaardbare belangeninvloeden te onderscheiden. Deel 1 van dit proefschrift draagt bij tot het aanpakken van de eerste uitdaging. Deel 2 draagt bij tot het aanpakken van de tweede uitdaging.

1. De rollen van belangen binnen wetenschappelijk onderzoek

Mijn bijdrage met betrekking tot de eerste uitdaging bestaat uit een bespreking van de rollen die verschillende soorten belangen spelen in redeneerprocessen binnen wetenschappelijk onderzoek. Meer bepaald bespreek ik belangeninvloeden op het aanvaarden van hypothesen, theoriekeuze, en het wetenschappelijk verklaren van fenomenen. De volgende belangeninvloeden worden geïdentificeerd. Ten eerste beïnvloeden niet-epistemische belangen of een hypothese al dan niet aanvaard wordt doordat ze meespelen in de beslissing of het ondersteunende bewijsmateriaal voldoende sterk is. Ten tweede, niet-epistemische belangen kunnen theoriekeuze beïnvloeden doordat ze mee bepalen (1) welke waarden gebruikt worden bij het beoordelen van theorieën, (2) hoe deze waarden geïnterpreteerd worden, en (3) hoe ze tegen elkaar worden afgewogen. Ten derde beïnvloeden epistemische en niet-epistemische belangen de constructie en evaluatie van wetenschappelijke verklaringen. Omdat wetenschappers willen dat hun verklaringen zo adequaat mogelijk zijn, en omdat de adequaatheid van een verklaring afhangt van de specifieke belangen (inclusief niet-epistemische belangen) die de verklaring hoort te dienen, gaan ze dergelijke belangen in rekening brengen bij het ontwikkelen en beoordelen van verklaringen.

Deze belangeninvloeden vormen de basis voor een kritiek op een gedachte die bij veel niet-filosofen aanwezig is, en die ook in de filosofische gemeenschap nog in zekere mate leeft: het ideaal van puurheid. Ik maak een onderscheid tussen twee versies van dit ideaal: (1) het volle ideaal van puurheid, dat stelt dat niet-epistemische belangen idealiter geen rol spelen in wetenschappelijke beslissingen zoals de aanvaarding van hypothesen en theoriekeuze, en (2) het partiële ideaal van puurheid, dat stelt dat niet-epistemische belangen idealiter geen *directe* rol spelen in dergelijke wetenschappelijke beslissingen. Ik bekritiseer het volle ideaal van puurheid door te argumenteren (1) dat niet-epistemische belangen in rekening moeten gebracht worden bij beslissingen over het al dan niet aanvaarden van een hypothese, en (2) dat de verwachting dat theoriekeuze niet beïnvloed wordt door niet-epistemische belangen te veeleisend is. Het doel van mijn bespreking van belangeninvloeden in de constructie en beoordeling van wetenschappelijke verklaringen is om het partiële ideaal van puurheid te ondermijnen. Ik bekritiseer dit ideaal door te argumenteren dat directe invloeden van niet-epistemische belangen op de aanvaarding van verklaringen moet toegelaten worden in wetenschap. De algemene conclusie is dat niet-epistemische belangen integraal deel uitmaken van redeneerprocessen binnen wetenschappelijk onderzoek.

2. Epistemische integriteit

Maar wanneer is een belangeninvloed in wetenschap dan problematisch? Aangezien het ideaal van puurheid geen goede basis vormt voor het onderscheid tussen aanvaardbare en onaanvaardbare belangeninvloeden in wetenschap, is er nood aan nieuwe criteria om dit onderscheid te maken. In deel 2 wordt een dergelijk criterium ontwikkeld, namelijk een criterium om epistemisch onaanvaardbare belangeninvloeden te identificeren. Meer bepaald introduceer ik een nieuw concept van epistemische integriteit dat kan fungeren als een dergelijk criterium. De onderliggende gedachte is dat als een bepaald belang ertoe leidt dat de epistemische integriteit van het onderzoeksproces geschaad is, dan heeft dit belang een epistemisch onaanvaardbare invloed op het onderzoeksproces.

De graad van epistemische integriteit van het onderzoeksproces wordt gedefinieerd als de mate waarin het onderzoeksproces voldoet aan bepaalde epistemische standaarden, namelijk die epistemische standaarden waarvan het publiek op legitieme wijze kan veronderstellen dat het onderzoeksproces eraan voldoet. Deze definitie wordt uitgebreid verduidelijkt, waarna ik aantoon dat ze de tekortkomingen van andere definities vermijdt. Verder argumenteer ik dat mijn concept van epistemische integriteit voldoet aan Rudolf Carnaps vier vereisten van explicatie: (1) het komt min of meer overeen met hoe de term ‘epistemische integriteit van het onderzoeksproces’ doorgaans gebruikt wordt, (2) het is exact, (3) het is vruchtbaar voor de ontwikkeling en rechtvaardiging van normen voor wetenschap en voor de analyse van cases uit de wetenschappelijke praktijk, en (4) het is zo simpel mogelijk (gegeven dat aan vereisten (1), (2) en (3) voldaan moet worden).

Ter illustratie van het nut van dit concept voor het identificeren van epistemisch onaanvaardbare belangeninvloeden in wetenschap, wordt het toegepast op meerdere voorbeelden uit verschillende onderzoeksvelden. Eerst toon ik aan dat de epistemische integriteit van verschillende soorten onderzoeksprocessen in het Amerikaanse Spaceshuttleprogramma geschaad was. Deze schade wordt verklaard op basis van de financiële en carrièrebelangen van de betrokkenen; deze belangen hebben de epistemische integriteit van de betreffende onderzoeksprocessen aangetast. Een dergelijke belangeninvloed is epistemisch onaanvaardbaar. Ten tweede behandel ik onderzoeksprocessen in klimaatwetenschap. Ik richt mij op controverses met betrekking tot het Intergovernmental Panel on Climate Change (IPCC). De epistemische integriteit van het IPCC-werk dat het mikpunt is geweest van kritiek, wordt beoordeeld. Waar epistemische integriteit werd aangetast, verklaar ik dit op basis van de niet-epistemische belangen van de uitvoerder(s) van het betreffende onderzoeksproces. Dit geeft aan dat de specifieke invloeden die deze belangen op de betreffende onderzoeksprocessen hadden, epistemisch onaanvaardbaar zijn.