To Spill, Filter and Clean

On problematic research articles, the peer review system, and organisational integrity procedures

Serge P.J.M. Horbach

To Spill, Filter and Clean

On problematic research articles, the peer review system, and organisational integrity procedures

Serge P.J.M. Horbach

Cover: Marlies Nillesen, Nijmegen

Press: Ridderprint, Alblasserdam

ISBN: 978-94-6375-684-6

To Spill, Filter and Clean

On problematic research articles, the peer review system, and organisational integrity procedures

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,
volgens besluit van het college van decanen
in het openbaar te verdedigen op

donderdag 13 februari 2020

om 16.30 uur precies

door

Serge Pascal Johannes Maria Horbach

geboren op 22 april 1993

te Roermond

Promotor
Prof. dr. H.A.E. Zwart

Copromotor
Dr. W. Halffman

Manuscriptcommissie

Prof. dr. H.W. de Regt	*(voorzitter)*
Prof. dr. S. de Rijcke	*(Universiteit Leiden)*
Prof. dr. B.J.R. van der Meulen	*(Universiteit Twente)*

Paranimfen
Axel Horbach
Marco Stevens

Table of Content

Chapter 1 - Introduction

1

She puts down her cup of tea to ask me full of amazement:
"Are you sure that is what happened?"
"They say it really happened that way, yes."
"But you wouldn't expect that in science, would you?"
"You might not, indeed. But well, apparently, it was not the first time it happened and I guess it won't be the last either."
"How come no one ever noticed this? How did this stay under the radar for so long?"
"That's a good question."
"And what has happened since it finally came to light?"

While doing my research I have had dozens of these conversations. It still amazes me how surprised people are when they hear about misconduct in science. I think this is telling in many ways. And as simple as the questions they typically ask may seem, many are still remarkably difficult to answer.

Science is considered by many one of the major drivers behind progress, development and knowledge in our society. Be it in the form of medical advancement, economic progress, technological developments or social understanding, science contributes to things we value in our society or daily lives. Throughout its history, science contributed to knowledge that found its way to exciting and important applications, as well as to knowledge on its way to do so. Both because of its proven utility to mankind and because of its specific methods and approach, science built a high status of trust as source of reliable knowledge. In this light, it is unsurprising that issues affecting science's trustworthiness or applicability have met with considerable indignation and (public) outrage, including among many of the people that I talked with, such as the one above.

During its history, science has gone through several periods of intense debate over its trustworthiness, methods and goals. Sociologists of science have indicated several conditions leading to hostility towards science (Merton, 1973). These conditions include: when the effects of applying scientific knowledge are deemed undesirable; when the authority of science is restricted by or clashes with political, economic or religious authority; or when the core values of science do not meet with those values in other institutions. Over the past century, major debates and conflicts include indignation over scientific practices under the Nazi regime, controversy over application of science in constructing nuclear weapons and other weapons of mass destruction, fear for uncontrollable consequences of research on genetic modifications, and debate about the status of scientific knowledge in a society of alternative facts and post-truth sentiments. Debates over the integrity and misconduct in science form a relatively recent addition to this list of concerns.

Starting with concerns about and the uncovering of major cases of scientific fraud, science met with considerable outrage during the 1980s. This was further fuelled by prominent cases of science being irreproducible, such as Pons and Fleischman's experiments on cold fusion (Collins and Pinch, 1993). Until the 1980s, the scientific enterprise was largely assumed to be relatively free of misconduct and various forms fraudulent behaviour (Guston, 2007; Horbach and Halffman, 2017). Although concerns regarding the ethics of research and its applications, e.g. in the light of nuclear energy, and debates about the use of science for political purposes, e.g. in the debate on risks of smoking, have been expressed for decades, cases of deliberate misconduct were thought to be rare. In addition, when fraudulent behaviour came to light, this was handled discretely. Allegations were not officially put on trial, but handled by some figure of authority in the administration. Usually, serious care was undertaken to contain the conflict as much as possible (Guston, 2007). This was partly because, in the 1970s, misconduct in science was commonly characterised as an issue of personal moral conduct rather than public policy (LaFollette, 1992). This also becomes clear from the fact that integrity of science was hardly discussed in policy or administrative settings. Even though science obtained massive federal financial support, and the 1978 United States *Science and the Congress* meeting hosted a substantial discussion on the regulation of science, the integrity of science did not appear on the agenda (Guston, 2007).

In order to proceed, we should be a little more precise in what we refer to when discussing the deceptively clear term 'science'. The term science is commonly used to denote a variety of distinct though related activities. Numerous attempts have been made to demarcate what is to count as science and what is not. These attempts either use criteria relating to how scientists *think,* distinguishing science from other forms of knowing by virtue of the systematic way of obtaining knowledge, or how scientists work and cooperate, relating to the social structures and ways of organising knowledge production (Halffman, 2019).

Following Merton, I distinguish science as (1) a set of methods by means of which knowledge is obtained, (2) an archive of accumulated knowledge resulting from the application of these methods, (3) a set of cultural values and traditions governing certain activities of knowledge creation, and (4) any combination of the foregoing (Merton, 1973). Later, other theorists and sociologists of science have come with alternative conceptualisations of science, including Luhmann's definition of science as a closed autopoietic system of communications (Luhmann, 1992), and Actor Network Theory's understanding of science as a seamless web of interaction between scientists, their objects of study and their environments (Callon, 1995). I will revisit some of these understandings of science, but for now it suffices to state that, in this dissertation, I will primarily, though not exclusively, be concerned with the second and third notions of science indicated by Merton and subsequent scholars in Science and Technology Studies (STS). In particular, I will focus on the *practices* constituting these interpretations of science, hence combining Merton's traditional conceptualisation of science with those of more contemporary scholars in STS and practice theory. When discussing the second notion, the archive of knowledge resulting from scientific efforts, I will commonly refer to the 'scientific literature', and study the practices involved in publishing, reviewing and disseminating research articles. With the term 'science' I will commonly refer to the third notion: the cultural structure and organisation as it shapes and is shaped by the practices of scientists in creating, discussing and disseminating knowledge. The main focus will therefore not lie explicitly with the methods of science, but rather with the cultures and practices that structure and facilitate these methods, as well as with the ways of reporting and disseminating its results.

With this understanding of science, I do not aim to set science apart as a fundamentally demarcated activity, distinct from all other social activities. Instead, I acknowledge that demarcating science is in itself problematic and boundaries between what is and what is not to count as science quickly become fuzzy. In addition, science is far from homogeneous, with very distinct practices constituting science in different disciplines and cultures, impacting on how research is conducted, how knowledge is validated and evaluated, and how results are disseminated. I will revisit this issue of diversity multiple times later in this dissertation. Last, in an attempt to clarify what I mean with the term 'science', I should add a note on the well-known difference between the term in the Anglo-Saxon and Continental European traditions. Throughout this dissertation I will follow the latter tradition, hence having the term 'science' encompass all forms of academic knowledge production, rather than restricting it to the natural sciences as is common Anglo-Saxon usage of the term. As such, I might sometimes use the terms 'science' and 'research' interchangeably.

1.1. Research integrity and misconduct

This dissertation studies research integrity as one of the cornerstones of good scientific practice. More specifically, it studies the relationship between current practices in organising and disseminating science and instances of problematic or dubious behaviour. In addition, it studies science and its self-regulating mechanisms, filtering undesirable records from the literature and acting when cases of undesirable behaviour come to light.

Fuelled by large, often spectacular cases of research misconduct, research integrity has come under the attention of scholars, science policymakers and the larger public alike. A substantial amount of scientific literature has emerged over the past decades, numerous policy reports, guidelines and codes of conduct have been written, and newspaper reports about misbehaving scientists appear at an increasingly regular basis. Combined with increasing concerns about the (ir)reproducibility of scientific experiments, about issues with research funding, and about scientific evaluation and reward structures, the global research enterprise is by some considered to be in a status of crisis.

One of the factors feeding this public debate is the still special status that science enjoys in contemporary society. Science is still seen as the primary source of knowledge that helps us move forward through difficult situations and that brings us economic progress, increased comfort and well-being, health or even freedom. In addition, science is one of the levers at a time when speculation, exaggeration and alternative facts run rampant. Last, and not least important, much of the science that has been compromised is paid with public money out of taxpayers' pockets. Even though science is going through harsh periods, scientists themselves are still highly esteemed by their fellow scientists as well as by (highly educated) members of society. In fact, for a wide spectrum of virtues scientists are more highly valued than other professionals who have been through higher education, such as lawyers, medical professionals, or managers (Veldkamp et al., 2017).

Just as we did with the term 'science' we should stop here for a moment to delineate what we mean with the term 'integrity'. Over the years, discussions about integrity and misconduct have included a host of topics and ended up ranging from very narrow to immensely broad. Traditional definitions of research misconduct focus on activities related to producing and disseminating a research article, namely the fabrication or falsification of data or results, and plagiarism of text, images or ideas (Martin, 2013; Biagioli et al., 2019). These three practices, together known as the core sins of science and labelled FFP, have formed the backdrop of most discussion on integrity and misconduct, especially since the United States based Office of Research Integrity came with a definition of research misconduct centring around these practices (Steneck, 2006; Office of Research Integrity, 1993). However, by now a host of other practices, grouped under the heading 'Questionable Research Practices' (QRP) have been introduced to this discussion. They include the 'less severe' forms of misrepresentation or gaming, including p-hacking, selective reporting and salami publishing (Sacco et al., 2018; Martin, 2013; John et al., 2012). In addition, several practices before and after the production and dissemination of research articles have been included in discussions on misconduct, including pirating of grant applications, coercive citation practices and faking peer review reports (Mongeon et al., 2016; Sacco et al., 2018).

On an even broader spectrum, several actors have proposed to further widen the concepts of integrity and misconduct in science to also include aspects usually referred to as research ethics, for instance encompassing the appropriate handling of laboratory animals. To make matters even more complicated, some have proposed to also include aspects of general ethics and appropriate behaviour under the common denominator of research integrity. These pleas have included the (in)appropriate supervision of (doctoral) students and even sexual harassment (Marin-Spiotta, 2018; Benya, 2019).

The range of notions that I have thus far linked to integrity and misconduct in science is to some extent arbitrary and constitutes a far from comprehensive list. Many more aspects could, and perhaps should, be mentioned. A rich and meticulously written overview of other aspects related to integrity and misconduct can be found in the introduction to a special issue on research misconduct recently published in *Research Policy* (Biagioli et al., 2019).

Controversy over definitions and the scope of integrity and misconduct in science are actually quite common. In our own work (Horbach and Halffman, 2017), analysing discourses in newspaper articles, policy reports and research articles, we demonstrate that the latter two frequently differ in their notion of integrity in various ways. Policy reports usually use rather narrow definitions of integrity, approach the issue from a norm-perspective, and relate it to financial matters (such as public spending). In contrast, research articles more commonly use a broader definition and a value-based approach, while linking the concept to authorship and methodological issues in science. Interestingly, the approaches towards integrity in science seem to be diverging, rather than coming into harmony, over the past years (Horbach and Halffman, 2017).

An important question to ask when delineating research integrity is: whose or what integrity are we actually talking about? Is it the integrity of science as a system that is at stake? Are we concerned about the integrity of individual scientists? Or is it perhaps the research literature's integrity we are discussing, accommodating all results and advice stemming from scientific research? As mentioned earlier, the focus of this dissertation will be on the research literature and science as a (cultural) system of specific practices. Hence, it will address the integrity of the published research record and the practices and cultural organisation constituting the scientific enterprise.

This approach slightly contrasts with the prevalent focus in recent and current debates on research integrity. The discussion about undesirable research practices still mainly focuses on the individual researcher as the source of all evil. This can be recognised in newspaper reports in which individual scientists are pilloried after being caught in violation of written or unwritten rules. It also shows in policy reports or codes of conduct in which scientists are taught how to behave and, sometimes more extensively, how they will be sanctioned in case they fail to live up to these rules. It even shows in the academic literature on research integrity and misconduct. This literature has traditionally focused on individual high-profile cases. Estimates of the number of misbehaving scientists, reasons for their misbehaviour and possible strategies to redirect their daily practices still prevail (Tijdink et al., 2016; Wicherts, 2017). Nevertheless, pleas for a more institutional or collaborative approach to considering research integrity are not absent (Martinson et al., 2013; Hackett, 1994). Indeed, some have rightly embedded discussions on research integrity in more fundamental debates on research funding, research evaluation and academic career trajectories (Stephan, 2012; Butler, 2003). Recently, a more integrative approach, acknowledging the responsibility and influence of organisations and their internal culture has gained considerable momentum. This is reflected amongst others in the recently revised Dutch code of conduct (KNAW et al., 2018) and in a shifted focus of research initiatives (e.g. Haven et al., 2019; Sacco et al., 2018).

Nevertheless, the discussion and research on integrity in science still leaves several topics relatively untouched and much may still be learned. In particular, few studies have focussed on systemic factors leading to problematic academic practices. With an emphasis on individual cases and individual culprits, research has hitherto only been moderately interested in identifying causes of misbehaviour that go beyond personal characteristics. For example, calls have been made for more knowledge on research cultures and practices in diverse academic disciplines and their way of tolerating or even (unintentionally) encouraging problematic research practices. In addition, while considerable attention has been given to the 'publish or perish' imperative in academia and the effects it has on research quality and integrity (Sarewitz, 2016), much can still be learned about this phenomenon. Ample qualitative and bibliometric evidence now indicates that publication pressure and funding structures impact on researchers' publication practices (Aagaard et al., 2015; Schneider et al., 2016). However, questions remain about how and why problematic research may be apparently more common in some research contexts than in others. This hints to the possibility of other factors playing an important role, including specificities of particular research cultures or research institutions.

Even more than increased attention for systemic causes, protagonists of integrity and reproducibility research have argued for more studies on potential ways of avoiding or filtering problematic research. Science is commonly considered to be a self-regulating and self-organising system. Sociologists of science in the tradition of Merton have long assumed that misconduct in science is rare, because researchers have many incentives to trace and filter problematic or erroneous statements (Koshland, 1987; Zuckerman, 1984). However, little is known about how effective these mechanisms are in performing this task. Some have argued for the performance of replication studies or expressed their faith in whistle-blowers detecting and filtering errors or fraud. Yet, the peer review system has been most central in these notions of self-regulation. Nonetheless, even though a tradition of research on peer review has existed since the work of Zuckerman and Merton (1971), few previous studies have focused on the relation between the form of review and the content of that being reviewed. In particular, scholars have recently urged for more knowledge on how these regulating mechanisms perform and what specific formats or models are particularly effective in repelling errors and fraud from the academic literature (Rennie, 2016). They identify open questions including on what we can expect from the peer review system in this process of self-regulation and how we can effectively design it to live up to these expectations. In addition, questions about the role of whistle-blowers, their identities and their motivations to act as such have been posed.

Last, following on questions regarding the detection of errors or fraud in research, perhaps even less attention has been paid to what happens when detection does take place. What mechanisms come into play when a researcher witnesses or accuses a peer of engaging in problematic research practices? And how do institutions and organisations live up to their role in providing safe and fair procedures for reporting and handling misconduct cases? These topics have currently received only marginal attention in empirical studies, while they may directly impact on individual researcher's career and well-being.

Revisiting the conversation that I had with a companion at the beginning of this chapter, we see that even some of the most basic questions, the first to appear on people's mind when hearing about cases of research misconduct, have thus far remained without satisfactory or systematic answer: "*How come no one ever noticed? How did this stay under the radar for so long? And what has happened since it finally came to light?"*

1.2. Questions and objectives

Considering these lacunas in current knowledge about research integrity and (responsible) research practices, this dissertation aims to contribute to the academic debate by addressing multiple topics. Particularly, this dissertation focusses on three related questions, which will each be addressed in a separate section:

1. What factors influence the uptake and spread of erroneous or fraudulent papers within the academic literature?
2. How can the editorial and peer review system be organised to prevent erroneous or fraudulent papers from entering the academic literature?
3. When misconduct is signalled, what mechanisms are at play in reporting, analysing and processing such cases within research performing organisations?

Hence, very briefly described, I will treat three topics: How problematic research gets *spilled* into the research literature, how self-regulating mechanisms may *filter* such articles from the literature, and how organisations may *clean* up cases that were not filtered. These topics will be discussed in three separate parts: *To Spill, To Filter* and *To Clean.*

Throughout this dissertation, I will take a practice perspective to answer these questions. Hence, I will focus on what involved actors *do*, rather than what norms, regulations or guidelines prescribe them to do, or what is considered right from an ethical or moral perspective. I will elaborate on the choice for this particular theoretical perspective later in this chapter.

The first part of the dissertation, *To Spill,* discusses a variety of practices that could be classified as 'questionable research practices' potentially resulting in a 'contamination' of the research literature. Throughout I will use the metaphor of spilling, as in spilling oil, to describe the way in which problematic research articles end up in the 'large ocean' of the academic literature. This part focuses mainly on the extent to which the literature has been contaminated, firstly addressing the question of what constitutes contamination, with a prime interest in two forms of such potential problematic research. Chapter 2 addresses the notion of text recycling or, with a somewhat more problematic term, self-plagiarism. This chapter addresses the controversy around the permissibility of reusing one's own previously written work, asking questions about the legitimacy of recycling and the potential consequences of it. It continues to ask questions about the prevalence and causes of unacceptable text recycling. It strives to identify systematic factors that influence the extent to which scientists recycle their previous work by correlating problematic recycling with research discipline, author productivity levels, number of co-authors, and existence of editorial policy on text recycling.

Chapter 3 then continues to address the use of misidentified cell lines in biomedical research. It first introduces the issue of misidentification in biomedical research and subsequently discusses the question of how information on misidentified cell lines has spread through the research literature. To discuss this topic, it is essential that we get a sense of the size and nature of the problem. This encompasses several questions. First, how many research articles have been based on misidentified or contaminated cell lines? How wide is their influence on the scientific literature? Second, what can we say about origins and trends in the contaminated literature? Is the problem dissolving or is it getting worse? And is it restricted to peripheral regions of the world's research, where perhaps protocols are less strict, or is it a truly global issue? Third, what could be appropriate ways to deal with the contaminated literature? Again, we will focus on institutional and organisational aspects of the issue, aiming to identify factors influencing the problem that exceed individual misbehaviour or sloppiness.

In the second part of this dissertation, *To Filter,* I will move beyond the spread of problematic research through the scientific literature and specifically address the self-regulating mechanisms of academic journals that may prevent such spreading. This part discusses the editorial and peer review system as prime example of such mechanisms. In chapter 4, I will first discuss the ways in which the peer review system has historically emerged by asking what peer review models have been developed and what has been the rationale for developing them. This leads to a classification of different editorial procedures. In the subsequent chapter, chapter 5, I will use this classification to assess peer review's ability to detect and filter problematic research. Based on empirical data on journals' review processes and retracted journal articles, I will assess the relation between editorial procedures and the relative number of articles going through this format that were later retracted. I hereby aim to investigate which review formats are particularly effective in filtering which kind of problematic research.

The last two chapters of this part focus on editorial innovations. While many initiatives to develop and strengthen the editorial and review process have been suggested, very little is known about whether these innovations succeed in convincing editors to implement them in their journal. In these chapters we will examine the rate in which journals innovate their editorial procedures and use ethnographic data to elaborate on the innovation process. This includes discussion about who makes decisions about whether or not to innovate the review process and about the factors that play a decisive role in editorial innovation, including research integrity and other editorial concerns.

This dissertation's last empirical part, *To Clean,* shifts attention from the academic literature to focus on research performing organisations. It will address self-regulating mechanisms at play in such organisations, particularly the role of whistle-blowers and organisational research integrity committees. Rather than on preventing problematic research from entering the literature, this part consists of analyses of how such cases are dealt with when they are spotted downstream. Chapter 8 first discusses the role of reporting alleged misconduct and deals with questions on what motivates researchers to (not) report witnessed cases of problematic behaviour and how institutional hierarchies and power structures play a role in this.

The last empirical chapter, chapter 9, then takes a final step to study cases in which allegations of research misconduct have been made. It focuses on the organisational responses to such allegations in universities. Through studying four cases in European universities, it aims to analyse how the respective organisations handle allegations of misconduct: Which actors are involved in the handling of such accusations? What mechanisms play a role in dealing with these cases? And what are the factors and motives behind initiating procedures or determining their outcome? Lastly, I aim to clarify the consequences of the specific ways in which cases of alleged misconduct are dealt with and I try to assess the extent to which organisations exploit the learning potential that such cases inherently possess.

1.3. Conceptual framework

1.3.1. Science and technology studies

To tackle these questions, this dissertation draws from insights from a plurality of theoretical and conceptual traditions. First and foremost, it has been influenced by insights from the field of Science and Technology Studies (STS). As one of its main characteristics, STS starts from an assumption that science and technology are utterly social activities. This is to say that scientists always belong to a community, that they are trained into the practices of this community, and that the community sets standards for inquiry and evaluation. Consequently, science, as well as technology, is shaped by processes in which social, practical and rhetorical work is crucial. Its actors, researchers, are not merely logical operators, but instead carry with them investments in skills, practices, prestige, knowledge and reputation, all of which influence the way in which they try to convince their colleagues of their ideas' value as well as the way in which they evaluate their peers' ideas. This has led STS to consider science and technology as set of diverse and fundamentally social practices, that may mean different things to different audiences. The aim of studying science is then to investigate how scientific knowledge is produced or *constructed* (Sismondo, 2010).

To be more precise, STS has been shaped by three notions of knowledge production (Wouters et al., 2013). First, it understands knowledge to be inscribed in and by technology and instruments, be it either social or material. That is to say that there is no direct path from nature to knowledge about nature, but that such path is created through the use of social and material technologies. Second, for knowledge to exceed the level of what one person 'knows' it has to be embedded in, and performed by, communities or infrastructures (Edwards, 2010). It is in the embedding of knowledge into such communities and infrastructures that the social character of knowledge becomes most apparent. Last, knowledge creation occurs in interaction between knowledge infrastructures, knowledge producers and knowledge 'users'. Knowledge is 'co-constructed' in an interplay which integrates, among many others, political, social, legal, ethical, bureaucratic, medical, technical and personal domains (Bowker and Star, 2000; Sismondo, 2010).

Some critics have argued that an STS approach rejects critical concepts such as truth, rationality and the reality of the material world. However, rather than rejecting those, STS, and specifically the Strong Programme that was at the root of its establishment (Bloor, 1991), shows how on their own terms, rationality and the material world have only limited value in describing why any particular scientific claim is believed over another. It argues that in order to understand why some scientific beliefs are maintained and others are not, we have to consider interpretations of those beliefs and the rhetorical work required to make such interpretations endure (Sismondo, 2010).

While STS shapes the general approach to science used in this dissertation, STS has informed my research in two more specific ways. First, STS' understanding of knowledge production has direct consequences for how to approach and conceive research misconduct in the process of knowledge creation. For instance, while fabrication, falsification and plagiarism are commonly understood as the major sins in science, plagiarism is often conceived as the least detrimental of the three, because it supposedly does not directly affect the ‘truth’ that science generates. However, being sensitive to the rhetorical work done by prestige, reputation and credibility in forming consensus about scientific claims, plagiarism is much more than just a threat to the public trust in science or an issue of giving fair credit. It interferes with the consensus-building process to establish theories and claims, giving an assertion more support than it merits (Penders, 2017). From this perspective, plagiarism is detrimental to the research and knowledge creating process itself, just as fabrication and falsification are. Equivalent reasonings apply to other issues, such as authorship disputes and dubious research practices such as text recycling. If science is profoundly social, then social transgressions profoundly affect science.

A second way in which STS informs and shapes this dissertation is by providing sensitivity to context and the situatedness of academic practices. STS understands scientists to be grouped in *epistemic cultures* that may have completely different orientations to their objects, social units of knowledge production, and patterns of interaction. In addition, different epistemic cultures may evaluate and obtain credibility for their research in diverse ways (Hessels et al., 2019). This leads to a differentiation of the normativity of research practices: while some ways of doing or reporting research may be dubious or considered fraudulent in one research discipline, it may be fully acceptable, common practice or even desirable in another. One of the goals of this dissertation will hence be to study these differences, which may in turn lead to a variety of recommendations towards diverse disciplines. A ‘one-size-fits-all’ approach, based on a common understanding of ‘the scientific method’ (assuming only one exists), will not be our frame of reference.

Third, in this dissertation I will return to some of the very traditional topics of early STS, including studies of peer review (e.g. Zuckerman and Merton, 1971). However, I will do this with a slightly different perspective. Traditional studies of peer review and refereeing had a strong focus on publishing and citing as a way of distributing and obtaining credit. I will add to this a perspective that studies the relation between peer review and the content of scientific knowledge being reviewed, in line with the traditional critique voiced by scholars from the sociology of scientific knowledge (SSK) on Merton’s work. This hence broadens the traditional STS perspective on peer review from a study of researchers to a study of research. Similarly, this dissertation builds on earlier STS work on dirt, contamination and cleanliness in science (e.g. Douglas, 1966; Mody, 2001). This work, inspiring this dissertation’s title, has analysed how less than theoretically optimal research materials, practices or settings come into being and how they have influenced knowledge production. This dissertation builds on these studies by extending their scope towards misconduct and QRP’s as potential pollutants of science.

Besides being informed by considerations from STS, this dissertation builds on several other theoretical frameworks. This includes insights from innovation studies and practice theory, as well as concepts from organisational theory and criminology, particularly the theories of neutralisation and sensemaking. When applying these theoretical frameworks, STS will continuously form my frame of reference, which I will explicate in the subsequent subsections.

1.3.2. Scientific publishing and innovation studies

In the second part of Part II, *To Filter*, I use concepts from innovation studies to describe transformations in the editorial and peer review process of academic journals. This part's chapters take an explicit STS perspective on innovation studies to recognise and study the motivations for journal editors to innovate editorial procedures. Distribution and implementation of innovations has been subject of analysis by scholars from a wide set of disciplines. In most of the traditional literature it has been described as a process of 'diffusion' (e.g. Greenhalgh et al., 2004; Peres et al., 2010), but we distance ourselves from this term as we consider it to be slightly misleading. Following the STS perspective on knowledge production and technological development, we consider adoption of new innovations the result of a social process and concrete actions, rather than a naturally occurring phenomenon, like the diffusion of chemical substances.

Several factors play a crucial role in the uptake and implementation of innovations. First, innovations are picked up by active knowledge seekers. These actors do so in an active process of adoption (Greenhalgh et al., 2004), hence innovations do not travel based on their own force and they are not collected by passive recipients. Second, actors are keener to search for innovations under specific circumstances. They predominantly do so when they experience a problem or recognise an opportunity to improve their current situation (Wisdom et al., 2014). Third, users *domesticate* innovations (Oudshoorn and Pinch, 2007; von Hippel, 1976). This is to say that they adapt an innovation to their own specific needs and circumstances. New initiatives have to be aligned with current practices and habits and hence shape and are shaped by the specific context in which they are to be implemented. Fourth, users' expectations of new technologies play a crucial role in their inclination to implement such innovations (Brown and Michael, 2003; Verbong et al., 2008; Van Lente, 1993).

Especially the latter three factors of this list critically depend on the specific context of implementation. In our case, this encompasses the editorial process of academic journals and the academic publishing landscape more generally. This scholarly publishing landscape is currently rapidly changing, raising concerns in public discourse on academic journals and the expectations of the editorial process. Both on an academic and a societal level, mounting frustration and critique have been expressed over the way in which public funds are increasingly transferred into the hands of ever fewer large commercial publishers (Kallio, 2017; Fletcher et al., 2019). The enormous profit margins of these publishers have been a constant point of concern, especially in an era where (online) publishing is increasingly easy and the added value of commercial publishers is blurring (Fletcher et al., 2019; Larivière et al., 2015). This has led to shifts in publishing strategies, including the Open Access movement (Laakso and Björk, 2012), and the rise of alternative publishing platforms such as preprint servers (Gunnarsdottir, 2005).

All of this has had major impact on various stakeholders' expectations of journals and the editorial process. Where it traditionally aimed to merely distinguish between 'good' and 'bad' science, making decisions about what is appropriate research to fill scarce journal pages with, new expectations have been developing. By now, a host of more contentious expectations of the peer review and editorial system have been voiced (Fyfe et al., 2017). The detection of fraud and errors is a prominent example of these. Supported by the introduction of digital and automated tools such as plagiarism scanners or software to check for flaws in statistics (Nuijten et al., 2016), scholars are increasingly expecting journals to filter problematic research upstream (Stroebe et al., 2012). With further developments of software tools to be expected, such as statistics or image scanners with possible use of Artificial Intelligence, these expectations might be prone to further shifts. A hint to this was expressed by Mirowski (2018) when he described a patent entitled 'Online peer review and method' Issued by the US Patent Office. The owner of the patent is none other than the for-profit mega-publisher Elsevier. The essence of the patent consists of a peer review process completely organised and executed by a computer program. How these developments and shifting expectations inform decisions about implementation of innovative editorial procedures, as well as how effective these procedures are in filtering problematic research, will be the focus of our studies in part II.

1.3.3. Organisation studies: between the system and the individual

In the final empirical part of this dissertation, I turn my attention away from the academic publication system to focus on another set of important actors in the research landscape, research performing organisations. As I mentioned before, the discourse on research integrity and misconduct has primarily focussed on individual perpetrators and their personal characteristics (Zwart and ter Meulen, 2019). Opposing this trend, scholars have repeatedly pointed to other factors, mainly addressing what have been labelled 'systemic' or 'ecosystem' characteristics, such as the pressure to publish and the introduction of new public management to higher education (Overman et al., 2016). Both perspectives, the individual and the systemic, deserve merit, and while issues on an individualistic level are currently reasonably well understood, many questions remain unanswered with respect to more systemic factors including the role of research evaluation and funding programs (Degn, 2018; Müller and de Rijcke, 2017). Nonetheless, a crucial layer of the academic landscape seems to be largely overlooked. This comprises the institutional and organisational setting in which researchers perform their daily tasks, where they get socialised into the academic profession and might be corrected when they transgress the local community's cultural and moral norms.

In the latter part of the dissertation, I will therefore draw on organisational theory to assess universities' responses, as organisations, to cases of alleged misconduct. Researchers usually have a great deal of autonomy in their work because they operate as professionals with a high degree of personal discretion, ideally demarcated from political requirements through the notion of "academic freedom". Yet, researchers are also employees, and they work in formal organisational surroundings where universities or other organisations, as employers, have the responsibility to act in situations of alleged misconduct. While notably absent in the discussion and analysis of misconduct in science, organisational characteristics have been widely studied in the context of other forms of misconduct, such as white-colour crime, corruption, financial fraud and medical errors (e.g. Vaughan, 1999; Thurman, 2001; Murphy and Dacin, 2011; Trinkle et al., 2017; Ashforth et al., 2008). These studies highlight the role of institutional characteristics on the emergence of questionable behaviour, as well as outline the ways in which micro, meso and macro levels interact to provide incentives, opportunities and motives for (im)proper behaviour. This literature emphasises the significance of social control, the presence of regulatory frameworks, the extent to which organisations are hierarchically structured, and the perceived level of managerial pressure on individual scholars in understanding the occurrence of organisational misconduct (Zuckerman, 1977; Vaughan, 1999; Vaughan, 2002; Murphy and Dacin, 2011). The last part, *To Clean,* will elaborate on these concepts to study how organisational settings and procedures affect the willingness of researchers to report cases of witnessed misconduct and how such allegations are subsequently dealt with by universities.

1.4. Methodology

In addressing the various research questions and objectives, this dissertation employs multiple methods. In particular, the dissertation combines various qualitative and quantitative approaches to study research practices, science, and its institutions. As far as the quantitative part is concerned, this dissertation has mainly drawn from insights and methods of scientometrics. This field, described as the quantitative and systematic study of scientific communication (Wouters, 1999), has provided tools to systematically analyse the scientific literature. I employed these tools specifically to analyse how information on misidentified cell lines spreads through the scientific literature and how different peer review models are related to retraction rates. This involved tracking down research records reporting misidentified cell lines from the scientific literature, as well as indicators on their origin, e.g. year of publication, research discipline and geographical origin; and impact, e.g. their number of citations. Similarly, a scientometric approach was used to study the phenomenon of retractions in science. Again, retracted articles were identified in the scientific literature, as well as information on their origin and impact. More specifically, retractions were linked to the peer review formats that assessed the article prior to publication. In this way, a novel database of retractions and peer review formats was established, providing specific tools for future research in this research area. In addition, I employed tools from scientometrics in chapter 2 to study large samples of research articles on the extent of text recycling in relation to various descriptive variables such as the number of co-authors, research discipline and productivity of the authors. Lastly, further quantitative analyses were performed in chapters 6 and 8, using survey data on editorial procedures and whistleblowing practices. These data were used to analyse which journals have implemented which specific editorial procedures, as well as to map innovation trajectories of such procedures. Chapter 8 uses survey data to correlate organisational and personal characteristics to researchers' willingness to report witnessed cases of academic misbehaviour.

We should note here that using scientometric approaches to study research misconduct has a slightly controversial, if not ironic connotation. Similar to what I have done in my dissertation, scientometrics can be used to trace, assess and evaluate several instances of dubious behaviour by mapping patterns in scholarly communication. However, with its traditional focus on citations and unidimensional output indicators, the field has been blamed for augmenting publication pressure, both for researchers and academic journals, ultimately incentivising problematic behaviour. Arguably, practices such as problematic text recycling or the formation of citation cartels are direct results of the fact that performance, of individuals or journals, is measured by simple productivity or citation counts (Mongeon et al., 2016). Especially the audit culture in science, leading to science's reward system heavily based on quantitative performance measurements is critiqued for being one of the factors nourishing questionable research practices and outright fraud (de Rijcke, 2019). It is only in this context that some of the novel questionable practices (such as salami slicing, self-plagiarism and the formation of citation cartels) have become meaningful. Nevertheless, scientometric approaches, if used in an informed and multidimensional way, can provide valuable insights in academic practices and hence contribute to the understanding of research integrity. Further methodological details are included in the chapters using these methods.

While quantitative methods allow studying phenomena at large scale and exposing geographical patterns and historical trends, they lack the ability to study specific particularities of research practices, to analyse motivations and practices' embedding in a cultural context. Therefore, the quantitative approaches in this dissertation are complemented with several qualitative methods to constitute a mixed methods approach. These qualitative methods comprise the use of in-depth interviews for case study analyses as well as ethnographic fieldwork at the editorial offices of academic publishers.

Using multiple methods allows me to analyse research and editorial practices from multiple perspectives. Ethnography brings with it a sensibility for specificity that most quantitative methods lack. First, it provides a tool capable of contextualising decisions, surfacing silenced voices, juggling disparate meanings and understanding the gap between words and deeds (Star, 1999; Hammersley and Atkinson, 2007). It makes visible how decisions were made, what struggles went into this decision making, as well as how problems were or were not solved through selecting a particular route. Second, it shows that people make meanings based on their circumstances, and that these meanings are inscribed into the judgments about their environments. Thereby, it allows complementing the quantitative data to yield a richer account of research and editorial practices and the organisational settings in which they are embedded. Not only can we now study what practices, forms of scholarly communication or organisational procedures are performed or executed, we are also able to grasp why they are engaged in, and why others are discarded. By considering motivations and reflecting on instances of decision making, we may obtain a richer understanding of the underlying mechanisms and causes of some of the patterns distilled through quantitative analyses. Of course, the ethnographic approaches also come with their drawbacks: its local character focussing on particularities of a specific field site complicate generalisation, the collection and analyses of data are continuously prone to interpretation and potential bias by the researcher, and collecting data is highly dependent on people’s openness, trust and willingness to collaborate. Therefore, the use of a mixed methods approach is important to allow for valuable triangulation of different forms of data collection, hence providing more robustly informed claims.

1.4.1. Reflection and collaboration

Originally trained as a mathematician, I was particularly well-equipped to perform the quantitative analyses for my dissertation. On the contrary, I was far less experienced in employing qualitative methods such as conducting in-depth interviews. Especially given the delicate conditions in which these interviews had to be conducted, this posed some demands on my skills of improvisation and adaptation. Interviewing those who had been closely involved in cases of alleged research misconduct was both one of the more challenging as well as rewarding exercises that led to the finalisation of this dissertation. I will reflect on the issue of studying sensitive topics and choosing appropriate methods to do so in the epilogue of this dissertation.

I should note that parts of this dissertation’s studies, most notably the work reported in chapters three, eight and nine, were conducted in the context of the European PRINTEGER project (PRINTEGER, 2016). This project, aiming to study and promote research integrity as an integral dimension of what it means to do good research, has offered a rich research environment of diverse research disciplines and cultures. With members from disciplines as diverse as criminology and law to sociology and medical ethics, based in eight European countries the project provided a sensitivity to this diversity and the way in which it affects research integrity challenges. Conversely, this dissertation's studies have contributed to further understanding the diversity and institutional relevance of research integrity matters. With the projects aim to not only understand but also foster a healthy research integrity climate among research organisations, a specific focus on applicability of research results was carried along to this dissertation. This resulted among others in the development of the Bonn-PRINTEGER statement (Forsberg et al., 2018). I will further reflect on the results of this dissertation's study and their implementation in actual research practices in chapter ten and the epilogue.

1.5. **Embedding and contributions**

This study hopes to contribute to the understanding of research integrity and thereby ultimately foster responsible research practices. Thereby it extends research on a particularly timely topic that has witnessed a considerable increase in attention over the past decade. Researchers, policy makers and a general public alike have increasingly recognised the importance of responsible research practices in performing and disseminating research, while simultaneously acknowledging that these practices are not self-evident. With a perceived higher pressure on researchers to perform fast, collaborative, newsworthy and high-quality research; increased scrutiny on how they do so; and intensified controversy over what constitute 'good' research practices, a better understanding of research and research integrity is urgently needed.

The augmented attention for this topic both from researchers, funders, policymakers and the public at large is reflected in numerous initiatives to update policy on research integrity (KNAW et al., 2018), enrol new funding programs (ZonMw, 2016), and large coverage of misconduct cases by the media. This interest of various stakeholders in issues of research integrity is also demonstrated by some of the responses to our work, which in themselves constitute an interesting source of data on this topic. I will reflect more elaborately on this phenomenon in the epilogue.

In this dissertation, I take a slightly different perspective on research integrity than is common in most of the contemporary and traditional literature. A substantial amount of research on responsible research practices and misconduct is currently done by scholars identifying with the newly emerging subfield of 'research-on-research' or 'metaresearch' (Enserink, 2018). These researchers study the research process and how it may be improved, but differ in several ways from the theoretical understanding of conceptualising science common in STS. Among others, scholars in this emerging subfield commonly start from realist or positivist assumptions about research, with less sensitivity for the inherently social and cultural nature of science. Second, metaresearchers commonly presupposes that research is performed according to textbook rules, hence paying less attention to the STS understanding that research practice is often much more complex and messier, that it depends on many practicalities and that it is seldom free of idiosyncrasies. In addition, much of the studies in metaresearch are performed by and focus on scholars in either biomedicine or psychology, leaving blank spots with respect to other research disciplines, as well as to the epistemic, methodological and cultural differences between the disciplines. In this dissertation, I aim to bridge the important work done by scholars in this subfield with more traditional studies and understandings of science from Science and Technology studies and scientometrics.

Following the perspectives and methods outlined above, this dissertation aims to contribute to the available literature on research integrity in multiple ways. First, as I have noted before, the prevalent discourse in studies on research integrity has focussed on individual cases of integrity breaches or on perpetrators' characteristics. Studies taking a different approach commonly focus on rules or codes of conduct (e.g. Godecharle et al., 2014). In this dissertation, I aim to widen this perspective on research integrity by focussing on research and editorial practices, studying how they shape and are shaped by the organisational, disciplinary or cultural setting in which they are embedded. Thereby I aim to distil patterns within or across research disciplines and networks of researchers, allowing for particularities and variability within these groups. While acknowledging that individual choices and practices remain crucial, I aim to analyse mechanisms and causes on an institutional or systemic level that transcends individual researchers. This includes links with organisational cultures, disciplinary contexts and research evaluation mechanisms.

Second, by distilling patterns in how problematic research spreads, how it might be filtered and how unfiltered cases are acted upon, this dissertation combines a fairly comprehensive set of factors and stages involved in irresponsible research practices. By also studying underlying mechanisms and motivations of actors to behave in particular ways, this allows for precise recommendations on how to foster responsible research and ultimately contribute to better science. Especially in the case of editorial practices, we specifically studied drivers and hurdles for implementing innovations. This may hence inform recommendations on how innovation processes might be steered to effectively address research integrity challenges.

Third, the dissertation aims to contribute to tailored recommendations with a view of specific contexts of implementation. Much of the existing research on research integrity, editorial practices and research misconduct has been carried out in the realm of biomedical science. While this is understandable from the perspective that breaches of integrity might have most severe consequences in this context, it has arguably led to uniformisation of solutions to integrity issues. In this dissertation's studies, we broaden this scope to include analyses of other disciplines and research contexts. This may expose the diversity of research integrity and editorial challenges in different contexts, but it also provides valuable input to tailor solutions to these challenges that fit the specific context's needs.

Fourth, this dissertation aims to theorise the editorial process and organisational contexts influencing crucial decisions regarding research integrity, such as whistleblowing and the handling of misconduct allegations. While some empirical work has been done on these topics, a theoretical understanding of them is currently lacking.

Fifth, by using a mixed-methods approach, through using both observational studies and qualitative analyses, it contributes to traditions of ethnographies of publishing (e.g. Hirschauer, 2010; Jacob, 2019) and scientometric analyses of research communication, but complements both by bridging these often sharply separated approaches. This may contribute both theoretically as well as methodically to both fields, performing a combination of methods that is in need of further exploration (de Rijcke, 2019).

Ultimately this dissertation hopes to contribute to a research atmosphere that is responsible, righteous, and trustworthy. In a societal context where trust in established institutions, including science, is under threat, this goal is particularly timely. The dissertation aims to answer some very basic questions popping up after every single case of research misconduct:

"*How come no one ever noticed? How did this stay under the radar for so long? And what has happened since it finally came to light?"*

Answering these questions both systematically, as well as with sensitivity to particular contexts, may further research on integrity issues. I aim to contribute to this not by pointing fingers at individual researchers, or by blaming 'the system as a whole', but rather by investigating how specific challenges to research integrity emerge and spread, how they are dealt with in specific contexts, and how particular actors, both individuals as well as organised groupings, may help to foster integrity in research. Let us now turn our attention to the first set of questions addressed in this dissertation: how research integrity challenges emerge and problematic research spills through the academic literature.

1.6. References

Aagaard K, Bloch C and Schneider JW. (2015) Impacts of performance-based research funding systems: The case of the Norwegian Publication Indicator. *Research Evaluation* 24: 106-117.

Ashforth BE, Gioia DA, Robinson SL, et al. (2008) Re-viewing orgnizational corruption. *Academy of Management Review* 33: 670-684.

Benya F. (2019) Treating Sexual Harassment as a Violation of Research Integrity. *Issues in Science and Technology* 35: 56-59.

Biagioli M, Kenney M, Martin BR, et al. (2019) Academic misconduct, misrepresentation and gaming: A reassessment. *Research Policy* 48: 401-413.

Bloor D. (1991) *Knowledge and social imagery*: University of Chicago Press.

Bowker GC and Star SL. (2000) *Sorting things out: Classification and its consequences*: MIT press.

Brown N and Michael M. (2003) A Sociology of Expectations: Retrospecting Prospects and Prospecting Retrospects. *Technology Analysis & Strategic Management* 15: 3-18.

Butler L. (2003) Modifying publication practices in response to funding formulas. *Research Evaluation* 12: 39-46.

Callon M. (1995) Four Models for the Dynamics of Science. In: Jasanoff S, Markle GE, Petersen JC, et al. (eds) *Handbook of Science and Technology Studies.* Thousand Oaks: SAGE, 29-63.

Collins H and Pinch T. (1993) *The Golem: What Everyone Should Know About Science,* Croyden: Cambridge University Press.

de Rijcke S. (2019) Wetenschap, op waarde geschat. *Faculteit der Sociale Wetenschappen.* Leiden: Leiden University.

Degn L. (2018) Academic sensemaking and behavioural responses – exploring how academics perceive and respond to identity threats in times of turmoil. *Studies in Higher Education* 43: 305-321.

Douglas M. (1966) *Purity and Danger: An Analysis of Concepts of Pollution and Taboo,* London: Routledge.

Edwards PN. (2010) *A vast machine: Computer models, climate data, and the politics of global warming*: Mit Press.

Enserink M. (2018) Research on research. *Science* 361: 1178-1179.

Fletcher R, Hall S, Corbera E, et al. (2019) The future of scholarly publishing: Paywalls and profits or a new plan? *Geoforum* 102: 1-4.

Forsberg E-M, Anthun FO, Bailey S, et al. (2018) Working with Research Integrity—Guidance for Research Performing Organisations: The Bonn PRINTEGER Statement. *Science and Engineering Ethics* 24: 1023-1034.

Fyfe A, Coate K, Curry S, et al. (2017) Untangling academic publishing: a history of the relationship between commercial interests, academic prestige and the circulation of research.

Godecharle S, Nemery B and Dierickx K. (2014) Heterogeneity in European Research Integrity Guidance: Relying on Values or Norms? *Journal of Empirical Research on Human Research Ethics* 9: 79-90.

Greenhalgh T, Robert G, Macfarlane F, et al. (2004) Diffusion of innovations in service organizations: Systematic review and recommendations. *Milbank Quarterly* 82: 581-629.
Gunnarsdottir K. (2005) Scientific journal publications: On the role of electronic preprint exchange in the distribution of scientific literature. *Social Studies of Science* 35: 549-579.
Guston DH. (2007) *Between politics and science: Assuring the integrity and productivity of reseach*: Cambridge University Press.
Hackett EJ. (1994) A social control perspective on scientific misconduct. *The Journal of Higher Education* 65: 242-260.
Halffman W. (2019) What Is Science? (And Why Does This Matter?). In: Turnhout E, Halffman W and Tuinstra W (eds) *Environmental Expertise: Connecting Science, Policy and Society.* Cambridge: Cambridge University Press, 11-35.
Hammersley M and Atkinson P. (2007) *Ethnography: Principles in practice*: Routledge.
Haven TL, Tijdink JK, Martinson BC, et al. (2019) Perceptions of research integrity climate differ between academic ranks and disciplinary fields: Results from a survey among academic researchers in Amsterdam. *Plos One* 14: e0210599.
Hessels LK, Franssen T, Scholten W, et al. (2019) Variation in Valuation: How Research Groups Accumulate Credibility in Four Epistemic Cultures. *Minerva* 57: 127-149.
Hirschauer S. (2010) Editorial Judgments: A Praxeology of 'Voting' in Peer Review. 40: 71-103.
Horbach SPJM and Halffman W. (2017) Promoting Virtue or Punishing Fraud: Mapping Contrasts in the Language of 'Scientific Integrity'. *Science and Engineering Ethics* 23: 1461-1485.
Jacob M-A. (2019) Under repair: A publication ethics and research record in the making. *Social Studies of Science* 49: 77-101.
John LK, Loewenstein G and Prelec D. (2012) Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling. *Psychological Science* 23: 524-532.
Kallio KP. (2017) Suble radical moves in scientific publishing. *Fennia - International Journal of Geography* 195: 1-4.
KNAW, NFU, NWO, et al. (2018) Nederlandse gedragscode wetenschappelijke integriteit.
Koshland DE. (1987) Fraud in Science. *Science* 235: 141-141.
Laakso M and Björk B-C. (2012) Anatomy of open access publishing: a study of longitudinal development and internal structure. *Bmc Medicine* 10: 124.
LaFollette MC. (1992) *Stealing into print: fraud, plagiarism, and misconduct in scientific publishing*: Univ of California Press.
Larivière V, Haustein S and Mongeon P. (2015) The Oligopoly of Academic Publishers in the Digital Era. *Plos One* 10: e0127502.
Luhmann N. (1992) *Die wissenschaft der gesellschaft*: Suhrkamp Frankfurt am Main.
Marin-Spiotta E. (2018) Harassment should count as scientific misconduct. *Nature* 557: 141-141.
Martin BR. (2013) Whither research integrity? Plagiarism, self-plagiarism and coercive citation in an age of research assessment. *Research Policy* 42: 1005-1014.
Martinson BC, Thrush CR and Lauren Crain A. (2013) Development and Validation of the Survey of Organizational Research Climate (SORC). *Science and Engineering Ethics* 19: 813-834.
Merton RK. (1973) *The sociology of science: Theoretical and empirical investigations*: University of Chicago press.

Mirowski P. (2018) The future(s) of open science. *Social Studies of Science* 48: 171-203.
Mody CCM. (2001) A Little Dirt Never Hurt Anyone. *Social Studies of Science* 31: 7-36.
Mongeon P, Waltman L and Rijcke Sd. (2016) What do we know about journal citation cartels? A call for information. In: CWTS (ed). Leiden: Leiden University.
Müller R and de Rijcke S. (2017) Exploring the epistemic impacts of academic performance indicators in the life sciences. *Research Evaluation* 26: 157-168.
Murphy PR and Dacin MT. (2011) Psychological pathways to fraud: Understanding and preventing fraud in organizations. *Journal of business ethics* 101: 601-618.
Nuijten MB, Hartgerink CHJ, van Assen MALM, et al. (2016) The prevalence of statistical reporting errors in psychology (1985–2013). *Behavior Research Methods* 48: 1205-1226.
Office of Research Integrity. (1993) *Definition of Research Misconduct*. Available at: https://ori.hhs.gov/definition-misconduct.
Oudshoorn NEJ and Pinch T. (2007) User-technology relationships: Some Recent Developments. In: Hackett EJ, Amsterdamska O, Lycnh M, et al. (eds) *Handbook for Social Studies of Science.* 3 ed. London: MIT Press, 541 - 567.
Overman S, Akkerman A and Torenvlied R. (2016) Targets for honesty: How performance indicators shape integrity in Dutch Higher Education. *Public Administration* 94: 1140-1154.
Penders B. (2017) Beyond Trust: Plagiarism and Truth. *Journal of Bioethical Inquiry*.
Peres R, Muller E and Mahajan V. (2010) Innovation diffusion and new product growth models: A critical review and research directions. *International Journal of Research in Marketing* 27: 91-106.
PRINTEGER. (2016) *Documents and Results*. Available at: https://printeger.eu/documents-results/.
Rennie D. (2016) Make peer review scientific. *Nature* 535: 31-33.
Sacco DF, Bruton SV and Brown M. (2018) In Defense of the Questionable: Defining the Basis of Research Scientists' Engagement in Questionable Research Practices. *Journal of Empirical Research on Human Research Ethics* 13: 101-110.
Sarewitz D. (2016) The pressure to publish pushes down quality. *Nature* 533: 147-147.
Schneider JW, Aagaard K and Bloch CW. (2016) What happens when national research funding is linked to differentiated publication counts? A comparison of the Australian and Norwegian publication-based funding models. *Research Evaluation* 25: 244-256.
Sismondo S. (2010) *An Introduction to Science and Technology Studies*: Wiley - Blackwell.
Star SL. (1999) The Ethnography of Infrastructure. *American Behavioral Scientist* 43: 377-391.
Steneck NH. (2006) Fostering integrity in research: Definitions, current knowledge, and future directions. *Science and Engineering Ethics* 12: 53-74.
Stephan PE. (2012) *How economics shapes science*: Harvard University Press Cambridge, MA.
Stroebe W, Postmes T and Spears R. (2012) Scientific Misconduct and the Myth of Self-Correction in Science. *Perspectives on Psychological Science* 7: 670-688.
Thurman A. (2001) Institutional responses to medical mistakes: ethical and legal perspectives. *Kennedy Institute of Ethics Journal* 11: 147-156.
Tijdink JK, Bouter LM, Veldkamp CLS, et al. (2016) Personality Traits Are Associated with Research Misbehavior in Dutch Scientists: A Cross-Sectional Study. *Plos One* 11: e0163251.

Trinkle BS, Phillips T, Hall A, et al. (2017) Neutralising fair credit: factors that influence unethical authorship practices. *Journal of Medical Ethics* 43: 368-373.

Van Lente H. (1993) Promising technology. The dynamics of expectations in technological developments. Delft: Twente University.

Vaughan D. (1999) The dark side of organizations: Mistake, misconduct, and disaster. *Annual review of sociology* 25: 271-305.

Vaughan D. (2002) Criminology and the sociology of organizations. *Crime, law and social change* 37: 117-136.

Veldkamp CLS, Hartgerink CHJ, van Assen MALM, et al. (2017) Who Believes in the Storybook Image of the Scientist? *Accountability in Research* 24: 127-151.

Verbong G, Geels FW and Raven R. (2008) Multi-niche analysis of dynamics and policies in Dutch renewable energy innovation journeys (1970–2006): hype-cycles, closed networks and technology-focused learning. *Technology Analysis & Strategic Management* 20: 555-573.

von Hippel E. (1976) The dominant role of users in the scientific instrument innovation process. *Research Policy* 5: 212-239.

Wicherts JM. (2017) The Weak Spots in Contemporary Science (and How to Fix Them). *Animals (Basel)* 7.

Wisdom JP, Chor KHB, Hoagwood KE, et al. (2014) Innovation Adoption: A Review of Theories and Constructs. *Administration and policy in mental health* 41: 480-502.

Wouters P. (1999) The Citation Culture. *Faculteit der Scheikunde.* Amsterdam: University of Amsterdam.

Wouters P, Beaulieu A, Scharnhorst A, et al. (2013) *Virtual knowledge: experimenting in the humanities and the social sciences*: MIT Press.

ZonMw. (2016) *Bevorderen Verantwoorde Onderzoekspraktijken - Over dit programma*. Available at: https://www.zonmw.nl/nl/onderzoek-resultaten/fundamenteel-onderzoek/programmas/programma-detail/bevorderen-van-verantwoorde-onderzoekspraktijken/.

Zuckerman H. (1977) Deviant behavior and social control in science. *Deviance and social change* 1: 87-138.

Zuckerman H. (1984) Norms and deviant behavior in science. *Science, Technology, & Human Values* 9: 7-13.

Zuckerman H and Merton RK. (1971) Patterns of evaluation in science: Institutionalisation, structure and functions of the referee system. *Minerva* 9: 66-100.

Zwart H and ter Meulen R. (2019) Addressing research integrity challenges: from penalising individual perpetrators to fostering research ecosystem quality care. *Life Sciences, Society and Policy* 15: 5.

Part I

TO SPILL

In the first part of this dissertation we will direct our attention to instances of problematic research. Using two examples of such research, I will discuss why such research might, or might not, be considered problematic, how conventions differ between different research disciplines and contexts, and how instances of problematic research spread or spill through the academic literature. Both chapters in this part use a quantitative approach to assess how many articles are affected by problematic research practices and to analyse differences in the contexts in which they appear, e.g. regarding research discipline or geographical origin.

In chapter 2 I will first study text recycling or 'self-plagiarism' as a particularly novel and still very contentious instance of questionable research practices. There has been heated debate about the appropriateness of reusing one's own text in future publications and the conditions under which this should be tolerated. While the conditions of fair text reuse have been specified more clearly in the wake of recent prominent cases of text recycling, the extent and causes of problematic text recycling remain unclear. Some consider it as a new way of gaming the reward system of science, while others still argue that, in some circumstances, recycling is appropriate or even desirable. After summarising this debate, chapter 2 goes on to analyse the frequency and patterns of occurrence of text recycling in articles published by researchers at Dutch universities.

Chapter 3 continues to discuss the slightly less recent and more established issue of using misidentified or contaminated cell lines in biomedical research. The usage of such cell lines in research has been alarmed for since the 1960s, but still remains a prominent issue of concern. To date, little is known about the size of the problem or the amount of research affected by it. In our study, we therefore make an estimation of the number of research articles reporting on research with misidentified cells. In addition, we trace how potentially false information in these 'contaminated' articles spreads through the literature by analysing citation patterns. We analyse trends in the publication and usage of contaminated articles, hence helping to understand the effectiveness of decades-old attempts to stop misidentification of cell lines.

Both chapters use quantitative, mainly scientometrically informed, methods to study the spread of problematic research through the research literature. Scientometric techniques are used to sample and trace articles, to classify them based on disciplinary and geographical origin, and to assess citation networks. In addition, chapter 2 depends on the usage of text similarity scanners to detect overlap between different research articles.

Together, the studies in this part of the dissertation provide the basis of our understanding about problematic research. It distils frequencies of occurrence, assesses differences in occurrence patterns and traces potential causes for researchers to engage in problematic practices. This will feed into later parts of this dissertation when we ask how problematic research may be filtered and how it may be acted upon and cleaned after detection.

Chapter 2 - The extent and causes of academic text recycling or 'self-plagiarism'

This chapter was published as:
Horbach, S. P. J. M., & Halffman, W. (2019). The extent and causes of academic text recycling or 'self-plagiarism'. Research Policy, 48(2), 492-502. doi: 10.1016/j.respol.2017.09.004

2

2.1 Introduction

Of the various forms of academic misconduct, text recycling or 'self-plagiarism' occupies a particularly contentious position. Scientists and commentators agree on the undermining effects of infringements of core conventions in research, such as falsification, fabrication and plagiarism (FFP) as well as a series of questionable research practices (QRP). These have been widely studied, leading to a growing body of literature on the nature (Steneck, 2006), prevalence (Fanelli, 2009), causes (Fanelli, Costas, & Lariviere, 2015) and consequences (Steen, 2011; Zhang & Grieneisen, 2013) of scientific misconduct.

Text recycling raises interesting further questions about the nature and causes of misconduct. Academic text recycling is the reuse of one's own writing in academic publications, ranging from a sentence to several pages, or even entire articles, without reference. Some authors object to the term 'self-plagiarism', as stealing from oneself is a legal oxymoron (Callahan, 2014; Chrousos, Kalantaridou, Margioris, & Gravanis, 2012; Thurman, Chervenak, McCullough, Halwani, & Farine, 2016). Other terms, such as '(unacceptable) text recycling' (Moskovitz, 2016) or '(unacceptable) duplication' (Thurman et al., 2016), have been proposed. In this chapter, we use the term 'text recycling'.

Text recycling is one of the new forms of misconduct specifically aimed at gaming or abusing the current science reward system. Journal editors and research leaders have warned about the spread of misconduct and the emergence of these novel forms (Bohannon, 2013; Martin, 2013). They point out that authors and editors employ various practices specifically aimed at increasing their publication and citation records as well as their journal impact factors. These practices include faking peer review reports (Callaway, 2015), the forming of journal citation cartels (Mongeon, Waltman, & Rijcke, 2016) and coercive citation strategies (Martin, 2013).

Recently, the alleged 'self-plagiarism' by one of the most prominent Dutch scientists attracted much attention in the Netherlands. For years, economist Peter Nijkamp published peer-reviewed journal articles at the astonishing rate of about one and a half per week. In 2013, it became apparent that he had been recycling large parts of his previous work, triggering accusations of 'self-plagiarism' (Retraction Watch, 2014). The allegations and particularly the conclusions of the investigating integrity committees led to a fierce debate about authors' fair use of their previously published texts (Breedveld, 2015). Currently, it has become evident that many questions about text recycling have yet to be fully answered and that there is still much uncertainty. While fair text re-use has been more clearly specified, the extent and causes of improper recycling remain unclear. A few small studies have endeavoured to shed light on these issues, but provide contradictory results. Their estimates of 'self-plagiarism' range from 3% (Bazdaric, Bilic-Zulle, Brumini, & Petrovecki, 2012) to 60% (Bretag & Carapiet, 2007) — a wide range that feeds intense speculation (Binder, 1990). Some of these studies even indicate that 'self-plagiarism' is more common than plagiarism (Sun, 2013). Not only is there uncertainty about the extent of improper text recycling, but questions have also been posed about its causes and ways to avoid it (Honig & Bedi, 2012; Martin, 2013; Scanlon, 2007; Sun, 2013). No satisfactory answers have as yet been forwarded.

It is this knowledge gap that this study aims to address. As such, this study provides important new insights into the 'self-plagiarism' phenomenon, specifically its extent in various research areas and its potential causes. Using the Turnitin plagiarism detection software, we analysed 922 research articles spread over four scientific disciplines, followed by manual inspection of the results.

This chapter is divided into two parts. The first provides the background, including an overview of the literature on a wide spectrum of concerns and questions related to 'self-plagiarism' and scientific misconduct in general. This part also describes the Nijkamp case in the Netherlands as well as its nature and consequences for individuals and the research system as a whole, following the subsequent debate's arguments for and against authors' right to reuse parts of their earlier texts. Thereafter, we identify several questions raised by the Nijkamp case and, based on the literature, formulate hypotheses on the causes of (problematic) text recycling. The second part empirically describes the extent of text recycling in four research domains. It provides an account of the methodological approach (section 2.5), the results obtained (section 2.6) and a reflection on hypotheses regarding the causes of text recycling (section 2.7).

2.2 The academic debate on text recycling

Although improper text recycling was first mentioned in the academic literature in the early 1990s (Binder, 1990; Samuelson, 1994), major contributions to the discussion have only recently been made (Chrousos et al., 2012; de Vasconcelos & Roig, 2015; Harriman & Patel, 2014; Joob & Wiwanitkit, 2016; Martin, 2013; Moskovitz, 2016; Roig, 2010). Since the discussion on misconduct and plagiarism in science only started in the 1980s (Horbach & Halffman, 2016), text recycling is a relatively new concern. In the current debate on text recycling, most researchers focus on the damage to the reader who is 'deceived by false claims of originality' (e.g. de Vasconcelos & Roig, 2015). Only marginal attention is paid to the implications of text recycling for the scientific enterprise as a whole, thus ignoring the consequences for co-authors, fellow scientists and even society, such as unfair competition due to skewed rewards and the abuse of publication resources and reviewers' efforts (Tramer, Reynolds, Moore, & McQuay, 1997).

The debate on 'self-plagiarism' is mostly concerned with its permissibility, on which there is as yet no consensus. Multiple scholars' opinions range from regarding 'self-plagiarism' as 'a serious offence' and 'academic misconduct' (Bretag & Mahmud, 2009; Martin, 2013) to claiming that 'it does not exist' (Callahan, 2014) and is 'unavoidable' (Chrousos et al., 2012).

Several arguments suggest that text recycling is unacceptable. Re-publication of texts could be considered an abuse of the scientific publication system, which is arguably already congested with less than essential publications. The abuse is especially relevant given the reliance on reviewers who offer their time to assess work that has already been reviewed. However, these arguments only hold for the recycling of large sections of texts, as the reuse of smaller text fragments does not burden the publication system. The major argument against text recycling is that it is a form of gaming the science reward system, with text recycling scientists claiming more productivity than their work actually warrants. In a research system in which the number of publications is often considered an indicator of 'quality' as well as a career promotion and grant allocation instrument, text recycling is a way to boost scores at the expense of other researchers through unfair competition for grants or positions. (This raises the bigger question on the extent to which proxies based on publication productivity are meaningful assessment criteria for job or grant allocation.) Lastly, text recycling also has potentially harmful consequences for society, especially in biomedical research. Tramer et al. (1997) point out that duplicate reporting of a drug's effectiveness yields erroneous results in its meta-analyses. As a result, the estimates of the treatment efficacy might be biased, which could potentially harm patients.

Besides these arguments against text recycling, some other authors favour it (Callahan, 2014). In addition to the argument that authors cannot steal from themselves, the reuse of particularly well-formulated expressions of standard methods, disclaimers, or even nuanced theoretical positions could arguably be justified. Nevertheless, even in such cases a reference to an earlier work can be added easily. In addition, some authors claim that reusing one's own work is unavoidable, especially in small research fields in which authors build on their own line of research (Chrousos et al., 2012). Publishing similar results for different audiences has also been presented as a justification for reusing previously published material (Nijkamp, 2015; Samuelson, 1994). In the aftermath of the Nijkamp case, another argument was made in favour of text recycling, with economists claiming that it has become standard practice in their field and can therefore not be held against an individual (Nijkamp, 2015; Westlund, Martin, & Stough, 2014).

Owing to its contentious nature, text recycling holds a remarkable position in the current debate on integrity and misconduct in science. In the current integrity debate, there is a general tendency to assume a universal understanding of integrity, with demarcations along the spectrum ranging from 'responsible research practices' via 'questionable research practices' to 'scientific misconduct' (Horbach & Halffman, 2016; Steneck, 2006). Based on this assumed collective understanding, there has been an extensive effort to measure the prevalence and causes of integrity breaches (e.g. Fanelli, 2009; He, 2013; Martinson, Anderson, & de Vries, 2005; Steen, 2011). Despite many obstacles to obtaining accurate results, (e.g. due to the limits of self-reporting), several estimates have been given of the prevalence of questionable research practices (QRP) and misconduct (Fanelli, 2009; John, Loewenstein, & Prelec, 2012). These results generally indicate that the prevalence of QRP greatly exceeds that of scientific misconduct, namely FFP. In addition, various scholars suggest potential causes of misconduct in science. These include:

- Scientific age: younger scientists are frequently considered to be at greater risk of misconduct, due to their lack of experience with accepted norms and practices; the rise of the internet and the subsequent culture of using material from this in essays, theses and articles; and because young scientists do not yet have established reputations in the field and thus have more to gain than older researchers (Fanelli et al., 2015; Honig & Bedi, 2012). Consequently, measures to prevent misconduct or to foster integrity are commonly aimed at junior or future scientists (Godecharle, Nemery, & Dierickx, 2013; Horner & Minifie, 2011; Kornfeld, 2012; Necker, 2014; OECD, 2010).
- Research culture: it is commonly suggested that the academic culture, most notably the pressure to publish and the focus on quantity rather than quality, increases the chance of scientists engaging in misconduct (Anderson, Ronning, De Vries, & Martinson, 2007; Fanelli, Costas, Fang, Casadevall, & Bik, 2017; Van Dalen & Henkens, 2012).
- Number of authors: it has been suggested that the increase in articles' average number of co-authors increases the incidence of misconduct. Adding more co-authors to a paper is thought to dilute the responsibility of each author, thus allowing some authors to cut corners (Bennett & Taylor, 2003; Sun, 2013).
- Clear rules and policies: it is generally suggested that codes of conduct and formal regulations regarding misconduct and protocols for handling suspect cases deter misconduct (Fanelli et al., 2017; Godecharle, Nemery, & Dierickx, 2014). Conversely, a lack of such codes and, specifically, a lack of consensus on the definitions of dubious practices are considered possible source of such practices.
- Lack of social control: it is hypothesised that a (perceived) lack of social control might increase the extent of unjustified research practices (Bohannon, 2013; Enders & Hoover, 2004; Fanelli et al., 2017; Stroebe, Postmes, & Spears, 2012). This social control takes various forms, including peer review, editorial evaluation, mentoring and societal evaluation.

The suggested potential sources of scientific misconduct are generally based on linking self-reported cases of scientific misconduct with contextual background information. Owing self-reporting's limitations, the results tend to have high levels of uncertainty. In this respect, text recycling holds a specific position in the spectrum of dubious behaviour in science, since it is relatively easy to measure with modern plagiarism software. Using plagiarism detectors, we directly tested the literature's claims about this specific questionable research practice. Our study not only provides insights into several open questions regarding text recycling, its perceived permissibility, and its incidence, but it also offers insights into the potential sources of text recycling and scientific misconduct in general.

2.3 Raising concern: the Nijkamp-case

The controversy surrounding Peter Nijkamp's alleged text recycling is interesting for several reasons. It illustrates the tensions that arise when a questionable, but not formally regulated, practice is challenged – in this case in the media. In the Dutch context, the Nijkamp controversy triggered a debate about the acceptability of various degrees of text recycling, resulting in the articulation of levels of recycling considered problematic, which we also used in our assessment. The controversy also raised several questions about the incidence and causes of text recycling, to which we endeavour to contribute.

The Peter Nijkamp case is extremely rich, with allegations of misconduct involving plagiarism, self-plagiarism, as well as data fraud; media interest; legal procedures; and regulatory action. The many ramifications of this case warrant a full-length article, but for brevity's sake we limit ourselves to the aspects directly involving 'self-plagiarism'. In this section, we provide a description of those case elements related to text recycling, including the formal allegations, the various integrity committees' investigation processes, the committees' final conclusions and the case's consequences. This overview highlights how the different arguments used in favour and against text recycling work in the context of a concrete case, illustrating the complexities and contradictions that arose when various investigating committees came to differing conclusions.

Peter Nijkamp is Emeritus Professor of Regional Economics and Economic Geography at the Faculty of Economics of the Vrije Universiteit (VU) in Amsterdam. During his career, Nijkamp became widely known as one of the most productive economists in the world. From 1975 onwards, he published over 2,300 research articles and more than 100 edited volumes. In addition, his work is highly cited, with a total of nearly 40,000 citations, giving him an h-index of 88 and an i10-index of 752 (obtained from Google Scholar on 8 September 2016). His enormous productivity and influence were rewarded with the Spinoza Award, the most prestigious scientific award in the Netherlands, in 1996. In addition, he was president of the governing board of the Netherlands Research Council (NWO) from 2002 until 2009, chairman of the Dutch Social Science Council (SWR) and vice-president of the Royal Netherlands Academy of Sciences (KNAW). A more stellar career in Dutch academia would hardly have been possible.

2.3.1 The allegations

The Nijkamp case began in May 2013, with an anonymous whistle-blower's allegation of plagiarism in the dissertation of Nijkamp's PhD student Karima Kourtit — plagiarism from Nijkamp's work and from that of others. Although the allegation did not target Peter Nijkamp directly, it did concern his work, since he co-authored most chapters in Kourtit's thesis. More accusations would follow, all from the same anonymous source, with attention shifting to Nijkamp.

During the Nijkamp case, several integrity committees were responsible for the official handling of the allegations. The committee handling the first allegation found Nijkamp and Kourtit guilty of inappropriate text recycling. Interestingly, the official investigation report referred to it as "plagiarism" (VSNU, 2013), most probably due to the lack of clear and accepted terminology for this topic. According to this committee, reuse of earlier material occurred via a process it labelled 'self-citation'. The process – which differs from the usual understanding of 'self-citation' as authors' references to their already published papers – proceeds as follows: Author A and B write an article; later, author A, together with author C, publishes a new article in which, without reference, he uses material from the first article. At a later stage, author C writes a thesis using passages from the article authored by A and C, again without reference (VSNU, 2013). This process is, among others, problematic, because author B's contribution is never credited. The use of the unfortunate term 'self-citation' for this process was later attributed to administrative issues with the publishing of a summary of the committee's report (Schuyt, 2014)

In November 2013, shortly after the presentation of the findings of the integrity committee handling the first allegation, the anonymous whistle-blower made a second allegation regarding plagiarism and self-plagiarism in 16 publications by Kourtit, most of which Peter Nijkamp co-authored. This allegation was addressed to the VU ombudsperson. Later, allegations of data fraud were made, but these are not discussed in detail here.

The initial committee investigating the second allegation deemed it partly founded (LOWI, 2015), but Nijkamp was cleared of all charges in this allegation after appealing his case at the national integrity office (LOWI) (LOWI, 2015). The LOWI committee decided that Nijkamp and Kourtit were not guilty of (self)-plagiarism, because classifying an act as such requires 'a clear intention to deceive' (LOWI, 2015). According to the LOWI, this was not present due to, among others, the published articles' nature (a book review, in which 'overlap with the original source can be expected') and the amount of copied material (in the alleged cases, no more than several sentences).

Owing to the vigorous objections from scientists and the media, the executive board of Nijkamp's university decided to launch another investigation into his publication practices. A new committee was appointed, charged with investigating not only those publications that were part of a formal allegation, but also with studying the publication and citation practices in Nijkamp's entire oeuvre (Zwemmer, Gunning, & Grobbee, 2015). The committee sampled all his publications after 1970 and scanned a selection with plagiarism-detection software (Zwemmer et al., 2015). Owing to technical difficulties and time constraints, the committee decided to test 261 of the more than 2,300 articles in Nijkamp's oeuvre. Of the 261 scanned articles, they found that 60 had significant overlaps with his prior publications without referring to them. The committee judged that the amount of overlapping passages gave the impression of 'systematic copy-pasting', which served as a strategy to increase the author's number of publications and could not be construed as part of an 'original oeuvre' (Zwemmer et al., 2015). Nijkamp criticized the committee's work harshly for, among others, its strategy of mechanically using plagiarism-detection software with only little human verification and interpretation of the results (Nijkamp, 2015).

Ultimately, Nijkamp was found responsible for committing 'self-plagiarism', but was cleared of all other charges. Multiple committees investigated most of the allegations in this case and frequently came to diverse conclusions. Some committees required a clear 'intention to deceive' in order to label a practice scientific misconduct, whereas others did not. In addition, some committees took the material's specific context into account, such as the type of article or the part of the article that contained recycled text, whereas others adhered to more strict definitions of (self-)plagiarism, more or less ignoring the context.

2.3.2 The consequences

The findings of the committees in the Nijkamp case had consequences for the actors involved as well as for the Dutch research system. In the end, Peter Nijkamp's university did not officially sanction him, but he suffered major reputational damage. The case drew a lot of attention (280 newspaper articles and multiple blogs) and created a large public outcry about the scientist (as well as the person) Peter Nijkamp. In addition, The Review of Economic Analysis retracted two of his papers on the grounds of 'self-plagiarism' (Retraction Watch, 2014), demonstrating that at least some people consider text recycling a severe form of misconduct.

Part of the confusion regarding the severity of text recycling might have been due to the lack of formal policy on this issue. In response to the Nijkamp case, the Royal Netherlands Academy of Arts and Sciences (KNAW) published an advice on correct citation practices (KNAW, 2014). The advice commented on plagiarism and, more specifically, on 'self-plagiarism'. In response, a specific paragraph on text recycling was incorporated into the Netherlands Code of Conduct for Academic Practice (VSNU, 2014;Schuyt, 2014). The Dutch were thus among the first to incorporate regulations regarding text recycling into their national policy statements (de Vasconcelos & Roig, 2015).

Besides the statement in the Netherlands, the Committee on Publication Ethics (COPE) also published a set of guidelines regarding text recycling (Harriman & Patel, 2014). Both policy statements agree that permissible reuse of an author's own material is highly dependent on the circumstances. They agree that reuse is allowed if "it concerns brief passages of introductory, theoretical or methodological explanation" (KNAW, 2014). However, reuse of parts of the results, conclusion or discussion sections are, in general, not allowed (Harriman & Patel, 2014). Both policy statements stress that reused passages should never create the suggestion that they constitute novel contributions and should always be properly referenced. Other codes of conduct stress the need to be open also about recycling material in, among others, an author's CV or list of publications (ESF/ALLEA, 2011).

Despite controversies about the permissibility of textual recycling, the new policy statements by VSNU and COPE acknowledge that, whatever the case, there are problematic forms of text recycling. We worked with the current Dutch criteria to identify such problematic text recycling. In the remainder of this chapter, we will address the extent to which such forms of text recycling appear in Dutch authors' publications and suggest reasons for their occurrence.

2.3.3 Questions raised by the Nijkamp case

Despite the many investigation committees studying the Nijkamp case and scientists' extended (national) debate, many questions are still unanswered. One of Peter Nijkamp's arguments in his defence was that he had not deviated from common practices and behaviours in his research area. In an open letter to the VU's rector, his fellow economists endorsed this argument, claiming that all of them would be guilty if subjected to the same norms (Westlund et al., 2014). This raises the question of how common inappropriate text recycling is and whether there are major differences between economics and other research fields. Does it specifically concern economists or is it more widespread? Secondly, many considered the act of text recycling as one of the explanations for Nijkamp's enormous productivity (Zwemmer et al., 2015). This raises the question whether productive scientists recycle text more often than their less productive colleagues do. Thirdly, unacceptable text recycling was argued to be due to a lack of clear rules and guidelines on this topic. This leads to the question whether guidelines on text recycling are an effective means of preventing it.

2.4 Testable claims about text recycling

In the literature and in the context of the Nijkamp controversy, several claims were made about the extent and causes of questionable research practices, and particularly about problematic text recycling. Given the limited amount of research on this topic, some claims were not substantiated, others were, and yet others used as a defence or as an accusation. Below are the claims we could test with our approach and from which we generated our research hypotheses.

2.4.1 Causes of (problematic) text recycling

Several causes of text recycling have been proposed. We will discuss these causes, categorising them as ‘individual causes’, which refer to aspects related to the individual author(s), and ‘systemic causes’, which refer to aspects of the research system. The discussion of these potential causes leads to several hypothesis on the occurrence of problematic text recycling, which will be further discussed in section 2.7.

Individual causes

Several causes of text recycling related to a specific author’s position and identity have been proposed. Firstly, multiple scholars suggest that articles’ number of authors influences their chance of containing plagiarised material. It has been suggested that the responsibility of each author is diluted when the number of authors is increased, which increases the chance of (self-)plagiarism (Bennett & Taylor 2003; Sun 2013).

Hypothesis 1: a high number of article authors increase the likelihood of problematic text recycling.

Secondly, it has been claimed that scientific age or career stage (i.e. the scientific career and position’s level of maturity) influences the likelihood of committing (self-)plagiarism (Fanelli et al., 2015; Honig & Bedi, 2012). Various academics point out that junior researchers have more incentives to (self-)plagiarise than senior researchers, because they have more to gain. In addition, it is believed that “ … graduate and post-doctoral students, […] are not aware of the problem or […] have trouble writing with ease and speed and feel that taking some material from here and there is something that won’t be noticed” (O'Hair & Neff, 2013). It is therefore hypothesized that (self-)plagiarism occurs more frequently among junior researchers (Honig & Bedi, 2012). This does not mean that scientists in the later phases of their careers do not engage in text recycling. Several scientists accused of ‘self-plagiarism’, such as Nijkamp, Breslow (Oransky, 2012) and Wansink (Chambers & Etchells, 2017), were in the later phases of their careers. Based on these cases, the contrary might also be plausible, but, based on the scientific literature, we hypothesise:

Hypothesis 2: authors of a young scientific age are more likely to recycle text improperly.

Systemic causes

We identify two proposed reasons for problematic text recycling related to the research system and policy.

Firstly, some scholars suggest a correlation between the clarity of journals or publishers' set norms and guidelines and the extent of text recycling in various research areas. It has been suggested that clear policy is needed to make authors aware of what is allowed and what not. The absence of clear policy might therefore increase the extent of problematic text recycling (Karabag & Berggren, 2012; Martin, 2013). While academies and other science-governing organisations' rules are relatively new, several journals have long had editorial policies against text recycling. Nevertheless, the Nijkamp controversy suggests that a lack of clear guidelines could lead to text recycling.

Hypothesis 3: a lack of clear editorial statements on text recycling increases the likelihood of it occurring.

Secondly, variations between research fields, which occur on multiple levels, might be relevant in this context. Firstly, some scholars have suggested that text recycling in the humanities is a more serious offence than in the natural sciences, because "the wording is the essence of the novelty" in the humanities (Chrousos et al., 2012). Based on this reasoning, the humanities could be expected to have more strict conventions about text recycling and, hence, lower incidence rates than other scientific domains. Similarly, the recycling of (highly technical) research protocols in the natural sciences is often considered less problematic, partly because their language is highly standardised. This suggests that the natural sciences might have substantial recycling incidences.

In addition, other than official policies on (self-)plagiarism, journal editors' willingness to address (alleged) cases of (self-)plagiarism might have an influence on the frequency with which recycled articles are published. Further, the (perceived) level of social control via editorial evaluation could potentially influence authors' writing strategies (Bennett & Taylor, 2003; Fanelli et al., 2017). Editors' crucial role in fostering research integrity and in maintaining the scientific literature's integrity has been commonly stressed (Council of Science Editors, 2012; Marusic, Katavic, & Marusic, 2007). It is commonly accepted that editors can (and should) not act as 'the policing force of the scientific community' (Marusic et al., 2007), but that they should nonetheless be proactive in fostering research integrity (Council of Science Editors, 2012). Enders and Hoover (2004) show that editors of (top) economic journals do not seem to be particularly strict with text recycling cases, while de Vasconcelos and Roig (2015) conversely argue that journal editors are generally very averse to unacceptable text recycling. After a survey of journal editors, Wager et al. confirm the latter view, finding that editors are very concerned about duplicate publication and plagiarism, which they regard as major threats to science's integrity (Wager, Fiack, Graf, Robinson, & Rowlands, 2009). It has been suggested that if editors are not willing to address improper behaviour, including unacceptable text recycling, authors may only suffer minor consequences. From this 'rational actor' perspective, authors could be tempted to recycle text if there are no sanctions. The combination of these factors leads us to our last hypothesis:

Hypothesis 4: problematic text recycling is more common in research disciplines in which phrases are more standardized and editors are less willing to address recycling.

2.5 Methods

2.5.1 Data collection and analysis

We selected four research areas to measure the extent of problematic text recycling and to study the disciplinary differences between various research domains. The selected domains are: biochemistry & molecular biology, economics, history and psychology. These domains cover a wide spectrum of research and can offer a perspective on the research community's diversity. Since multiple scholars have found that publication cultures, competitive pressures and policy formulations are highly heterogeneous in different countries (Fanelli et al., 2015; Godecharle et al., 2014), we decided to focus our research on a single country. Owing to the great uproar about 'self-plagiarism' after the Nijkamp case in the Netherlands, we decided to focus our study on authors affiliated with Dutch universities.

We collected research articles by authors in the specified research areas. We grouped the sampled articles into the categories 'productive' and 'less productive' authors to study the influence of the authors' productivity on text recycling. The sample articles were scanned using the Turnitin plagiarism detection software (Turnitin, 2006). The results were subsequently subjected to full-text inspection to allow us to delete all forms of acceptable overlap. We provide detailed descriptions of the data collection and the analysis techniques below.

2.5.2 Collection of research papers per research area

Biochemistry: We used the Web of Science to create a list of the most productive scientists affiliated with a Dutch university (search on research area = biochemistry & molecular biology, country = Netherlands, and timespan = 2010-present). Using Sciencedirect, we retrieved papers by the top six authors on the list. To identify the less productive authors, we searched Sciencedirect by research area for papers (biochemistry and molecular biology) and by authors' affiliation with Dutch universities, and then only selected papers in which none of the authors were ranked in the top 25 of the Web of Science productivity ranking. All the selected papers were published between 2010 and November 2016, and were selected on the basis of their publication date (newest papers first). We omitted Sciencedirect entries comprising announcements of upcoming papers in subsequent issues, or short article abstracts.

Economics: As with the previous research area, we only used the most productive Dutch economists as listed in the 2013 Economische en Statistische Berichten (ESB) list of top economists ("economentop 40") (Phlippen, 2013). This list provides a more accurate overview of Dutch economists' productivity than the Web of Science. We then again searched Sciencedirect for the top six authors on this list (for the 'productive' category) and for economists affiliated to Dutch universities, selecting only those papers that did not include any author in the top 40 (for the 'less productive' category).

History: We used the same process as we did for biochemistry, with a slight alteration regarding the search engine due to Sciencedirect's lack of history articles. Research articles were not collected via Sciencedirect, but via Google Scholar and the database of the Dutch history journal Nederlands Tijdschrift voor Geschiedenis (all articles from 2010 onwards). We categorised the articles according to productive (top of the list) and non-productive authors (not in the top 25) as indicated by the Web of Science.

Psychology: Similar to the biochemistry process, with obvious changes in search terms from 'biochemistry & molecular biology' to 'psychology', while searching both the Web of Science and Sciencedirect.

We sampled between 125 and 135 articles from all the areas and from the 'productive' category as well as the 'less productive' category. We could only select 50 articles by productive authors in the time period 2010-present due to the lack of a large database and the general low number of articles in the research area 'history'. In all the cases in which articles by a specific author were sampled, we used the spelling of the author names used in the ESB's productivity list (in the case of economics) or the Web of Science (in the other cases).

2.5.3 Collection of policy guidelines

We collected academic journals' editorial policy guidelines to analyse hypothesis 3. We distinguished three categories of journals: journals with the highest impact factor in their research area ('top journal'), journals in which our sample of articles most frequently appeared ('most frequent'), and journals that published one of our sample articles with problematic text recycling ('containing problematic recycling'). A journal was acknowledged as containing statements on plagiarism if these refer to the inclusion of previously published work, or if they specifically use the word 'plagiarism'. If a journal specifically points out that authors are also not allowed to copy text from their own previously published work, it was classified as having statements on 'self-plagiarism'. These guidelines were collected via the journals' and the publishers' webpages, because authors are directed to them when submitting a manuscript and they provide instructions for preparing manuscripts.

2.5.4 Analysis

We uploaded all the documents to the plagiarism detection software Turnitin (Turnitin, 2006). This software provides three levels of scanning strictness: 'compliant', 'standard' and 'strict'. We used the 'standard' level for all our investigations. The collected articles were tested for textual overlap against an internal Turnitin database comprising more than 62 billion webpages and 165 million scientific articles (Turnitin, 2006). Hence, the sampled articles were not only tested against other articles in our sample, but also against a wide variety of journal articles and other sources, including books and book chapters.

We subsequently checked all the individual papers' passages flagged as overlapping by inspecting the relevant articles. Our aim was to avoid the fierce criticism of mechanical analysis, which the Nijkamp case had engendered (Nijkamp, 2015). This process also distinguishes our work from previous research on text recycling (e.g. in Sun, 2013). In agreement with the COPE and VSNU new policy statements (Harriman & Patel, 2014; VSNU, 2014), the following were considered acceptable overlaps:

- Full texts or abstracts of the same articles in the original journal or electronic database
- Author information (name, address, affiliations, etc.)
- List of references
- Overlaps correctly cited with reference to the original
- Overlaps with one of the authors' master or PhD thesis
- Overlaps with unpublished working papers
- Overlaps with articles published after the studied article's publication date (in order to ensure no overlap was counted twice)
- Overlapping (technical) descriptions in the methodological sections of the paper containing a reference to a previous paper.

Deciding on the appropriateness of textual overlap was not always straightforward. The fourth criterion, claiming that an overlap is appropriate if it is correctly referenced, specifically led to discussion. In such cases, we decided to label the overlap as appropriate, thereby rendering all contentious cases 'unproblematic'. Consequently, the resulting analysis is a conservative estimate of the extent of text recycling.

Articles were considered problematic if, after deletion of all acceptable overlap, they contained at least 10% passages identical to those in previously published articles. The threshold of 10% was chosen in keeping with previous research on text recycling (Bretag & Carapiet, 2007; Bretag & Mahmud, 2009). The data collection and analysis are schematically depicted in the graphical abstract.

In some cases, the Turnitin software was unable to scan an uploaded article, for example, due to particularly troublesome text formats. Table 2.1 shows the number of articles we eventually retained. We stored information on all the articles regarding the research area, the category (productive/ less productive), the journal in which it was published, the year in which it was published, the number of authors and the extent of text recycling (problematic if above 10% after close study, unproblematic otherwise). In addition, table 2.1 provides information on the number of first authors in our analysis. (Section 2.6 presents a more specific analysis of the number of authors in our sample.)

Research area	Category					
	Productive		Less productive		Total	
	Number of articles	Number of 'first authors'	Number of articles	Number of 'first authors'	Number of articles	Number of 'first authors'
Biochemistry	128	104	112	107	240	211
Economics	133	78	125	106	258	184
History	48	25	142	127	190	152
Psychology	125	91	109	103	234	194
Total	434	298	488	443	922	741

Table 2.1: number of articles scanned per research area and category

2.6 Results

This section presents the results of the empirical analysis described in section 2.5. The presentation of the results is based on a discussion of figure 2.1 and tables 2.2 to 2.4, which provide information on text recycling (table 2.2), its connection with policy statements (table 2.3) and its connection with the number of authors (table 2.4). Section 2.7 presents the conclusions drawn from these results and a reflection on the hypotheses in section 2.4.

Research area	Number of articles			Number of articles containing problematic text recycling			Percentage of articles containing problematic text recycling		
	Productive	Less productive	Total	Productive	Less productive	Total	Productive	Less productive	Total
Biochemistry	128	112	240	6	2	8	4.7%	1.8%	3.3%
Economics	133	125	258	27	9	36	20.3%	7.2%	14.0%
History	48	142	190	1	0	1	2.1%	0.0%	0.5%
Psychology	125	109	234	10	1	11	8.0%	0.9%	4.7%
Total	434	488	922	44	12	56	10.1%	2.5%	6.1%

Table 2.2: the extent of recycling in various research areas by productive and less productive authors affiliated to Dutch universities

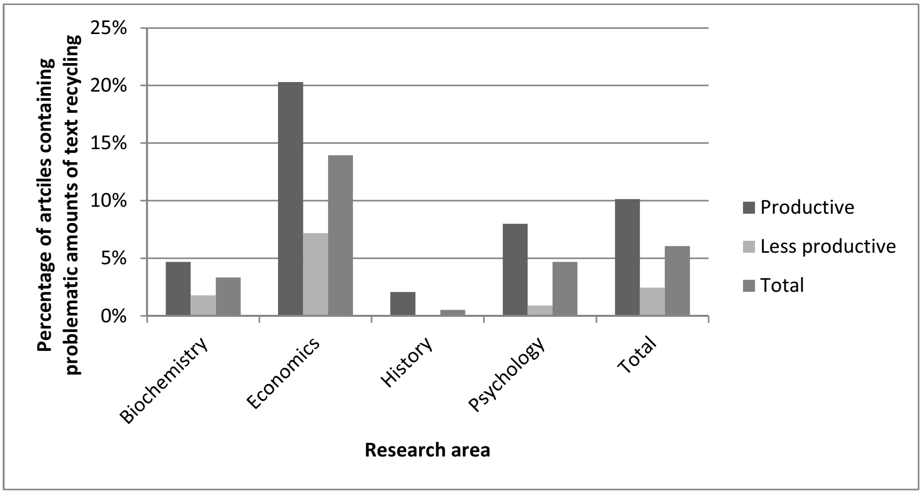


Figure 2.1: the extent of problematic text recycling in various research areas by productive and less productive authors affiliated to Dutch universities

Table 2.2 and figure 2.1 present an overview of the extent of problematic text recycling in the various research areas by productive authors and their less productive colleagues. In the table, ‘articles containing problematic text recycling’ refers to the articles deemed ‘problematic’, as defined in section 2.5.

Table 2.2 presents several interesting results. Firstly, it indicates the general extent of problematic text recycling, which occurs in slightly more than 6% of the entire sample.

Secondly, the table's last column demonstrates the major differences between the extent of text recycling in the various research areas. Whereas text recycling seems fairly common among economists, it is hardly found among historians. At least 10% of the text in as many as one in seven publications by Dutch economists has been published before. In fact, over 40% of some economists' articles contain problematic forms of text recycling when assessed according to current standards.

Lastly, it is apparent from the table that the degree of text recycling differs between productive and less productive authors. Productive authors recycle their previous work significantly more often than their less productive colleagues (10.1% vs. 2.5%, $p < 0.0001$). Moreover, this general pattern is visible in all of the individual research areas.

Research area	Top journal			Most frequent			Containing problematic recycling		
	Tot	plagiarism	text recycling	Tot	plagiarism	text recycling	Tot	plagiarism	text recycling
Biochemistry	10	9	6	5	5	1	3	3	1
Economics	10	4	4	5	5	0	17	15	0
History	10	3	3	5	2	2	1	0	0
Psychology	10	8	3	5	5	0	8	8	0

Table 2.3: policy statements on (self-)plagiarism in scientific journals' policy guidelines. Three categories of journals are distinguished: journals with the highest impact factor in their research area (top journals), journals in which our sample of articles were most frequently published (most frequent) and journals that published at least one of our sample articles containing self-plagiarism (containing self-plagiarism).

Table 2.3 demonstrates the frequency with which scientific journals have specific statements on (self-)plagiarism in their policy guidelines, which was identified with the procedure described in section 2.5.3. The table shows that statements on text recycling are rather uncommon in journals' policy guidelines, as the vast majority of journals in our sample do not explicitly mention text recycling. Statements on text recycling are more common in influential journals (those with high impact factors). The journals in which our sample are most commonly published and those publishing recycled articles almost uniformly lack statements on text recycling. In contrast, nearly all journals prohibit plagiarism.

Assuming that the most influential journals represent common practices in their research area, we conclude that the fields of biochemistry and psychology pay most attention to (self-)plagiarism.

Research area	All articles		Articles with problematic recycling		Articles without problematic recycling		Total		
	Productive	Less product.	Productive	Less product.	Productive	Less product.	all	Problem. recycling	No prob. recycling
Biochemistry	7.98	6.12	6.17	4	8.07	6.15	7.11	5.63	7.16
Economics	3.22	2.87	3.09	3.22	3.26	2.84	3.05	3.12	3.04
History	1.06	1.24	1	0	1.06	1.24	1.19	1.00	1.20
Psychology	5.52	3.83	4.18	5	5.68	3.82	4.73	4.25	4.76
Total	5.06	3.57	3.68	3.5	5.21	3.35	4.16	3.64	4.19

Table 2.4: average number of authors per article in various categories of our sample

Table 2.4 presents an analysis of the average number of authors in various categories of our sample. It provides information on the number of authors of either recycled or non-recycled articles and either falling into the category of productive or less productive authors. From the analysis, we conclude that articles containing problematic text recycling have, on average, more authors than those not containing problematic recycling. This holds for all research areas, except economics, in which the values are nearly identical. Analysis of the entire sample demonstrates that self-plagiarised articles have significantly fewer authors than non-self-plagiarised articles (3.64 vs. 4.19, $p < 0.05$). Remarkably, this pattern is clearest in the productive authors category. In this category, the articles containing problematic recycling have far fewer authors than those not containing problematic recycling (3.68 vs. 5.21, $p < 0.005$).

Lastly, we would like to mention that our methodological approach enabled us to not only find cases of text recycling by the authors, but also by other scholars. Hence we could not only track 'self-plagiarism', but also actual 'plagiarism'. However, whereas we found 56 cases of problematic text recycling by the original article author, we only found one case that tended towards plagiarism. In this case, it was one of the articles in our sample plagiarised by authors not affiliated with Dutch universities and, hence, outside our sample.

2.7 Conclusion

This research set out to study the extent of problematic text recycling in various categories of scientific authors. The results of our analysis indicate problematic text recycling in 6.1% of the articles published by authors affiliated with Dutch universities. However, the results show strong differences between the extent of problematic text recycling in various research areas, with high rates in economics and very low rates in history. In addition, the results demonstrate that text recycling is significantly more common among productive authors than among their less productive colleagues (10.1% vs. 2.5%, $p < 0.0001$).

In the remainder of this section, we will comment on the hypotheses on the occurrence and causes of problematic text recycling as derived from the literature in section 2.4.

2.7.1 Number of authors

The first hypothesis states that a high number of article authors increases the likelihood of text recycling. However, our results show the contrary. On average, the 56 articles in our sample containing problematic recycling have 3.66 authors, whereas the 866 articles not containing problematic recycling have 4.18 authors on average. Further, if we consider the individual research areas, the same pattern appears: the number of authors of recycled articles vs. the number of authors of articles not containing recycled text is 5.63 vs. 7.16 (biochemistry), 3.12 vs. 3.04 (economics), 1.00 vs. 1.20 (history) and 4.25 vs. 4.76 (psychology). In addition, this pattern is clearest in the productive authors category. In this category, the articles containing problematic recycling have significantly fewer authors than those not containing problematic recycling (3.68 vs. 5.21, $p < 0.005$). Again, this difference is also evident in the individual research areas, with problematically recycled articles written by productive authors in biochemistry having, on average, 4.00 authors compared to 6.18 authors of those without problematic recycling. In psychology and economics, the ratios are respectively 4.18 vs. 5.38 and 3.09 vs. 3.26. Hence, we argue that a high number of authors reduce the likelihood of committing problematic text recycling. A possible explanation could be that a high number of authors amplify the internal control of the content and the origin of a manuscript.

Based on the results, we argue that productive authors publishing a manuscript with only few co-authors are a category specifically at risk of recycling their previous work.

2.7.2 Scientific age

The second hypothesis predicts more problematic text recycling by authors of a young scientific age. Our data, however, show that productive (in all cases senior) researchers show significantly more signs of text recycling than less-productive (often more junior) researchers. Our data therefore do not support the hypothesis.

We surmise that senior researchers might recycle text more frequently for multiple reasons. Firstly, cynicism about the research and peer review system is identified as one of the factors enhancing the willingness to commit academic misconduct (Clair, 2015). Senior researchers might have grown more cynical about the system than junior researchers. Consequently, they might be more likely to participate in dubious behaviour in general and undertake text recycling in particular.

Secondly, senior scientists might be more confident or aware of the low probability of being caught out. From a rational actor perspective on committing misconduct, a researcher is less likely to participate in dubious behaviour if he perceives the consequences as severe. In theory, the consequences of committing (self-)plagiarism are severe, ranging from dismissal to irreparable reputational damage. However, in practise, these consequences or penalties are hardly ever applied (Hoover, 2006). Given that the definitions of 'self-plagiarism' are widely contested, it is extremely difficult for any agency, journal, editor, or whatever institution to build a successful case against a 'self-plagiarist'. In addition, there is no consensus on whose responsibility it is to act against a 'self-plagiarist'. Moreover, even clear-cut cases of (self-)plagiarism usually go unexposed and unpunished (Hoover, 2006).

Senior researchers might also be more aware that, in practice, the consequences of recycling text are minimal. This provides a possible explanation for senior researchers more commonly committing text recycling, which they might perceive as one of the more 'safe' options of shortcutting the academic system's pressures.

2.7.3 Editorial policy statements

The third hypothesis concerns the relation between editorial policy and the extent of problematic text recycling, stating that the absence of clear policy statements on text recycling increases its likelihood. Our data present only limited evidence for this claim. In this regard, we distinguish between (a) policy statements in high impact journals arguably setting the tone for their research area and (b) policy statements in the journals in which a specific paper is published.

In the first case, the existence of editorial policy statements does not seem to have a clear correlation with the extent of problematic text recycling. Journals in the area of economics do not very actively publish policy statements on (self-)plagiarism, while journals on history are even less active. The latter journals hardly ever include any statement about the acceptability of (self-)plagiarism in submitted articles or how they respond to such cases. Nevertheless, articles by historians show very little sign of text recycling, whereas economists demonstrate a significantly higher extent of problematic text recycling. In addition, journals in the field of biochemistry are most likely to include statements on (self-)plagiarism in their editorial policy instructions, but biochemists do not show the lowest extent of problematic text recycling in their articles.

In contrast, we notice that, of all the journals in our sample that published an article containing recycled text, nearly all had statements about plagiarism in their policy guidelines, while almost none had specific statements concerning text recycling. In addition, of all the journals present in our sample, journals on history (the area with the lowest extent of problematic text recycling) most commonly published statements on text recycling in their editorial policy guidelines. This suggests that statements in the policy report of the journal in which an article is published decrease the likelihood of authors recycling their texts. However, such statements might also indicate that editors of these journals are keener to detect and reject recycled manuscripts, a topic which will be discussed more in depth below.

2.7.4 Willingness of editors

Lastly, the fourth hypothesis asserts that a high level of standardized language in research areas and editors' lacking willingness to act against (alleged) cases of problematic text recycling increase the likelihood of its occurrence. Our results partly support this hypothesis. As expected, the extent of problematic recycling by historians, who are part of the humanities, is very low. This is in line with the hypothesis and is due to the conviction that 'wording is the essence of novelty' in this research area. This reasoning predicts a high incidence of text recycling in biochemistry, due to the high level of standardization in the language used for research protocols and methods sections. Contrary to this hypothesis, however, the incidence of problematic text recycling in biochemistry is relatively low.

With respect to editors' perceptions, Enders and Hoover (2004) show that editors of (top) economic journals do not seem to be very strict regarding plagiarism cases, whereas De Vasconcelos & Roig (2015) argue that, in general, journal editors are very keen on addressing 'self-plagiarism'. A survey of (mainly) medical journal editors demonstrated that redundant publication and plagiarism are their primary concern regarding publication ethics (Wager et al., 2009). Although no specific data are available for the attitude of journal editors in the other research areas, the high frequency of text recycling and, in the area of economics, editors' arguable lack of willingness to address it, support the hypothesis. In addition, in paragraph 7.3 we already discussed that text recycling tends to occur relatively often in journals with no editorial policy statements on this form of misconduct. Arguably, the editors of these journals are relatively unconcerned about 'self-plagiarism' and are less keen to address it.

Consequently, our data suggest that journal editors' perceptions of the severity of text recycling and their willingness to address it have a major influence on the frequency of text recycling in published journal articles. However, a thorough study of journal editors' perceptions is required to address this topic further.

2.8 Discussion

Judging by the results of our study, the inappropriate reuse of textual material in research is definitely a form of misconduct that deserves serious attention and consideration. Whereas, in the current debate on textual reuse, the focus is primarily on the reader who is 'deceived by false claims of originality', we argue that, beyond this, text recycling should concern the research system as a whole. The inappropriate reuse of prior research puts science's current reward system under stress, potentially disrupting the system by harming various actors, such as co-authors, colleagues, reviewers and editors.

Our study suggests that the rates of problematic text recycling are substantial. Occurring in more than 6% of our sample articles, it seems to be significantly more common than other, more serious, forms of misconduct, such as plagiarism, falsification and fabrication (Fanelli, 2009). In addition, we identified several causes and risk factors that increase the extent of problematic text recycling, thereby suggesting potential measures to avoid this. Such measures include the confirmation and effective implementation of rules in journal policies, enhanced social control of authors, attention to prolific authors' publication practices, in addition to increasing young researchers' awareness of research integrity. These measures may all serve to either actively prevent or quickly detect the improper recycling of previously published text.

Besides indicating specific risks factors, the results of our study also identify potential differences in scientific disciplines' publication cultures. This suggests that a one-size-fits-all approach to preventing improper text recycling may well be too disrespectful of the diversity of practices in the various research fields. Consequently, specific measures should be sought for specific disciplines.

The Nijkamp case clearly highlights a need for specific measures and the shortcomings of a one-size-fits-all approach. In the Netherlands, this case demonstrated the rather contentious nature of text recycling in academic publishing, with even various integrity committees judging similar cases of text recycling differently. Specifically, the case points out several aspects of the published material that might be subject to debate, such as the type of article being published, the section of the article containing recycled material and the presence of a clear intention to deceive. In addition, the results of our analysis confirm the statements expressed in the Nijkamp case that text recycling is a rather common phenomenon in some academic disciplines. This seemingly widespread occurrence of text recycling, the variations between research fields, and text recycling being openly criticized (and sanctioned) by some, while accepted (or even promoted) by others, may require further discussion and carefully tailored measures.

A number of factors may limit this study's findings. Firstly, different research disciplines exhibit different publication practices. For example, books and book chapters are more common media for history publication than for the biomedical sciences. In our research, we limited the sampling of text to academic journal publications, thereby potentially omitting other relevant forms of academic publishing. We stress, however, that the majority of these sources (such as books and book chapters) are present in the Turnitin database against which our sample of texts was tested. Consequently, potential overlaps with these sources were visible in our analysis.

Secondly, the qualitative step in our analysis, in which textual overlap was either classified as appropriate or non-appropriate, leads to conservative estimates of the extent of text recycling. Furthermore, this step could have introduced some variation between the research areas. This might have occurred, for example, because 'grey' cases of text recycling — when it was not obvious whether the recycling was problematic or not — were all labelled 'unproblematic'. However, given the large number of sampled papers and the fact that no systematic differences appeared between the research areas in this respect, we are confident that such variations were kept to a minimum.

Lastly, our analysis focuses specifically on text recycling by researchers at Dutch academic institutions. While this limitation was necessary to keep the parameters of the specific national context stable, this does raise questions regarding the generalisability of our findings. Relevant differences between the Netherlands and other countries might take the form of different publication practices and different levels of competitive pressures. In respect of publication practices, the Netherlands was one of the first countries to articulate national policy statements on the permissibility of text recycling. In the light of our results, which suggest that a formal policy reduces the extent of unjust text recycling, the extent of text recycling in other countries could be even higher. In addition, various analyses of publication practices and research culture show that researchers in the Netherlands are not at a higher risk of engaging in misconduct or questionable research practices (Fanelli, 2016; Fanelli et al., 2015).

Text recycling is particularly pernicious where research funding is allocated between university departments on the basis of productivity indicators, as in some parts of Dutch academia. While extensive text recycling may seem a mere quirk of a particular field's publication culture, simple output indicators will over-estimate the productivity of research groups in fields with high levels of text recycling, skewing the allocation of resources in their favour.

With the publication system gradually turning into one large 'meta-journal' in which articles are increasingly available to all scholars, the act of text recycling is becoming rather silly. If previously published material is easily accessible, the need to reuse text is quickly diminished, as a reference to the original material can be easily added. Hence, the act of recycling one's own text arguably serves little purpose other than boosting one's publication record. Our results thus also reaffirm that assessing quality by means of productivity is problematic and may give rise to undesirable gaming. As has been previously shown, any performance measurement has a limited lifespan: it will cease to be effective after some time, because professionals learn to play with it, or because the beneficial effects of performance measurement are realized or wear off (De Bruijn, 2002; Hicks, Wouters, Waltman, De Rijcke, &

Rafols, 2015). Substantial text recycling may, among many others, be considered an indication that the era of the current science reward system is reaching its limit.

2.9 Postscript

Since the publication of this chapter as article in *Research Policy*, several developments have taken place both nationally and internationally regarding academic text recycling. Both in the Netherlands and in Europe, new codes of conduct for research integrity have been published that now explicitly mention text recycling as a problematic research practice (ALLEA, 2017; KNAW et al., 2018). Even though both guidelines are not very elaborate in detailing the conditions under which recycling is (not) acceptable, they do state that such practice is not acceptable at least in some cases.

Besides, several actors have critically responded to our article, questioning our methodology and demarcation of ‘problematic’ text recycling. I will come back to this issue in the epilogue of this dissertation and refer to (Lukkezen, 2019) and (Horbach & Halffman, 2019) for an example of such criticism and our response to it.

2.10 References

ALLEA. (2017). The European Code of Conduct for Research Intregrity - Revised Edition. In A.-A. E. Academies (Ed.). Berlin, Germany.

Anderson, M. S., Ronning, E. A., De Vries, R., & Martinson, B. C. (2007). The perverse effects of competition on scientists’ work and relationships. *Science and Engineering Ethics, 13*(4), 437-461.

Bazdaric, K., Bilic-Zulle, L., Brumini, G., & Petrovecki, M. (2012). Prevalence of Plagiarism in Recent Submissions to the Croatian Medical Journal. *Science and Engineering Ethics, 18*(2), 223-239. doi: 10.1007/s11948-011-9347-2

Bennett, D. M., & Taylor, D. M. (2003). Unethical practices in authorship of scientific papers. *Emergency Medicine, 15*(3), 263-270.

Binder, C. (1990). SELF-PLAGIARISM AND MULTIPLE PUBLICATIONS - NEGLIGENCE OR EPIDEMIC DISEASE. *Acta Endocrinologica, 122*(1), 7-7.

Bohannon, J. (2013). Who's Afraid of Peer Review? *Science, 342*(6154), 60-65.

Breedveld, P. (2015). Leidse hoogleraar wil internationale discussie over Nijkamp. from http://www.advalvas.vu.nl/nieuws/leidse-hoogleraar-wil-internationale-discussie-over-nijkamp

Bretag, T., & Carapiet, S. (2007). A Preliminary Study to Identify the Extent of Self-Plagiarism in Australian Academic Research. Plagiary: cross-disciplinary studies in plagiarism, fabrication, and falsification. *MPublishing, University of Michigan Library, Ann Arbor, 2*.

Bretag, T., & Mahmud, S. (2009). Self-Plagiarism or Appropriate Textual Re-use? *Journal of Academic Ethics, 7*(3), 193-205. doi: 10.1007/s10805-009-9092-1

Callahan, J. L. (2014). Creation of a Moral Panic? Self-Plagiarism in the Academy. *Human Resource Development Review, 13*(1), 3-10. doi: 10.1177/1534484313519063

Callaway, E. (2015). Faked peer reviews prompt 64 retractions. *Nature News*.
Chambers, C., & Etchells, P. (2017). Fresh concerns raised over academic conduct of major US nutrition and behaviour lab. *The Guardian*. Retrieved from https://www.theguardian.com/science/head-quarters/2017/mar/02/fresh-concerns-raised-over-academic-conduct-of-major-us-nutrition-and-behaviour-lab
Chrousos, G. P., Kalantaridou, S. N., Margioris, A. N., & Gravanis, A. (2012). The 'self-plagiarism'oxymoron: can one steal from oneself? *European journal of clinical investigation, 42*(3), 231-232.
Clair, J. A. (2015). Procedural Injustice in the System of Peer Review and Scientific Misconduct. *Academy of Management Learning & Education, 14*(2), 159-172. doi: 10.5465/amle.2013.0243
Council of Science Editors. (2012). CSE's White Paper on Promoting Integrity in Scientific Journal Publications. from http://www.councilscienceeditors.org/wp-content/uploads/entire_whitepaper.pdf
De Bruijn, H. (2002). Performance measurement in the public sector: strategies to cope with the risks of performance measurement. *International Journal of Public Sector Management, 15*(7), 578-594.
de Vasconcelos, S. M. R., & Roig, M. (2015). Prior Publication and Redundancy in Contemporary Science: Are Authors and Editors at the Crossroads? *Science and Engineering Ethics, 21*(5), 1367-1378. doi: 10.1007/s11948-014-9599-8
Enders, & Hoover, G. A. (2004). Whose Line Is It? Plagiarism in Economics. *Journal of Economic Literature, 42*(2), 487-493. doi: doi: 10.1257/0022051041409066
ESF/ALLEA. (2011). The European Code of Conduct for Research Integrity. Strasbourg: European Science Foundation & All-European Academies.
Fanelli, D. (2009). How Many Scientists Fabricate and Falsify Research? A Systematic Review and Meta-Analysis of Survey Data. *Plos One, 4*(5), 11. doi: 10.1371/journal.pone.0005738
Fanelli, D. (2016). *Is Dutch science at greater risk?* Paper presented at the NRIN Research conference 2016 - Fostering Responsible Research Practices, Amsterdam.
Fanelli, D., Costas, R., Fang, F. C., Casadevall, A., & Bik, E. M. (2017). Why Do Scientists Fabricate And Falsify Data? A Matched-Control Analysis Of Papers Containing Problematic Image Duplications. *bioRxiv*. doi: 10.1101/126805
Fanelli, D., Costas, R., & Lariviere, V. (2015). Misconduct Policies, Academic Culture and Career Stage, Not Gender or Pressures to Publish, Affect Scientific Integrity. *Plos One, 10*(6), 18. doi: 10.1371/journal.pone.0127556
Godecharle, S., Nemery, B., & Dierickx, K. (2013). Guidance on research integrity: no union in Europe. *Lancet, 381*(9872), 1097-1098.
Godecharle, S., Nemery, B., & Dierickx, K. (2014). Heterogeneity in European Research Integrity Guidance: Relying on Values or Norms? *Journal of Empirical Research on Human Research Ethics, 9*(3), 79-90. doi: 10.1177/1556264614540594
Harriman, S., & Patel, J. (2014). Text recycling: acceptable or misconduct? *Bmc Medicine, 12*, 148. doi: 10.1186/s12916-014-0148-8
He, T. W. (2013). Retraction of global scientific publications from 2001 to 2010. *Scientometrics, 96*(2), 555-561. doi: 10.1007/s11192-012-0906-3

Hicks, D., Wouters, P., Waltman, L., De Rijcke, S., & Rafols, I. (2015). Bibliometrics: the Leiden Manifesto for research metrics. *Nature, 520*, 429-431.

Honig, B., & Bedi, A. (2012). The fox in the hen house: A critical examination of plagiarism among members of the Academy of Management. *Academy of Management Learning & Education, 11*(1), 101-123.

Hoover, G. A. (2006). A game-theoretic model of plagiarism. *Atlantic Economic Journal, 34*(4), 449-454.

Horbach, S. P. J. M., & Halffman, W. (2016). Promoting Virtue or Punishing Fraud: Mapping Contrasts in the Language of 'Scientific Integrity'. *Science and Engineering Ethics*, 1-25. doi: 10.1007/s11948-016-9858-y

Horbach, S. P. J. M., & Halffman, W. (2019). Rejoinder to 'Is there a Peter Nijkamp effect?' by Jasper Lukkezen. *Research Policy, 48*(2), 505. doi: https://doi.org/10.1016/j.respol.2018.10.024

Horner, J., & Minifie, F. D. (2011). Research Ethics II: Mentoring, Collaboration, Peer Review, and Data Management and Ownership. *Journal of Speech Language and Hearing Research, 54*(1), S330-S345. doi: 10.1044/1092-4388(2010/09-0264)

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling. *Psychological Science, 23*(5), 524-532. doi: 10.1177/0956797611430953

Joob, B., & Wiwanitkit, V. (2016). Scientific misconduct and self-plagiarism. *American Journal of Obstetrics and Gynecology, 214*(4), 549-550. doi: http://dx.doi.org/10.1016/j.ajog.2015.12.011

Karabag, S. F., & Berggren, C. (2012). Retraction, dishonesty and plagiarism: analysis of a crucial issue for academic publishing, and the inadequate responses from leading journals in economics and management disciplines. *Journal of Applied Economics and Business Research, 2*(3), 172-183.

KNAW. (2014). CORRECT CITATION PRACTICE, Academy Advisory Memorandum. from https://www.knaw.nl/en/news/publications/correct-citation-practice?set_language=en

KNAW, NFU, NWO, TO2-federatie, Verneniging Hogescholen, & VSNU. (2018). Nederlandse gedragscode wetenschappelijke integriteit.

Kornfeld, D. S. (2012). Perspective: Research Misconduct: The Search for a Remedy. *Academic Medicine, 87*(7), 877-882. doi: 10.1097/ACM.0b013e318257ee6a

LOWI. (2015, 25-02-2015). LOWI Advies 2015, nr. 2.

Lukkezen, J. H. (2019). Is there a Peter Nijkamp effect?: A comment on The extent and causes of academic text recycling or 'self-plagiarism' by Serge Horbach and Willem Halffman. *Research Policy, 48*(2), 503-504. doi: https://doi.org/10.1016/j.respol.2018.10.023

Martin, B. R. (2013). Whither research integrity? Plagiarism, self-plagiarism and coercive citation in an age of research assessment. *Research Policy, 42*(5), 1005-1014. doi: http://dx.doi.org/10.1016/j.respol.2013.03.011

Martinson, B. C., Anderson, M. S., & de Vries, R. (2005). Scientists behaving badly. *Nature, 435*(7043), 737-738. doi: 10.1038/435737a

Marusic, A., Katavic, V., & Marusic, M. (2007). Role of editors and journals in detecting and preventing scientific misconduct: strengths, weaknesses, opportunities, and threats. *Med. & L., 26*, 545.

Mongeon, P., Waltman, L., & Rijcke, S. d. (2016). What do we know about journal citation cartels? A call for information. Retrieved from https://www.cwts.nl/blog?article=n-q2w2b4

Moskovitz, C. (2016). Self-Plagiarism, Text Recycling, and Science Education. *Bioscience, 66*(1), 5-6. doi: 10.1093/biosci/biv160

Necker, S. (2014). Scientific misbehavior in economics. *Research Policy, 43*(10), 1747-1759. doi: 10.1016/j.respol.2014.05.002

Nijkamp, P. (2015). 'Zelfplagiaat'-rapport van de Commissie-Zwemmer: Broddelwerk over fictieve peanuts. from http://www.advalvas.vu.nl/sites/default/files/dossiers/Peter%20Nijkamp%20uitgebreide%20verantwoording%20rapport%20Zwemmer%2017%20maart%202015.pdf

O'Hair, R. A. J., & Neff, J. (2013). Is There "Self Plagiarism" and "Salami" Publishing in JASMS? *Journal of the American Society for Mass Spectrometry, 24*(7), 986-987. doi: 10.1007/s13361-013-0674-1

OECD. (2010). Best Practices for Ensuring Scientific Integrity and Preventing Misconduct. from http://www.oecd.org/sti/sci-tech/40188303.pdf

Oransky, I. (2012). JACS makes it official, retracting Breslow "space dinosaurs" paper for "similarity to his previously published reviews". *Retraction Watch,*. Retrieved from http://retractionwatch.com/2012/05/20/jacs-makes-it-official-retracting-breslow-space-dinosaurs-paper-for-similarity-to-his-previously-published-reviews/

Phlippen, S. (2013). Economentop 40. In Economische en Statistische Berichten,,(Vol. 4674). Amsterdam: ESB.

Retraction Watch. (2014, 02-07-2014). Retractions arrive in plagiarism scandal involving economist Nijkamp. from http://retractionwatch.com/2014/07/02/retractions-arrive-in-plagiarism-scandal-involving-economist-nijkamp/

Roig, M. (2010). Plagiarism and self-plagiarism: What every author should know. *Biochemia Medica, 20*(3), 295-300.

Samuelson, P. (1994). SELF-PLAGIARISM OR FAIR-USE. *Communications of the Acm, 37*(8), 21-25.

Scanlon, P. M. (2007). Song From Myself: An Anatomy of Self-Plagiarism. *MPublishing, University of Michigan Library, Ann Arbor, 2*.

Schuyt, K. (2014). *Tussen fout en fraude: Integriteit en oneerlijk gedrag in wetenschappelijk onderzoek*. Leiden: Leiden University Press.

Steen, R. G. (2011). Retractions in the medical literature: how many patients are put at risk by flawed research? *Journal of Medical Ethics, 37*(11), 688-692. doi: 10.1136/jme.2011.043133

Steneck, N. H. (2006). Fostering integrity in research: Definitions, current knowledge, and future directions. *Science and Engineering Ethics, 12*(1), 53-74. doi: 10.1007/pl00022268

Stroebe, W., Postmes, T., & Spears, R. (2012). Scientific Misconduct and the Myth of Self-Correction in Science. *Perspectives on Psychological Science, 7*(6), 670-688. doi: 10.1177/1745691612460687

Part I – *To Spill*

Sun, Y. C. (2013). Do journal authors plagiarize? Using plagiarism detection software to uncover matching text across disciplines. *Journal of English for Academic Purposes, 12*(4), 264-272. doi: 10.1016/j.jeap.2013.07.002

Thurman, R. H., Chervenak, F. A., McCullough, L. B., Halwani, S., & Farine, D. (2016). Self-plagiarism: a misnomer. *American Journal of Obstetrics and Gynecology, 214*(1), 91-93. doi: http://dx.doi.org/10.1016/j.ajog.2015.09.004

Tramer, M. R., Reynolds, D. J. M., Moore, R. A., & McQuay, H. J. (1997). Impact of covert duplicate publication on meta-analysis: a case study. *British Medical Journal, 315*(7109), 635-640.

Turnitin. (2006). from http://www.turnitin.com/en_us/

Van Dalen, H. P., & Henkens, K. (2012). Intended and unintended consequences of a publish-or-perish culture: A worldwide survey. *Journal of the American Society for Information Science and Technology, 63*(7), 1282-1293.

VSNU. (2013). Beschuldiging van plagiaat. from http://www.vsnu.nl/wi-2013.html

VSNU. (2014). The Netherlands Code of Conduct for Academic Practice, revised version.

Wager, E., Fiack, S., Graf, C., Robinson, A., & Rowlands, I. (2009). Science journal editors' views on publication ethics: results of an international survey. *Journal of Medical Ethics, 35*(6), 348-353.

Westlund, H., Martin, J. C., & Stough, R. R. (2014). Letter of support for Professor Peter Nijkamp.

Zhang, M. H., & Grieneisen, M. L. (2013). The impact of misconduct on the published medical and non-medical literature, and the news media. *Scientometrics, 96*(2), 573-587. doi: 10.1007/s11192-012-0920-5

Zwemmer, J., Gunning, J. W., & Grobbee, R. (2015, 11-02-2015). Rapport over verwijzingen in het werk van prof. dr. P. Nijkamp.

Chapter 3 - The ghosts of HeLa: How cell line misidentification contaminates the scientific literature

This chapter was published as:
Horbach, S. P. J. M., & Halffman, W. (2017). The ghosts of HeLa: How cell line misidentification contaminates the scientific literature. Plos One, 12(10), e0186281. doi: 10.1371/journal.pone.0186281

3

3.1. Introduction

The misidentification of cell lines is a stubborn problem in the biomedical sciences, contributing to the growing concerns about errors, false conclusions and irreproducible experiments (Ioannidis, 2005; Schulz, Cookson, & Hausmann, 2016). As a result of mislabelled samples, cross-contaminations, or inadequate protocols, some research papers report results for lung cancer cells that turn out to be liver carcinoma, or human cell lines that turn out to be rat (Masters, 2010; Stacey, 2000). In some cases, these errors may only marginally affect results; in others they render results meaningless (Masters, 2010).

The problems with cell line misidentification (ICLAC, 2016) have been known for decades, commencing with the controversies around HeLa cells in the 1960s (Gartler, 1967, 1968; Gold, 1986; Nelson-Rees & Flandermeyer, 1976; Nelson-Rees, Flandermeyer, & Hawthorne, 1974). In spite of several alarm calls and initiatives to remedy the problem, misidentification continues to haunt biomedical research, with new announcements of large-scale cross-contaminations and widespread use of misidentified cell lines appearing even recently (A. Capes-Davis et al., 2010; Grens, 2016; Ye, Chen, Qin, Liu, & Zheng, 2015). Although no exact numbers are known, the extent of cell line misidentification is estimated between one fifth and one third of all cell lines ("Identity crisis," 2009; Masters, 2010). (Although currently only 488 or 0.6% of over 80,000 known cell lines have been reported as misidentified, most cell lines are used infrequently (Artimo et al., 2012).) In addition, misidentified cell lines keep being used under their false identities long after they have been unmasked (Lacroix, 2008), while other researchers continue to build on their results. Considering the biomedical nature of research conducted on these cell lines, consequences of false findings are potentially severe and costly (Freedman, Cockburn, & Simcoe, 2015), with grants, patents and even drug trials based on misidentified cells (Boonstra et al., 2010). Several case studies performed by the International Cell Line Authentication Committee (ICLAC) highlight some of the potential consequences of using misidentified cell lines (Hall, 2016; Vaughan, Glanzel, Korch, & Capes-Davis, 2017). Especially in the last decade, the gravity of the problem has been widely acknowledged, with several calls for immediate action in journal articles (Allison, Brown, George, & Kaiser, 2016; A. Capes-Davis et al., 2010; de Oliveira, Marques, & Losi, 2016; Masters & Iclac, 2012; Stacey, 2000), requirements for grant applications (e.g. National Institutes of Health, 2015; National Institutes of Health, 2016) and even an open letter to the US secretary of health (Nardone, 2007).

The current calls for action and remediation activities are almost exclusively concerned with avoiding future contaminations, such as through systems for easier verification of cell line identities. Various solutions have been proposed (American Type Culture Collection Standards Development Organization Workgroup ASN-0002, 2010; Lorsch, Collins, & Lippincott-Schwartz, 2014; Yu et al., 2015), among others employing genotypic identification through short tandem repeats (STR) (Grens, 2015). In addition, authors are expected to check overviews of misidentified cells (such as (Artimo et al., 2012; A. Capes-Davis et al., 2010; Liang-Chu et al., 2015; Yu et al., 2015)) before conducting their experiments. However, little attention is currently paid to the damage that has already been done through the past distribution of research articles based on misidentified cells. Although systems such as retractions and corrections are available to alert other researchers of potential problems in publications, these systems are rarely used to flag problems with cell lines (Casadevall, Steen, & Fang, 2014; Vaughan et al., 2017). Even if future misidentifications could be avoided completely – which is not likely given the track record of earlier attempts – these ‘contaminated’ articles will therefore continue to affect research.

Before any action can be taken, it is essential that we get a sense of the size and nature of the problem of contaminated literature. This raises several questions. First, how many research articles have been based on misidentified or contaminated cell lines? How wide is their influence on the scientific literature? Second, what can we say about origins and trends in the contaminated literature? Is the problem getting better, or restricted to peripheral regions of the world’s research, where perhaps protocols are less strict? Third, what could be appropriate ways to deal with the contaminated literature? To answer these questions, we searched the literature for research papers using cell lines that are known to have been misidentified. In order to put the results of this search in perspective, we analysed the precise complications of misidentification for three particular cell lines.

3.2. The process of distributing cell lines

To study the scale of literature contamination, we need to understand the process of setting up, distributing, and publishing about cell lines. This process is illustrated in Figure 3.1.

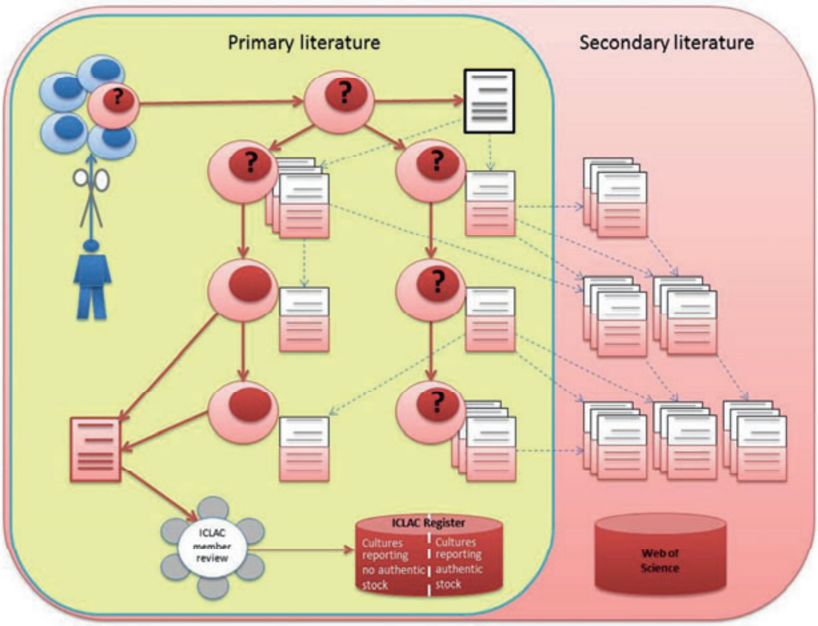


Figure 3.1 The creation, distribution and literature of a cell line: a cultured sample of cells (blue cells) may produce an immortal cell line (red cells), sometimes announced in 'an establishing paper' (in white). Cells may then be distributed to other researchers and reported in research papers, the 'primary literature'. If misidentification of cells is reported in 'a notifying paper' (in red, bottom left), this may raise questions about the entire cell line (question marks) and the papers based on it, since misidentification commonly occurs at the source. Notifying papers should be reported to ICLAC, which will decide whether cell lines should be added to the ICLAC misidentified cell line register. Meanwhile, the contaminated primary literature is cited (dotted lines) by 'secondary literature', spreading the contamination further.

Briefly, setting up a novel cell line commences with a tissue sample from an organism, human or other. If this culture grows successfully, the establishment of a new cell line is sometimes reported in what we will call an 'establishing paper'. Subsequently, scientists may share or obtain this cell line, either via their personal network or via cell banks. These scientists may perform research on this cell line and publish their findings in scientific journals, thereby setting up what we call the *primary literature* based on a cell line.

At some point, a cell line may be found to be misidentified. This observation may be published in a 'notifying paper', leading to a registration in the International Cell Line Authentication Committee's (ICLAC) database of cross-contaminated or misidentified cell lines (A. Capes-Davis et al., 2010). Based on available data, cell lines can be added to one of two tables. The first is reserved for cell lines with no known authentic stock, and the second for cell lines where authentic stock is known to exist.

In this chapter, we focus on the first category only: cell lines without any reported original stock. In this case, it must be assumed that all primary literature could be based on false grounds and should at least be treated with caution. In addition, we tried to estimate the size of the *secondary literature*: those articles referring to the primary literature, hence potentially building on questionable materials.

3.3. Materials and Methods

3.3.1. Gathering data

Version 8.0 of the ICLAC list of misidentified cell lines (A. Capes-Davis et al., 2010) was published in December 2016 (http://iclac.org/databases/cross-contaminations/). From this list we only used table 1, listing those cell lines for which no authentic stock of the originally presumed cell line is reported. This list holds 451 cell lines. The identification of research articles using any of these misidentified cell lines was more difficult than expected. Although any article based on research using one of the listed cell lines supposedly mentions this cell line, the information is frequently not incorporated in one of the searchable fields in large databases (such as title, abstract or keywords), not even in otherwise well-documented biomedical databases such as PubMed. Therefore, the exact number of papers based on misidentified cell lines cannot be established. However, we could identify articles that either refer to the establishing article of a misidentified cell line, or that name the cell line in their title, abstract or key words. This search was carried out within the Web of Science database, because this platform allows for detailed citation analysis. We used the following two search methods to obtain conservative estimates of the number of research articles based on misidentified cell lines:

<u>Method 1</u>: For any cell line on the ICLAC list we tried to find an original article that reports the establishment of the cell line. These ‘establishing articles’ were searched using the Cellosaurus database (Artimo et al., 2012), and subsequent references herein to the German collection of microorganisms and cell lines (DSMZ) database, the American Type Culture Collection (ATCC) database, and the European Collection of Authenticated Cell Cultures (ECACC) database. The DSMZ, ATCC and ECACC databases were consulted to check for references to any establishing article on one of the cell lines. Establishing articles were found for 255 cell lines. Subsequently, the original articles found in this fashion were searched in the Web of Science database. All references to the establishing articles were collected. We considered a reference to this original article as a good proxy for the usage of a cell line, since typically the original papers are focused on reporting the establishment of the cell line only (as checked in our three case studies described below).

It could be argued that we went too far back in time. It is a common phenomenon that articles have a limited ‘lifespan’, the time during which they receive citations, which would mean that older establishing articles would currently be poorly cited as the cell lines they announced are no longer relevant. To test for this we considered the citation lifespan of all establishing articles published before 1980. They turned out to have an average citation lifespan of over 40 years and the majority of them still received citations in 2016 or 2017. Hence we believe there are good reasons to be inclusive of even relatively old establishing papers.

Method 2: We searched the WoS database for all articles stating the names of one of the 451 listed cell lines in their title, abstract or keywords, as well as one of the words: 'cell(s)' or 'cell line(s)'. In order to provide for more accurate search results, we only searched articles in the 25 research fields (as defined by WoS) that were most common among the results of the first search method. Thereby we excluded articles that use misidentified cell lines, but were classified in fields in which research on cell lines is less common. We note that even though the research areas as defined by WoS may overlap, classifying articles in more than a single research area, this has not lead to double counting in our analysis: all articles included in one of the 25 identified research areas were only counted once.

3.3.2. Verification of data

Several strategies were employed to verify the validity of the data and to reduce the number of 'false positives' (i.e. those articles ending up in our sample, without reporting on research with misidentified cell lines). Regarding search method 1, we closely verified all establishing articles that resulted in at least 100 hits of primary articles in our database (n = 41). In this verification we found one article that was actually a notifying paper instead of an establishing paper and hence deleted it from our search. In addition, four articles were found that report on the establishment of several cell lines, some of which are not listed in the ICLAC database. In two of these cases, the establishing article reports on both the establishment of the contaminated cell line as well as the contaminating cell line (in the cases of EJ138 and HPB-MLT). We decided to not delete those establishing articles from our database, because they yield only few false positives (order of magnitude of tens) in our database.

Regarding search method 2: Due to some names of cell lines that could easily be confused with other meanings (such as 'WISH', 'CaVe' or 'EU-1'), this search created noise. Therefore we used an iterative process to delete this noise. This process proceeded as follows: we observed a random sample of 100 articles, sampled by ordering the articles on publication data and selecting every 200th hit. We checked whether the articles found actually used one of the listed cell lines, this was done by first reading the abstract of the found articles and, in case of doubt, further consulting the full text. If the research reported in the article did not use one of the listed cell lines, the search term that found this article was either replaced by: "*name of cell line* cell*" (e.g. "WISH cell*" instead of "WISH"), which was done in 26 cases, or the search term was deleted from the query (in cases of very generic words such as in the case of the 'OF' cell line), which was done in 15 cases (see Supporting Information table S1 for a list). (In this search, the asterisk (*) signifies a wildcard, i.e. the term 'cell*' will find any word starting with 'cell'.) We continued this process until the samples did not contain false positives due to structural issues in the search terms.

Subsequently, the process of randomly selecting 100 articles was iterated four times and was executed independently by both authors. Concluding from the results in the random samples, our search method provides reliable results. Nevertheless, the results inevitably contain remaining false positives, the extent of which is estimated to a maximum of 10% of the contaminated primary literature, judged by our verification through random samples of the set, in which 6.5% of the articles was found to consist of false positives The set of false positives contains, among others, articles using cells in the ICLAC register that are nonetheless reported with their correct origin (as reported for KB cells by Vaughan et al. (2017)).

All claims in this chapter are based on a dataset of articles found through either method 1 or 2. Both searches were performed without additional software tools, but with manual searches working with complex Boolean search strings. To gather information on the secondary literature, i.e. those articles citing the articles in the primary literature, we used the standard WoS 'citation report'. In the secondary literature we excluded self-citations in order to observe the actual 'spreading' of the contaminated literature. The exclusion of self-citations is a standard option in the WoS' citation report.

3.3.3. Case studies

In order to verify the collected data and to get a deeper understanding of how knowledge based on misidentified cell lines spreads through the literature, we performed three case studies in which we tracked the publications concerning a single cell line or a family of cell lines. All three are misidentified cell lines for which no original stock was reported and were selected at random from the ICLAC database. The case studies were performed on the cell lines: ALVA-31, a family of thymic cell lines (F2-4E5, F2-5B6, P1-1A3 and P1-4D6), and JCA-1. The results of the case studies indicate that our search method indeed renders accurate data, with only very few 'false positives', and rather conservative estimates.

3.3.4. Analyses of the contaminated literature's origins

We performed several analyses on the contaminated primary literature's origins based on WoS data, analysing their temporal and geographical origin and the distribution over research areas. The development over time uses the WoS publication date of the definitive version of the article; hence electronic versions may have been published prior to this date. The WoS goes back to 1945, but is incomplete for the first decades of the database. For the geographical origin of the research records, we employed the WoS category 'Country/Territory', which is based on the affiliation of the authors. The origins of the contaminated primary literature are compared with the total literature on research involving cell lines. This total literature comprises the articles that mention any word starting with 'cell' (i.e. cell, cells, cellular, etc.) in their title, keywords or abstract (hence not only misidentified cells), and belong to the 25 WoS-defined research areas that were most common among the dataset of contaminated primary literature. This reference group was also used to estimate what fraction of the relevant total literature is contaminated (see under the heading Contamination of the scientific literature).

3.4. Results

3.4.1. Contamination of the scientific literature

Using ICLAC's *Database of Cross-Contaminated or Misidentified Cell lines* (A. Capes-Davis et al., 2010), we searched the scientific literature with the *Web of Science* (WoS) (Thomson Reuters, 2016) to identify research articles based on misidentified cell lines. Using complementary search strategies (see methods), we were able to identify 32,755 articles (on August 4th, 2017) based on cell lines that are currently known to be different from the cell lines reported in these publications. As we only searched for cell lines known to be misidentified, this constitutes a conservative estimate of the scale of contamination in the primary literature. Moreover, to avoid false positives, we excluded several cell lines, such as the ones with non-unique identifiers or the cell lines for which verified stock is still in circulation. With non-unique identifiers we refer to names of cell lines that do not only refer to the cell line but (potentially) also to other phenomena. For example the case of the 'OF' cell line or the 'WISH' cell line. With 'non-unique identifier' we hence do not refer to cell lines that have multiple names or names with multiple spellings (such as the Intestine 407 cell line, which is also called 'Intestine407', 'Int-407' and 'Int407'). In cases of multiple spellings of cell line names, we stuck to the spelling indicated in the ICLAC database. Thereby we probably missed many articles using these cell lines in search method 2, again leading to conservative estimates.

In addition, research based on misidentified cell lines has a wide impact on the scientific literature, as it appears that these research papers are comparatively highly cited. WoS does not allow for precise total numbers, but we can give indications of this 'secondary contamination' of the literature. Analysing citations to primary contaminated articles, we found 46 papers with more than a thousand citations and over 2600 contaminated articles with over a hundred citations. Furthermore, over 92% of the contaminated papers are cited at least once, which is more than average for biomedical literature (Larivière, Gingras, & Archambault, 2009). In total, we can conservatively estimate the citations to the primary contaminated primary literature at over 500,000, excluding self-citations, thereby leaving traces in a substantial share of the biomedical literature. Even though it is clear that articles may receive citations for many reasons, including negative or even ritual citations, and hence not all citing articles contain (critical) errors, the amount of research potentially building on false grounds remains worrisome.

A spreadsheet with all results can be found in the Supplementary Material. This table lists all cell lines in the ICLAC database and the number of articles in the primary and secondary literature reporting on these cell lines, both for search method 1 and 2. In addition, the mean citation rate for articles in the primary literature is given as well as information on the temporal distribution of the secondary literature (the first and last year in which articles are published as well as the year in which most of the secondary literature on this cell line appeared). Given the fact that citation distributions tend to form (truncated) bell-shaped curves, this information provides reasonable insight in the temporal distribution of the secondary literature. The data is listed per cell line and not summarised, as this approach could lead to double counts.

The total number of research articles on cells can be estimated between 4.5 and 5 million (see methods). Therefore, the contaminated primary literature makes up a little under 0.8% of the total literature on cells, while the (potentially) contaminated secondary literature can be estimated in the order of 10% of the total research output in this area. However, we should stress that our aim is to measure the size of the problem. The sample undoubtedly contains false positives and is hence not suitable to identify individual contaminations.

3.4.2. Closer inspection of primary literature

An objection to our findings might be that our general search methods do not provide a proper overview of how specific misidentified cell lines actually affect research. To get a deeper understanding of how knowledge based on misidentified cell lines spreads through the literature, we present three case studies in which we tracked the publications concerning a single cell line or a family of cell lines. All three are misidentified cell lines for which no original stock was reported and were selected at random from the ICLAC database.

<u>ALVA-31:</u> This cell line was originally established in 1993 as a human prostate carcinoma (Loop, Rozanski, & Ostenson, 1993), but was found to be identical to a different human prostate carcinoma, the PC-3 cell line, in 2001 (van Bokhoven, Varella-Garcia, Korch, Hessels, & Miller, 2001; Varella-Garcia, Boomer, & Miller, 2001). We found 56 articles referring to ALVA-31, which are in turn cited by 2615 articles. Of these 56 primary articles, 22 were published after the misidentification of the ALVA-31 cell line was discovered. On closer inspection of those 22 articles, it appears that the ALVA-31 cell line was actually used in 20 of them, while only two articles mention the cell line’s misidentification. Remarkably, the most recent articles describing research based on ALVA-31 cells are published in 2016, fifteen years after the misidentification was reported.

In this case, one could argue that it might do little harm to use ALVA-31 cells, while actually working with PC-3 cells, because both are human prostate carcinoma and share many characteristics. However, in some cases, even researchers themselves argue that the precise identity of ALVA-31 is essential: "To exclude a cell type-specific effect, we extended ALVA-31 studies to other human PCa cell types" (Puto, Brognard, & Hunter, 2015). Subsequently, the authors explain how they used PC-3 cells in additional studies to 'exclude cell type-specific effects'; in effect comparing two identical cell lines.

Thymic cell lines: In a 1994 report, the establishment of a group of novel thymic cell lines (F2-4E5, F2-5B6, P1-1A3 and P1-4D6) (Fernandez et al., 1994) was announced. In a report by (MacLeod et al., 1999), the cell lines were found to be misidentified, having been derived in fact from a liver carcinoma. In total, 69 articles were found that refer to these cell lines, in turn cited by 2092 articles. Of the primary articles, 43 were published after the report by MacLeod et al. and the most recent one was published only in late 2016 (Lins, Vieira, Rosa, & Smaniotto, 2016). Of the fifteen most recent articles referring to the 1994 report, thirteen actually refer to it because they use the cell lines, all thirteen reporting research on thymic cells, without mentioning any knowledge of the misidentification of these cell lines. The other two articles refer to the establishing article for the sake of the method used in it to establish novel cell lines.

JCA-1: The JCA-1 cell line was originally established in 1990 (Muraki et al., 1990) and found to be misidentified in 2001 by van Bokhoven, Varella-Garcia, Korch, and Miller (2001), who showed that the cells in fact are derived from a bladder carcinoma rather than a prostate carcinoma. We found 64 articles referring to the establishing paper or explicitly mentioning JCA-1 in their title, key words or abstract. In turn, these articles are cited by 3352 articles. Of the primary articles, 18 appeared after the report by van Bokhoven et al. In contrast to the cell lines discussed previously, there seems to be no contemporary usage of JCA-1 in scientific research: the most recent article describing research using this cell line dates from 2009. However, also in this case, several articles were published reporting to use 'prostate cancer cell lines', after it became known that JCA-1 actually originated from bladder carcinoma. In fact, as we verified in the full text, of the 18 articles published after the report by van Bokhoven, Varella-Garcia, Korch, and Miller (2001), only 3 show awareness of the fact that the line had been misidentified. In contrast, 14 simply stated to have used the JCA-1 cell line, the vast majority explicitly referring to them as prostate cancer cells.

As these case studies show, merely listing a cell line as misidentified does not deter scientists from using it. This constitutes additional evidence for the claim that avoiding future contaminations does not form a complete solution to the issue of cell line misidentification. Instead, demonstrably misidentified cell lines continue to have an impact on research, either directly because scientists keep using them, or indirectly because scientists build on previous research employing misidentified lines. (Additional case studies on the consequences of using misidentified cell lines can be found at the ICLAC webpage (ICLAC, 2016).)

3.4.3. A transitory problem?

One might wonder whether the contamination of the research literature is mainly a problem of the past, given that the first concerns about misidentified cell lines were expressed half a century ago (Gartler, 1967, 1968) and that numerous initiatives have tried to alleviate the problem since.

Based on the set of 32,755 records of primary contaminated literature, we analysed the publication dates of the articles. The majority of the articles, 57%, were written since 2000 and the number of articles using misidentified cell lines is still growing (see Figure 3.2). Clearly, the problem is definitely not one of the past, but is very relevant to contemporary science, with 58 new articles based on contaminated literature appearing even as recently as February 2017.

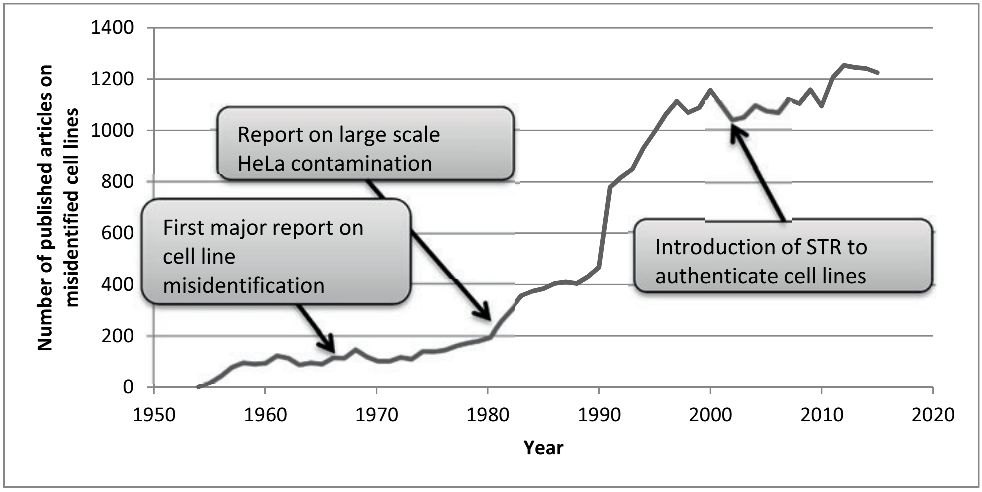


Figure 3.2 The distribution of the contaminated primary literature over the years. The graph includes references to the first report on intraspecies cell line misidentification (Gartler, 1968), a major list of misidentified cell lines based on HeLa contaminations (Nelson-Rees, Daniels, & Flandermeyer, 1981) and the introduction of Short Tandem Repeat (STR) as technique for cell line authentication (Masters et al., 2001).

Figure 3.2 indicates three moments in history when cell line contamination became evident. First, through the work of Stanley Gartler it became possible to detect intraspecies cell contamination, after which several of such contaminations involving HeLa cells were reported in *Nature* in 1968 (Gartler, 1967, 1968). Second, cell culture contamination was put on the global research agenda by the work of Walter Nelson-Rees et al. in the 1970s (Nelson-Rees & Flandermeyer, 1976; Nelson-Rees et al., 1974), culminating in a list of contaminated cell cultures in *Science* in 1981 that demonstrated large-scale contamination of cell cultures by HeLa cells (Nelson-Rees et al., 1981). From this point on, it could be expected that most scientists working in those areas of research frequently employing cell cultures, were aware of the potential issues with their research material. However, the vast majority of research papers based on misidentified cell lines was published after this point in time. Even after the introduction of STR in 2001 (Masters et al., 2001), the annual number does not decrease.

Similar to the primary literature, the number of articles in the secondary literature is also still growing. In 2016, over 40,000 papers were published that referred to primary contaminated literature. In addition, from the information in the Supplementary Material, we conclude that the majority of misidentified cell lines continue to contaminate the secondary literature in 2017 (251 cell lines for search method 1 and 232 cell lines for search method 2), while dozens of cell lines created most of their secondary literature in the past two years (38 for search method 1 and 87 for search method 2). Moreover, we conclude that many cell lines (108 for search method 1, 87 for search method 2) have generated contamination in secondary literature for a period of more than 25 years, with articles appearing long after it became known that the cell line was misidentified. Hence the contamination of the literature through reference to articles using misidentified cell lines remains a very topical problem.

3.4.4. A peripheral problem?

Another objection to our findings could be that cross-contamination occurs particularly in regions with new or emerging research communities, in which levels of training or access to testing facilities may be limited. For example, several recent publications indicate levels of cell line contamination for China between 25% (Ye et al., 2015) and 46% (Huang, Liu, Zheng, & Shen, 2017) and demonstrate that of all 'new' cell lines developed in China 85% actually turned out to be HeLa cells (Ye et al., 2015).

However, the majority of the articles using misidentified cell lines originate from countries holding well-established research traditions (e.g. US, Japan, Germany). Relative to their share of total research output, authors from these countries often perform research on misidentified cell lines. In fact, mainly due to their enormous share of total literature on cell lines, over 36% of all contaminated primary literature stems from the US. Figure 3.3 shows the percentage of contaminated primary articles as a fraction of the total number of articles on cells per country (see Supplementary Materials for data). It includes the 25 countries with the largest share of the contaminated primary literature. In this list, we see countries holding excellent research reputations ranking high. Hence, the problem does not only occur in regions with low standards of quality and diligence in research, but is also a problem in countries that hold excellent research reputations. Nevertheless, an analysis of the literature for the past five years showed a dramatic rise of China's share in the contaminated literature, confirming recent worries expressed in the literature (Ye et al., 2015).

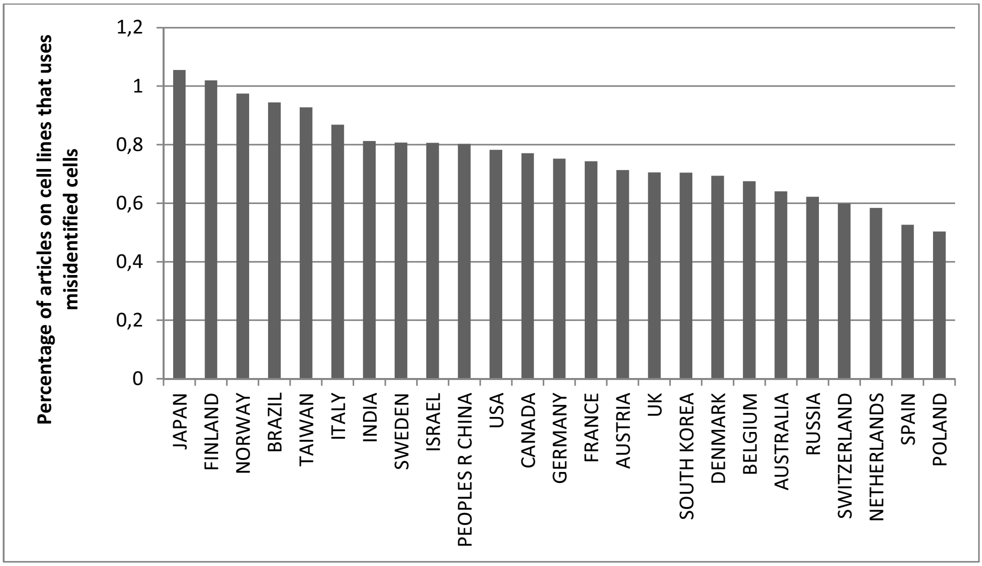


Figure 3.3 The percentage of contaminated primary articles as a fraction of the total number of articles on cells per country. The figure includes the 25 countries with the largest absolute number of articles in the contaminated primary literature.

Last, we analysed which research disciplines were most affected by the use of misidentified cell lines. Figure 3.4 shows the distribution of contaminated articles over the various research areas as defined by WoS. Among the contaminated primary literature, oncology, biochemistry/molecular biology, pharmacology and cell biology are most affected, confirming concerns about medical applications.

However, analysis of citations obtained by the primary literature indicates that the secondary literature spreads to a much more diverse range of research areas. The articles in the secondary literature originate also in fields rarely using cell lines for their research, such as psychiatry, engineering and agriculture science, see Figure 3.4. Consequently, the impact of misidentified cell cultures may spread to non-biomedical fields and affect scientists that are not as trained to judge the validity of research on misidentified cell lines.

Part I – *To Spill*

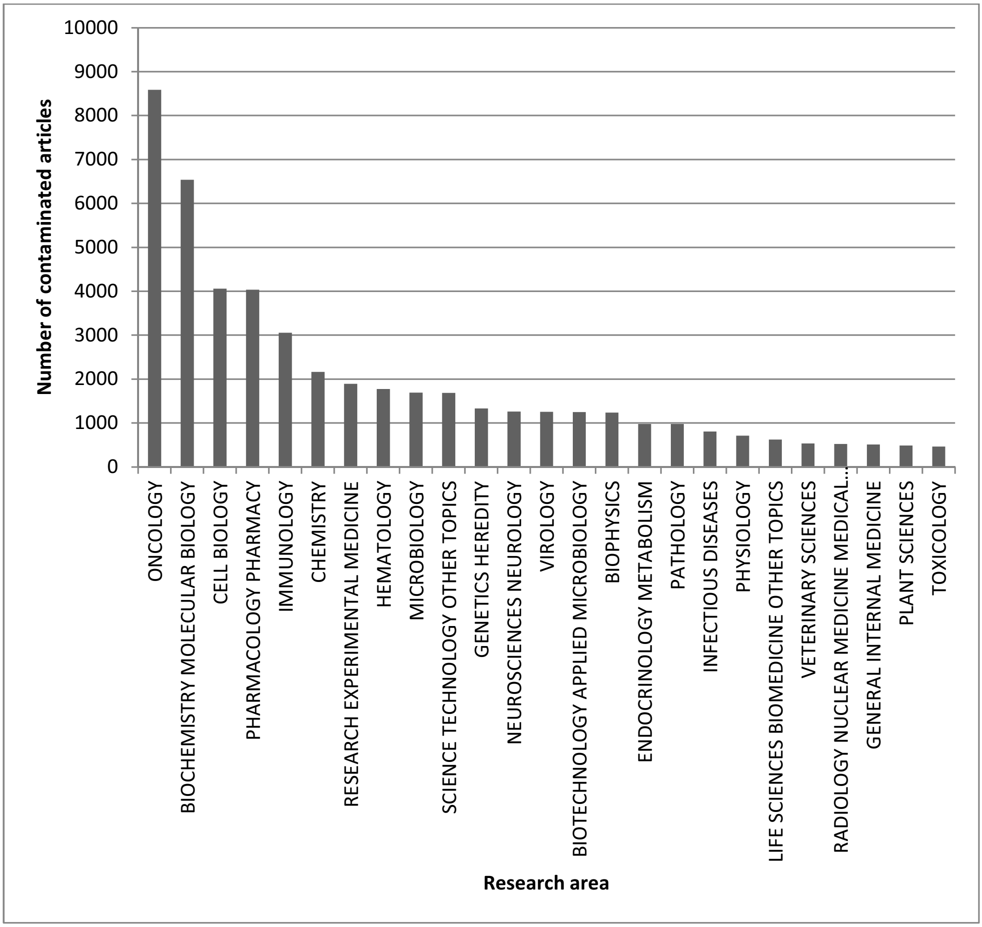


Figure 3.4 The distribution of contaminated primary literature over the research areas as defined by Web of Science. Only the 25 most affected research areas are included.

3.5. Discussion

3.5.1. Potential counter arguments

Our results seem to present worrying problems for the biomedical sciences. Although the issue of misidentified cell lines has long been known, its effect on the scientific literature has not been properly recognised, let alone properly treated (Drexler, Dirks, Matsuo, & MacLeod, 2003; Perkel, 2011). However, various arguments have been presented to suggest that papers based on misidentified cell lines are still valuable and that no remedial action is needed.

First, it has been asserted that, in some cases, the origin or specific characteristics of a cell line might be of little influence on the results of an experiment. Indeed, in some cases all that a researcher needs is ‘a cell line’, independent of type, origin or status. In fact, this argument was already mentioned by Gartler in 1968 to put his findings into perspective (Gartler, 1968). To be sure, we acknowledge that not all 32,755 articles that we found contain critical errors. However, this is not a valid argument *not* to label articles that employ misidentified cell lines, for two reasons. To begin with, it is currently up to every individual scientist to judge the status of an article every time they cite or read it, by first checking the ICLAC database of misidentified cell lines to see whether any of the cell lines used in an article are in this database, and subsequently judging the influence that the misidentification may have on the results. This is a cumbersome and unlikely assumption about how researchers cite their literature, given the low levels of awareness indicated in our analysis. In addition, the secondary literature spreads into research fields that do not commonly use cell lines for their research. It is particularly doubtful whether scientists from these fields are aware of the potential issues with research on cell lines and whether they are in the position to make informed decisions about the validity of the claims in research articles based on misidentified cell lines.

Second, it has been argued that no remedial action is needed, as the problem will be addressed by new verification techniques. Similarly, it has been argued that the problem is already widely known, that scientists may be expected to implement effective lab protocols and be sufficiently critical about cell lines and their literature. Hence contaminated literature should have faded beyond the time horizon of literature considered relevant for current research and have disappeared from the relevant research record (Masters, 2002). However, there is no sign of any ‘fading away’ of the problem. As we demonstrated, both the number of articles using misidentified cell lines, and the number of articles referring to them are still growing. Moreover, as demonstrated in the case studies, scientists show little awareness of the fact that cell lines may be misidentified. The citation analysis of the primary literature shows that articles keep being cited long after misidentifications have been reported, with over 40,000 articles citing contaminated research articles in 2016, including hundreds of citations to primary contaminated literature published decades ago.

3.5.2. Practical measures

Over the past decades thorough attention has been paid to the improvement of authentication testing for cell lines. Authentication of cell lines, and hence the ability to demonstrate cross contamination, became possible by the introduction of genetic markers by Gartler in 1967 (Gartler, 1967). Subsequently, many techniques for cell line authentication were introduced, starting off with inspection of banded marker chromosomes (Nelson-Rees et al., 1974), and the visualization of chromosomal pattern and architecture in general (MacLeod & Drexler, 2002), subsequently followed by the methods of Human Leukocyte Antigen (HLA) typing (O'Toole, Povey, Hepburn, & Franks, 1983), enzyme polymorphisms (O'Brien, Shannon, & Gail, 1980) and DNA polymorphisms (Gilbert et al., 1990). More recently, the techniques of 'DNA fingerprinting' (Jellreys, 1985) and the usage of locus-specific probes were introduced. Finally, this led to the now accepted standard method of short tandem repeat profiling (Masters et al., 2001). As has been pointed out recently, the techniques for proper cell line authentication are now widely available (Amanda Capes-Davis & Neve, 2016). However, implementation of these techniques is still falling short for multiple reasons, including time and financial constraints, lack of training and lack of (international) standards (Amanda Capes-Davis & Neve, 2016).

Despite measures to authenticate new and existing cell lines (Yu et al., 2015), research based on the wrong cells is still present in the literature and in fact continues to be published. Some form of precautionary labelling of contaminated articles seems unavoidable. However, this remedial action should be proportionate and not cause unnecessary damage. For some individual scientists, research departments, or scientific journals, rash measures could turn out to be painful. Indeed, some researchers have authored over a hundred articles in our set of contaminated primary literature. Even though the problem with these articles almost exclusively (Normile, 2014) falls under the heading of '*honest error*', with no intention to deceive, notifying all those articles as potentially erroneous, or worse: retracting them, would have a disproportionate impact on several scientists' careers. This would undermine, rather than support, an effective clean-up operation. However, in addition to catching cell line contamination at the source, initiatives to label contaminations 'downstream' in the published literature are direly needed. We can make several suggestions.

First, notifications should be posted alongside previously published articles using misidentified cell lines. This could be done in the form of 'expressions of concern', which are described as "*Neither retractions nor corrections, they alert readers that there may be an issue with a paper, when the full story is not yet clear.*" (Oransky, 2017) If clear and uncontended, the consequences of the misidentification for the article's conclusions could be reported, but otherwise the expression of concern could merely state: "Cell line X in this study is known to be misidentified and is actually Y. See Z for more information." The interpretation of this warning is then entirely up to the expert reader. Such notifications would also serve to preserve as much valuable data as possible: data reported on a misidentified cell line might still be entirely valid, provided the real origin of the cell line is clear. Hence it might be a waste of funds and efforts to automatically dismiss these data. In cases where the use of these cell lines leads to (severely) false conclusions that could have a major impact on future research, articles could be retracted. For recent cases, a system of self-retractions, as proposed by Fanelli (2016), could be employed.

Second, to allow for simple future identification of articles using misidentified cell lines, we recommend that authors mention the employed cell lines in easily searchable parts of their article, such as the keywords or abstract. Some journals have already suggested measures in this direction, but implementation seems to be slow (Grens, 2015). However, some journals have currently installed a system of Research Resource Identifiers (RRIDs), which might assist in tackling the cell line misidentification issues (Singh Chawla, 2015). Alternatively, a system of cross-reference between databases of cell lines and scientific journal publications could be set up. Linking the NCBI databases of 'BioSamples' and 'research articles' would be a natural candidate for such a system. In similar ways, the Cellosaurus database, the ICLAC database and Research Resource Initiative are already cross-linked.

In addition, better use could be made of paper trails for cell line provenance (Freshney, 2002). A clear and complete overview of the origin of a cell line, as well as the various verification tests, the experiments that it has been part of, and the results that these yielded, would be of great benefit in examining the status and quality of a cell line. In addition, this would allow for easy identification of potentially erroneous research when a cell line is found to be misidentified.

Besides being of use in terms of recognition of erroneous research, the paper trail might also serve other purposes, such as mapping the existing knowledge on a certain cell line (thereby also allowing for simple identification of knowledge gaps) and providing a stage for the publication of negative results of experiments on cell lines. The publication of such results has long been proposed as a way of fostering integrity in research (McElreath & Smaldino, 2015).

Nearly half a century after the first concerns about misidentified cell lines, the initiatives to improve authentication need to be complemented by attention to the already contaminated literature. Our analysis shows that the task is sizeable and urgent.

3.6. Postscript

Since the publication of this chapter as article in *PLoS*, several studies have continued to assess the incidence and distribution of articles reporting on misidentified cell lines. Babic et al. (2019) studied the effectiveness of using RRIDs to prevent reporting on misidentified cell lines. Using text-mining methods to analyse the methods section of articles in Pubmed Central they find over 300k mentions of cell lines in 150k articles, of which 8.6% is misidentified. This is a considerably higher than the share of articles we estimated to report on misidentified cells. In papers using RRIDs they find only 3.3% of cell lines to be misidentified, hence concluding that RRIDs are an effective way of reducing reporting on problematic cells. In the work by Hepkema (2019), the effectiveness of journal guidelines is assessed. In addition, she compares our search method to identify problematic articles with Babic et al.'s. Through this comparison she finds that correctly identifying articles on misidentified cells is potentially more arduous than we expected and error margins, which we estimated at 10%, might be larger than anticipated.

3.7. References

Allison, D. B., Brown, A. W., George, B. J., & Kaiser, K. A. (2016). A tragedy of errors. *Nature, 530*(7588), 27-29.

American Type Culture Collection Standards Development Organization Workgroup ASN-0002. (2010). Cell line misidentification: the beginning of the end. *Nat Rev Cancer, 10*(6), 441-448. doi: 10.1038/nrc2852

Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E., . . . Stockinger, H. (2012, Jul). ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res.* 2012/06/05. Retrieved Web Server issue, 40, from http://web.expasy.org/cellosaurus/

Babic, Z., Capes-Davis, A., Martone, M. E., Bairoch, A., Ozyurt, I. B., Gillespie, T. H., & Bandrowski, A. E. (2019). Incidences of problematic cell lines are lower in papers that use RRIDs to identify cell lines. *Elife, 8*, e41676. doi: 10.7554/eLife.41676

Boonstra, J. J., van Marion, R., Beer, D. G., Lin, L., Chaves, P., Ribeiro, C., . . . Altorki, N. K. (2010). Verification and unmasking of widely used human esophageal adenocarcinoma cell lines. *Journal of the National Cancer Institute, 102*(4), 271-274.

Capes-Davis, A., & Neve, R. M. (2016). Authentication: A Standard Problem or a Problem of Standards? *PLOS Biology, 14*(6), e1002477. doi: 10.1371/journal.pbio.1002477

Capes-Davis, A., Theodosopoulos, G., Atkin, I., Drexler, H. G., Kohara, A., MacLeod, R. A., . . . Freshney, R. I. (2010). Check your cultures! A list of cross-contaminated or misidentified cell lines. *Int J Cancer, 127*(1), 1-8. doi: 10.1002/ijc.25242

Casadevall, A., Steen, R. G., & Fang, F. C. (2014). Sources of error in the retracted scientific literature. *Faseb Journal, 28*(9), 3847-3855. doi: 10.1096/fj.14-256735

de Oliveira, D. E., Marques, C. S., & Losi, V. C. (2016). "Cell identity" crisis: Another call for immediate action. *Cancer Letters, 381*(1), 122-123. doi: 10.1016/j.canlet.2016.07.031

Drexler, H. G., Dirks, W. G., Matsuo, Y., & MacLeod, R. A. F. (2003). False leukemia-lymphoma cell lines: an update on over 500 cell lines. *Leukemia, 17*(2), 416-426. doi: 10.1038/sj.leu.2402799

Fanelli, D. (2016). Set up a 'self-retraction' system for honest errors. *Nature, 531*(7595), 415-415.

Fernandez, E., Vicente, A., Zapata, A., Brera, B., Lozano, J. J., Martinez, C., & Toribio, M. L. (1994). Establishment and Characterization of Cloned Human Thymic Epithelial - Cell Lines - Analysis of Adhesion Mulecule Expression and Cytokine Production *Blood, 83*(11), 3245-3254.

Freedman, L. P., Cockburn, I. M., & Simcoe, T. S. (2015). The Economics of Reproducibility in Preclinical Research. *PLOS Biology, 13*(6), e1002165. doi: 10.1371/journal.pbio.1002165

Freshney, R. I. (2002). Cell line provenance. *Cytotechnology, 39*(2), 55-67.

Gartler, S. M. (1967). Genetic markers as tracers in cell culture. *National Cancer Institute Monograph, 26*, 167-195.

Gartler, S. M. (1968). Apparent HeLa Cell Contamination of Human Heteroploid Cell Lines. *Nature, 217*(5130), 750-750. doi: 10.1038/217750a0

Gilbert, D. A., Reid, Y., Gail, M. H., Pee, D., White, C., Hay, R. J., & O'Brien, S. J. (1990). Application of DNA fingerprints for cell-line individualization. *American journal of human genetics, 47*(3), 499.

Gold, M. (1986). *A Conspiracy of Cells: One Woman's Immortal Legacy-And the Medical Scandal It Caused*: SUNY Press.

Grens, K. (2015). The Great Big Clean-Up. *Scientist, 29*(9), 50-55.

Grens, K. (2016, 31-08-2016). Popular Tumor Cell Line Contaminated. *The Scientist.* from http://www.the-scientist.com/?articles.view/articleNo/46929/title/Popular-Tumor-Cell-Line-Contaminated/

Hall, E. (2016). RGC-5 is not rat retinal ganglion. *Case studies.* from http://iclac.org/case-studies/rgc-5/

Hepkema, W. (2019). *The effect of journal guidelines on the use of misidentified cell lines in scientific literature.* (Master of Science), Radboud University, Nijmegen.

Huang, Y., Liu, Y., Zheng, C., & Shen, C. (2017). Investigation of Cross-Contamination and Misidentification of 278 Widely Used Tumor Cell Lines. *Plos One, 12*(1), e0170384. doi: 10.1371/journal.pone.0170384

ICLAC. (2016). Definitions. 2017, from http://iclac.org/resources/definitions/

Identity crisis. (2009). *Nature, 457*(7232), 935-936. doi: 10.1038/457935b

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Med, 2*(8), e124.

Jellreys, A. (1985). Individual-specific 'fingerprints' of human DNA. *Nature, 316*, 4.

Lacroix, M. (2008). Persistent use of "false" cell lines. *Int J Cancer, 122*(1), 1-4. doi: 10.1002/ijc.23233

Larivière, V., Gingras, Y., & Archambault, É. (2009). The decline in the concentration of citations, 1900–2007. *Journal of the Association for Information Science and Technology, 60*(4), 858-862.

Liang-Chu, M. M., Yu, M., Haverty, P. M., Koeman, J., Ziegle, J., Lee, M., . . . Neve, R. M. (2015). Human biosample authentication using the high-throughput, cost-effective SNPtrace(TM) system. *Plos One, 10*(2), e0116218. doi: 10.1371/journal.pone.0116218

Lins, M. P., Vieira, L. F. D., Rosa, A. A. M., & Smaniotto, S. (2016). Growth hormone in the presence of laminin modulates interaction of human thymic epithelial cells and thymocytes in vitro. *Biological Research, 49*. doi: 10.1186/s40659-016-0097-0

Loop, S. M., Rozanski, T. A., & Ostenson, R. C. (1993). Human Primary Prostate Tumor-Cell Line, Alva 31: A New Model for Studying the Hormonal-Regulation of Prostate Tumor-Cell Growth. *Prostate, 22*(2), 93-108. doi: 10.1002/pros.2990220202

Lorsch, J. R., Collins, F. S., & Lippincott-Schwartz, J. (2014). Fixing problems with cell lines. *Science, 346*(6216), 1452-1453. doi: 10.1126/science.1259110

MacLeod, R. A. F., Dirks, W. G., Matsuo, Y., Kaufmann, M., Milch, H., & Drexler, H. G. (1999). Widespread intraspecies cross-contamination of human tumor cell lines arising at source. *Int J Cancer, 83*(4), 555-563.

MacLeod, R. A. F., & Drexler, H. G. (2002). Authentication and Characterization. In J. R. W. Masters & B. O. Palsson (Eds.), *Human Cell Culture: Cancer Cell Lines Part 3: Leukemias and Lymphomas* (pp. 371-397). Dordrecht: Springer Netherlands.

Masters, J. R. W. (2002). HeLa cells 50 years on: the good, the bad and the ugly. *Nature Reviews Cancer, 2*(4), 315-319.

Masters, J. R. W. (2010). Cell line misidentification: the beginning of the end. *Nature Reviews Cancer, 10*(6), 441-448. doi: 10.1038/nrc2852
Masters, J. R. W., & Iclac. (2012). End the scandal of false cell lines. *Nature, 492*(7428), 186-186.
Masters, J. R. W., Thomson, J. A., Daly-Burns, B., Reid, Y. A., Dirks, W. G., Packer, P., . . . Arlett, C. F. (2001). Short tandem repeat profiling provides an international reference standard for human cell lines. *Proceedings of the National Academy of Sciences, 98*(14), 8012-8017.
McElreath, R., & Smaldino, P. E. (2015). Replication, Communication, and the Population Dynamics of Scientific Discovery. *Plos One, 10*(8), e0136088. doi: 10.1371/journal.pone.0136088
Muraki, J., Addonizio, J. C., Choudhury, M. S., Fischer, J., Eshghi, M., Davidian, M. M., . . . Chiao, J. W. (1990). Establishment of New Human Prostatic-Cancer Cell-Line (JCA-1). *Urology, 36*(1), 79-84. doi: 10.1016/0090-4295(90)80319-i
Nardone, R. M. (2007). An Open Letter Regarding the Misidentification and Cross-Contamination of Cell Lines: Significance and Recommendations for Correction. from http://www.cellbankaustralia.com/resources/Nardone-Open-Letter.pdf
National Institutes of Health. (2015). Enhancing Reproducibility through Rigor and Transparency, NOT-OD-15-103. Bethesda.
National Institutes of Health. (2016). Implementing Rigor and Transparency in NIH & AHRQ Research Grant Applications, NOT-OD-16-011. Bethesda.
Nelson-Rees, W. A., Daniels, D., & Flandermeyer, R. R. (1981). Cross-contamination of cells in culture. *Science, 212*(4493), 446-452.
Nelson-Rees, W. A., & Flandermeyer, R. R. (1976). HeLa cultures defined. *Science, 191*(4222), 96-98.
Nelson-Rees, W. A., Flandermeyer, R. R., & Hawthorne, P. K. (1974). Banded marker chromosomes as indicators of intraspecies cellular contamination. *Science, 184*(4141), 1093-1096.
Normile, D. (2014, Jul. 2, 2014). Nature retracts controversial stem cell papers. from http://www.sciencemag.org/news/2014/07/nature-retracts-controversial-stem-cell-papers
O'Brien, S. J., Shannon, J. E., & Gail, M. H. (1980). A molecular approach to the identification and individualization of human and animal cells in culture: isozyme and allozyme genetic signatures. *In Vitro Cellular & Developmental Biology-Plant, 16*(2), 119-135.
O'Toole, C. M., Povey, S., Hepburn, P., & Franks, L. (1983). Identity of some human bladder cancer cell lines.
Oransky, I. (2017). Stuck in limbo: What happens to papers flagged by journals as potentially problematic? *Retraction Watch.* Retrieved 28-02-2017, from http://retractionwatch.com/2017/02/28/stuck-limbo-happens-papers-flagged-journals-potentially-problematic/
Perkel, J. M. (2011). Curing Cell Lines. *Biotechniques, 51*(2), 85-90.
Puto, L. A., Brognard, J., & Hunter, T. (2015). Transcriptional Repressor DAXX Promotes Prostate Cancer Tumorigenicity via Suppression of Autophagy. *Journal of Biological Chemistry, 290*(25), 15406-15420. doi: 10.1074/jbc.M115.658765

Schulz, J. B., Cookson, M. R., & Hausmann, L. (2016). The impact of fraudulent and irreproducible data to the translational research crisis - solutions and implementation. *Journal of Neurochemistry, 139*, 253-270. doi: 10.1111/jnc.13844

Singh Chawla, D. (2015). Researchers argue for standard format to cite lab resources. *Nature, 29*, 2015. doi: doi:10.1038/nature.2015.17652.

Stacey, G. N. (2000). Cell contamination leads to inaccurate data: we must take action now. *Nature, 403*(6768), 356-356. doi: 10.1038/35000394

Thomson Reuters. (2016). Web of Knowledge. from https://apps.webofknowledge.com/

van Bokhoven, A., Varella-Garcia, M., Korch, C., Hessels, D., & Miller, G. J. (2001). Widely used prostate carcinoma cell lines share common origins. *Prostate, 47*(1), 36-51. doi: 10.1002/pros.1045

van Bokhoven, A., Varella-Garcia, M., Korch, C., & Miller, G. J. (2001). TSU-Pr1 and JCA-1 cells are derivatives of T24 bladder carcinoma cells and are not of prostatic origin. *Cancer Research, 61*(17), 6340-6344.

Varella-Garcia, M., Boomer, T., & Miller, G. J. (2001). Karyotypic similarity identified by multiplex-FISH relates four prostate adenocarcinoma cell lines: PC-3, PPC-1, ALVA-31, and ALVA-41. *Genes Chromosomes & Cancer, 31*(4), 303-315. doi: 10.1002/gcc.1149

Vaughan, L., Glanzel, W., Korch, C., & Capes-Davis, A. (2017). Widespread use of misidentified cell line KB (HeLa): Incorrect attribution and its impact revealed through mining the scientific literature. *Cancer Research*. doi: 10.1158/0008-5472.CAN-16-2258

Ye, F., Chen, C. G., Qin, J., Liu, J., & Zheng, C. Y. (2015). Genetic profiling reveals an alarming rate of cross-contamination among human cell lines used in China. *Faseb Journal, 29*(10), 4268-4272. doi: 10.1096/fj.14-266718

Yu, M., Selvaraj, S. K., Liang-Chu, M. M. Y., Aghajani, S., Busse, M., Yuan, J., . . . Neve, R. M. (2015). A resource for cell line authentication, annotation and quality control. *Nature, 520*(7547), 307-+. doi: 10.1038/nature14397

Part II

TO FILTER

In the second part of this dissertation, we shift our attention from how dubious research practices may intentionally or unintentionally spill through the academic literature to how journals may prevent this. In part I – *To Spill* we analysed the occurrence patterns of some problematic research practices, showing that they occur at different rates in different research context. Some research disciplines are more vulnerable to some questionable practices and specific geographical reasons seem to be more at risk. This raises questions about how such detrimental research practices may best be avoided or dealt with.

In *To Filter*, we will consider one specific way of avoiding that questionable research practices do harm to the scientific enterprise by studying journal's editorial processes as filtering mechanisms. This act of filtering and self-regulation of science has been a key issue in the debate on research misconduct. However, little research is conducted to empirically study such aspects. Multiple self-regulation mechanisms have been proposed, some of which we will revisit in part III of this dissertation. Still, the peer review system in particular is considered an essential gatekeeper of both quality and integrity in science.

However, the allocation of this responsibility to the peer review system is not self-evident. In fact, it is fairly recent and still remains controversial. The emergence of the expectation of peer review to act as a filtering mechanism will be the first point of attention in part II of this dissertation. Chapter 4 describes the establishment of peer review's current models by reviewing the scientific literature on peer review and by adding recent developments based on information from editors and publishers. It analyses the rationale for developing new review formats and discusses how they have been implemented in the current system, resulting in a systematisation of peer review procedures.

In chapter 5 we subsequently use this systematisation of review procedures to assess their effectiveness in flagging problematic research. To do so, we use a variety of data sources. A survey among journal editors provides us information about the type of review used by specific journals, and the Retraction Watch database of retracted journal articles provides a proxy to problematic research. Combining this survey data and knowledge about retracted journal articles, we aim to assess the correlation between retraction rates and specific review procedures. Just as we did in part I – *To Spill, r*esults are verified for disciplinary differences. This time we also assess variation between reasons for retraction.

The last two chapters of this part continue to analyse innovations in editorial procedures, as well as their implementation in journals' editorial process. The variety of different actors' expectations of the editorial system may have led to the development of specific new editorial procedures, with even more innovations to be expected in the future. However, little is known about whether these innovations manage to convince other journal editors. We therefore set out to study implementation patterns across publishers and research disciplines, as well as the motivations for editors or publishers to engage in novel review procedures.

We use two complementary methods to study this. First, chapter 6 uses the results of the survey introduced in chapter 5 to assess patterns of innovation and implementation across a wide spectrum of journals. It uses the information provided by journal editors to historically trace the usage of review procedures and the reasons to change them. This quantitative approach is supplemented with ethnographic data in chapter 7. This final chapter of *To Filter* reports on ethnographic fieldwork done at the editorial offices of large commercial publishers. It describes how these publishers organise their editorial process, who makes decisions about this organisation and what factors influence these decisions.

In combination, the chapters in *To Filter* analyse what different models of peer review exist, how effective they are in preventing retractions of journal articles, where different models are implemented, and the reasons for (not) implementing innovative review procedures. By combining a plurality of research methods, including literature review, interviews, survey and scientometric approaches, as well as ethnography, part II of this dissertation aims to present a multi-perspective description of editorial peer review. This may lead to informed recommendations for journal editors and academic publishers on how to organise their editorial process and ways of effectively implementing specific review procedures.

Chapter 4 - The changing forms and expectations of peer review

This chapter was published as:
Horbach, S. P. J. M., & Halffman, W. (2018).
The changing forms and expectations of peer review. Research Integrity and Peer Review, 3(1), 8. doi: 10.1186/s41073-018-0051-5

4

4.1. Background - Quality and integrity in science

Recently, there has been heated debate on the quality, credibility and integrity of scientific literature. Due to a perceived increase in scientific fraud and irreproducible research, some claim the publication system, or even science in general, to be in crisis (Begley & Ioannidis, 2015; de Oliveira, Marques, & Losi, 2016). This rising concern has become obvious in the media, in policy initiatives, as well as in scientific literature. Concerned scientists as well as policymakers increasingly express their worry about data manipulation, plagiarism, or questionable research practices that affect the functioning of science (Horbach & Halffman, 2017b).

A key issue in the debate on scientific integrity has been the extent to which processes of institutional self-regulation are able to track and prevent misconduct (e.g. Hiney, 2015; Stroebe, Postmes, & Spears, 2012). It has long been assumed that misconduct could hardly occur in the sciences due to well-established self-regulating mechanisms (Koshland, 1987). Sociologists of science in the tradition of Merton assumed that any form of research misconduct would sooner or later come to light due to scientists' motivation to challenge competing knowledge claims via the peer review system, replication studies, or the presence of a whistle-blower, at least in as far as misconduct involves the misrepresentation of the research process (Zuckerman, 1984).

The system of peer reviewing research papers in particular has long been central to these notions of self-regulation (Horner & Minifie, 2011). However, the expectation and ability of the peer review system to detect fraudulent and erroneous research is contentious and has developed and changed over time. While some currently argue that 'safeguarding the scientific integrity of published articles' is one of peer review's core responsibilities (Guston, 2007; LaFollette, 1992; Rennie, 2003; Stroebe et al., 2012), others argue that the system was never designed, nor meant to do so (Biagioli, 2002; Smith, 2006; Südhof, 2016). Some even claim that peer review "ensures the state of good science", and "assures that science is trustworthy, relevant and valuable" (Cronin, 2005; Ware, 2008; Weller, 2001), while others regard these claims as mere 'myths', and find peer review to be conservative, biased and putting a burden on (unpaid and unrecognised) reviewers (Biagioli, 2002; Smith, 2006; Stroebe et al., 2012; Südhof, 2016).

Nevertheless, most scholars seem to agree that peer review serves as a filter in distinguishing between 'good' and 'bad' science (Pontille & Torny, 2015; Tennant et al., 2017). Despite an ever-growing number of concerns about its effectiveness, fairness and reliability (Easterbrook, Gopalan, Berlin, & Matthews, 1991; Franke & Kaul, 1978; Lee, Sugimoto, Zhang, & Cronin, 2013; Lock, 1985; Smith, 2006; Südhof, 2016; Weller, 2001; Yankauer, 1990), peer review is still considered the best available practice to ensure the quality and correctness of the scientific literature. However, the devil is in the detail: specific features have been added to the peer review process in the expectation they would address specific problems obscured by blanket notions such as 'quality'. Currently there is a clear need for a systematic analysis of peer review forms and their underlying concerns, especially in light of a wave of experimentation fuelled by new internet technologies.

Ever since being established, journal peer review has developed in a quite disorderly fashion so that currently it comes in many shapes and sizes (Burnham, 1990; Tennant et al., 2017). For various reasons, different journals and publishers tend to adhere to different forms of peer review. Among others, the increased specialisation in areas of science (Biagioli, 2002; Rennie, 2003), the rapid growth of science (Bornmann, 2011; Walker & Rocha da Silva, 2015), the changing financial foundation and incentives in scientific publishing (Guédon, 2001; Guston, 2007; LaFollette, 1992; Larivière, Haustein, & Mongeon, 2015), and the advent of novel technological possibilities (BioMed Central, 2017; Gunnarsdottir, 2005; Larivière et al., 2015; Solomon, 2002) all have had a major impact on the structure of peer review. By now, so many forms of peer review exist that some claim we can no longer call it a single system (Biagioli, 2002; Pontille & Torny, 2015; Rennie, 2003). While peer review is used in many contexts, including in grant assessment and career advancement, we will focus here on peer review of journal articles only. In addition, we will focus on the aspects directly affecting the review of a paper's content in the editorial process (i.e. the intellectual exercise), rather than on the (technical) infrastructure that facilitates it (i.e. contemporary digital review submission systems or the analogue predecessors in which reviews were communicated via e-mail.)

This chapter has three objectives. First, we describe the diversity of current peer review practices and innovations in section 4.2. We review the academic literature to analyse the various rationales for developing these new forms, and discuss how they have been implemented. In doing so, we add some of the latest innovations to a new overview that improves on existing ones. Second, using our updated overview, we will identify some common patterns in the various peer review forms in a typology that systematises this diversity. This typology, presented in section 4.2.4, can serve as a useful tool for future research on peer review instruments, e.g. in considering the quality and effectiveness of review forms. Third, in section 4.3, we will pay detailed attention to the emergence of novel expectations some have of peer review, specifically for maintaining 'the integrity of science's published record'. We will also indicate how these expectations have inspired peer review innovations.

We will demonstrate that these new expectations are not always entirely compatible with one another and hence lead to tensions in the current academic debate about what peer review can and should do. Underlying this debate, we note a growing expectation that the scientific literature will serve as a database of established knowledge, rather than as a collection of research reports, pointing to more fundamental disagreement about the nature of scientific knowledge. At least some of the expectations of peer review are not just about the practicalities of 'how to make it work better'; many also expect the process to address the functions of the publication system and even what it means to publish an account of a research project.

4.2. The historic development of peer review

4.2.1. The appearance of peers

Many accounts of the peer review process' origins locate its beginnings in the 17th century, coinciding with Henry van Oldenburg's establishment of an academic journal (Biagioli, 2002; Burnham, 1990; Kronick, 1976; Rennie, 2003). However, historians of science have increasingly rejected this claim. In fact, they argue that many journals did not introduce peer review in the sense of 'peers judging the publishability of a manuscript' until after the Second World War (Baldwin, 2015, 2017; Fyfe, McDougall-Waters, & Moxham, 2015). Earlier, decisions on acceptance or rejection would commonly be made by a single editor or a small editorial committee, frequently based on their personal preferences (Baldwin, 2015). In fact, the term 'peer review' only emerged in the scientific press in the 1960s and even then was initially used to describe grant review processes, rather than journal article reviewing (Baldwin, 2017; Moxham & Fyfe, 2017).

The practice of assessing or commenting on manuscripts prior to publication primarily arose in learned societies in the early and mid-19th century (Moxham & Fyfe, 2017). In their early forms, reviews were commonly performed by other society members and hardly intended to act as a gatekeeping mechanism. Instead, comments or reports about manuscripts were aimed, for instance, at increasing the public visibility of science or evaluating new findings in service of the king (Csiszar, 2016). Only in the late 19th century, by the time some review practices were well-established (Moxham & Fyfe, 2017), was the referee gradually "reimagined as a sort of universal gatekeeper with a duty to science" (Csiszar, 2016). Despite some early concerns, the system remained in use and was slowly adopted by independent journals, also outside the scope of academic societies.

In the late 19th century, the *British Medical Journal* (BMJ) was one of the independent journals to pioneer the novel practice of using external reviewers to assess submitted manuscripts. Since 1893, its editor-in-chief, Ernest Hart, called upon the specialised knowledge of a reviewer, whom he labelled as "an expert having knowledge and being a recognised authority in the matter". Although Hart acknowledged the fact that such a system was labour intensive, requiring "heavy daily correspondence and constant vigilance to guard against personal eccentricity or prejudice", he believed that his system of selecting outside reviewers was "the only system that seems adequate to the real needs of professional readers" (Burnham, 1990).

In bringing outside expertise to the review process, extending its scope to actual peers, rather than a closed group of editorial committee members, the peer review process began to take the shape that is still very common today. However, this system of employing other peers than the journal's or publisher's committee members, only became regular practice after the Second World War (Baldwin, 2017), with a major journal such as *Nature* adopting such a peer review system as late as 1973 (Baldwin, 2015).

In addition, differences between scientific fields were substantial. From the outset, (external) reviewing practices were considered time-consuming, costly and labour intensive. Especially in fast-developing fields, peer reviews were considered so burdensome that they prohibited quick knowledge exchange, and so made journals reluctant to use review mechanisms akin to those in learned societies (Baldwin, 2013; Moxham & Fyfe, 2017). Moreover, different publishing formats, e.g. monographs as opposed to journal articles, have resulted, even today, in distinct review practices in different research fields (Moxham & Fyfe, 2017; Pontille & Torny, 2015).

Several factors have been at the heart of journals' and societies' rationales for starting to use external reviewers in their review practices. Specialisation and growth in science were two such motivating factors. As growing numbers of manuscripts covering a wider range of topics and specialisations were submitted, editors had to select which they would publish and were less and less capable of judging all submitted work themselves. This led to them soliciting external, expert opinions (Biagioli, 2002; Burnham, 1990; Manske, 1997). Other factors, including a shift in the role of science in society, could have been equally important in establishing review systems. Specifically, the practice of external referees assessing and judging submitted manuscripts was taken up most prominently in the United Kingdom and North America, while other regions remained very hesitant until well after the Second World War (Csiszar, 2016). And then, even between the UK and USA, there are differences. In the USA, review practices were perceived (among others) as mechanisms for providing scientific legitimacy that would answer to growing requirements of public accountability. These expectations were less pronounced in other regions, which partly explains the slower development of external review systems (Baldwin, 2015; Csiszar, 2016). However, the gradual spread of publications being peer-reviewed as a quality indicator supervised by research managers provided a strong incentive for researchers to publish in peer-reviewed journals.

In spite of currently being revered in some sciences, peer review still has a remarkably short history. The work of luminaries such as Einstein, for example, was often published without being peer reviewed (Kennefick, 2005). Peer review practices were varied and often contentious. In the debates on peer review, specific concerns led to innovations and modifications, to which we will now turn our attention.

4.2.2. The concern for fairness and bias

4.2.2.1. Blind justice

After the system using external reviewers became widely implemented in the 1960s and 70s, developments in peer review succeeded each other with increasing speed. The first major developments concerned the level of anonymity in review. Initial peer review practices (nearly) always disclosed authors' identities to editors and reviewers, whereas authors knew the identity of the editor-in-chief, but not necessarily of the editorial committee or invited outside reviewers (Moxham & Fyfe, 2017). Already in the 1950s, in the framework of sociology journals, the matter of blinding authors' and reviewers' identities was raised. The *American Sociological Review* was the first to install regulations in which authors were required to attach a detachable cover page to their manuscript so that their identities could be obscured. The rest of the paper had to "bear the title as a means of identification, but not name and institution" (American Sociological Review, 1955). From sociology, the anonymization of authors spread to other social sciences and the humanities.

Starting in the 1970s and continuing to the present, various researchers have examined the bias in selecting and accepting manuscripts of authors of different demographics and status (Zuckerman & Merton, 1971). In response to this debate, various categories describing different forms of author and reviewer anonymity in peer review were established in the mid-1980s (Pontille & Torny, 2014; Prechelt, Graziotin, & Fernández, 2017). These categories are still in place and frequently show up in discussion regarding peer review:

		Reviewer	
		Anonymised	**Identified**
Author	**Anonymised**	Double-blind	'Blind review'
	Identified	Single-blind	Open review

Table 4.1: Forms of peer review blinding, Reproduced with permission from (Pontille & Torny, 2014) licensed under a CC BY-NC-ND 3.0 License

The single-blind and double-blind systems have continued to be the most common forms of evaluating articles, with a tendency to use the single-blind format in the biomedical and natural sciences, and a the double-blind system more frequently in the social sciences and humanities (Pontille & Torny, 2014; Ware, 2008; Weller, 2001). In addition, a *triple-blind* review process has been proposed, in which the identity of the author is not only concealed from the reviewers, but also from the handling editors (Rojas, 2007). Currently, a few journals use this system, but it remains fairly uncommon in designing review processes (Tennant et al., 2017).

The rationale for developing the system of double-blind review was simple: in the new system only the journal's secretariat would know the author's identity, therefore peer evaluation and editorial committee decisions would rely only on the content of the manuscript and not on the reputation of the author or his/her institute (Pontille & Torny, 2014). Subsequently, when author anonymisation spread to other social sciences and humanities, a different rationale emerged. The extension was introduced not only on editorial initiative as had been the case when the *American Sociological Review* established the system in sociology, but also resulted from demands for fair and equal treatment of minority groups in science, most notably women (Benedek, 1976). As such, this development is part of a broader societal movement, including the second feminist wave, that demands equity between different members of society (Whelehan, 1995).

The call for more equal treatment of minority groups was strengthened by various assessments of bias in peer review. Although evidence of such bias remains slightly indecisive (Tennant et al., 2017), there are strong indications that it exists, especially regarding gender and status/affiliations. This was confirmed in a famous study by Peters and Ceci (1982), in which they resubmitted published manuscripts with different authors' and institutions' names and paraphrased titles to the very same journals that had published them. The vast majority of the manuscripts (8 out of 12) was rejected on grounds of poor quality or 'methodological flaws' (Peters & Ceci, 1982). Similar effects were reported in later studies (Okike, Hug, Kocher, & Leopold, 2016; Ross, Gross, Desai, & et al., 2006). The initial report by Peters and Ceci initiated a fierce debate, with dozens of letters in response. Specifically, the perception that manuscripts were judged not merely on their content, but also according to 'circumstantial' factors such as the author's affiliation, background and personal characteristics invoked debate leading to the spread of double-blind review (Pontille & Torny, 2014). This format of review now presents a way of combatting referees' bias. However, in the digital age, critics have repeatedly pointed to the ineffectiveness of blinding author identities as a simple Google-search commonly enables identifying the authors of a 'blinded' manuscript.

4.2.2.2. Transparency: in reviewers we trust?

Interestingly, the issue of reviewer bias as a threat to the quality and fairness of peer review has not only led to the establishment of double-blind peer review, but also to its radical opposite: the system of open review. Currently, the term 'open review' is used for many different models and encompasses a wide variety of characteristics of peer review. A recent systematic review of the definitions for 'open peer review' demonstrates that scholars use the term to indicate processes in which, among others, the identity of the authors and reviewers are public, the review reports themselves are openly available, or the review process allows reviewers and/or authors to interact with each other (Ross-Hellauer, 2017). In this chapter, we use the term 'open review' merely to indicate that the identity of the authors and reviewers are mutually known to each other.

Open review gained momentum in the late 1990s, with the decision of the *British Medical Journal* to publish both reviewer names and reviews (Smith, 1999). Other initiatives followed, most notably in the biomedical sciences (Amsen, 2014).

The rationale for choosing an open system of peer review is transparency. Its advocates argue that open review leads to more constructive feedback, reduces reviewers' bias and gives credit to the reviewer (Godlee, 2002). Thereby it addresses some of the same concerns as those raised by the double-blind format, but with a radically opposite strategy. In addition, open review could reduce the chance of reviewers taking unfair advantage of their position as reviewer, either by plagiarising the manuscript under review, unjustly delaying its publication or advising rejection for unjust reasons (Godlee, 2002; Pontille & Torny, 2015; Tennant et al., 2017; Walker & Rocha da Silva, 2015).

The system of open peer review claims to contribute to reviewer evaluation, in response also to questions regarding the integrity or fairness of reviewers, rather than the integrity or quality of the evaluated manuscript. This is especially pertinent in systems that communicate reviewers' identities not only to the authors, but also to the general readership. In addition, formats of open review, in which the review reports are published alongside the article, provide another measure to increase transparency and therefore invoke scrutiny of reviewers. The emergence of the open review format hence allows surveillance of a system that has criticism as its major task.

In contrast, opponents of the system have stressed that open review could pose a threat to the quality of reviewing. This would especially be a concern when junior researchers are to review manuscripts by more senior colleagues, fearing professional reprisal if they submit negative reviews. In general, scholars have expressed concern about reviewers being milder in open review forms, thereby leading to more, and potentially poorer, manuscripts being published (Ross-Hellauer, 2017).

4.2.3. Technological advances in peer review

From the 1990s onwards, various technological advances paved the way for novel development of the peer review system. This opened possibilities which include new timing of the process, such as post-publication peer review (see 2.3.1); publishing more articles, while allowing a shift of review criteria from importance to rigour (see 2.3.2); the advent of automated checks and similar software tools (see 2.3.3); further specialisation of peer review (see 2.3.4); and more communication during the review process (see 2.3.5). Using these headings, we will attempt to describe the bewildering experimentation that erupted in the age of the internet. As we will show, these changes were not just driven by technological possibilities, but also by the interplay between technological potential and specific concerns about peer review's imperfections.

Even so, besides opening up possibilities for a wide range of novel peer review formats, arguably the most important development brought on by the advent of digitization, lies in the technical infrastructure facilitating review. This mainly affected the possibility of contacting and finding suitable reviewers much more quickly than before. Accessing researcher's webpages and email addresses allowed for much faster circulation of manuscripts and review reports, potentially increasing the speed and efficiency of the review process enormously. In the remainder of this section we will focus on the intellectual aspects that, facilitated by new technologies, affect the actual review process.

4.2.3.1. The timing of peer review in the publication process

Traditionally, peer review occurs between the submission and publication of a manuscript. In this format, editors receive a manuscript and possibly send it to outside reviewers or an editorial committee, who advise whether a manuscript is good enough to be published. Over the last two decades, two new forms of peer review have emerged that change the chronology of the reviewing. Firstly, there is a format in which manuscripts are evaluated after publication, the *post-publication peer review*, and secondly, a system in which articles are reviewed prior to submission to the journal, a format called *registered reports*.

4.2.3.2. Post-publication review and preprint servers

In the 1990s, several studies demonstrated that peer review is potentially biased, slow, unreliable and inconsistent (e.g. Easterbrook et al., 1991; Eckberg, 1991; Lock, 1985; Peters & Ceci, 1982), thereby nourishing the desire for alternative models and the formation of preprint archives. Especially the system's indolence and inconsistency were indicated as reasons for the formation of post-publication peer review. Preprint servers were established, based on already existing archives of print-based mail exchanges in high-energy physics. Even though some forms of disseminating preprint articles have been in place since the 1960s (Larivière et al., 2014), the advent of the internet and digital technologies enabled the establishment of large and fast-operating archives in which authors could freely upload their manuscripts, thereby bypassing publishers. In these archives, manuscripts usually go through a minor evaluation to check whether they meet minimal standards of academic writing (Gunnarsdottir, 2005; Walker & Rocha da Silva, 2015). Subsequently, the actual review is done by community members who comment on the manuscript, either via personal or public communication. Authors can then improve the manuscript and upload new versions to the archive (Bohlin, 2004; Gunnarsdottir, 2005). Originating in physics, astronomy and mathematics, the preprint servers have found their way to other scientific disciplines, with similar servers set up for biology, engineering and psychology (Tennant et al., 2017).

At first, these preprint servers were mainly used by authors to make preliminary versions of their articles available, before submitting the final version to a peer-reviewed journal. However, with the enormous increase in submissions to preprint archives recently (Walker & Rocha da Silva, 2015), these servers have themselves become a major communication channel in, which some authors use as a sole venue for their manuscripts (Fitzpatrick, 2009). This fast dissemination method allows scholars to keep up with each other's work, provides a way of crediting the first author(s) for presenting novel findings and thereby solving priority issues, and allows readers to comment on early drafts of a paper. Ideally this results in exchanging ideas and improving the manuscript (Bohlin, 2004; Gunnarsdottir, 2005). However, despite an increased number of papers being deposited in arXiv and other preprint servers, the proportion of scientific literature made available in this fashion is still very low, and limited to only a few academic fields (Walker & Rocha da Silva, 2015).

Besides being used in preprint servers, post-publication review has gradually also been taken up by journals and publishers. The first journal to implement this format was *Electronic Transactions in Artificial Intelligence* in 1997 (Fitzpatrick, 2009; Pöschl, 2012). Introducing this new review form served mainly to accelerate knowledge distribution. Especially in the last few years, a number of journals have switched to this post-publication model of peer review. Finally, several independent platforms such as *PubPeer* were established, in which post-publication review of any published manuscript can be done, independent of what kind of review it went through during the publication process (Knoepfler, 2015). These platforms will be discussed in more depth in section 4.2.5.2.

Besides responding to concerns of speed and consistency, introducing open archives resulted in several new expectations of peer review. Rather than being a selection or gatekeeping mechanism, according to some scholars, reviewing should be transformed into a filtering process that presents relevant literature to researchers in the right fields: "… peer review needs to be put not in the service of gatekeeping, or determining what should be published for any scholar to see, but of filtering, or determining what of the vast amount of material that has been published is of interest or value to a particular scholar" (Fitzpatrick, 2010). Hence, the peer review system should not be thought of as a way of stopping 'irrelevant' research from being published, but merely as a way of directing the right literature to the right reader. By lowering the threshold for publishing manuscripts, including those reporting negative results, this system also serves as a response to the apparent bias in published manuscripts towards positive results (Dickersin, 1990). Some consider countering this bias an important measure to restore the integrity of the scientific literature (van Assen, van Aert, Nuijten, & Wicherts, 2014).

The system of publishing articles prior to being reviewed serves to enhance research integrity in two additional ways. Firstly, the publication of preprints can improve the detection of fraudulent research. There are several cases in which authors, often after previous rejections from journals, alter their data and/or conclusions to deliver a more positive result. Such cases of spin or data manipulation are more easily detected if preprints of a manuscript have been published. In this way, preprints serve as a means of detecting authors' improper behaviour. Secondly, preprints also serve a function in recognizing reviewer misbehaviour, such as plagiarising manuscripts under review or delaying review to obtain an advantage in priority issues.

Besides these advantages, establishing preprint servers and introducing electronic publishing in general have had a major effect on the costs of publishing and of obtaining access to scientific literature. Continuing a trend started by large publishing companies that created a publishing market in the 1980s, the introduction of electronic publishing in the mid-1990s brought a massive increase in the number of journals, articles and citations (Larivière et al., 2015). This number shows a concentration of articles and citations in the outlets of large commercial publishers. In the fields of both medicine and natural science, as well as in the social sciences, large commercial publishers bought journals from smaller publishers and established new journals themselves, in order to drastically increase their market share in academic publishing (Fyfe et al., 2017). One of its consequences has been a sharp increase in journal prices and the establishment of 'big deals' with (university) libraries (Larivière et al., 2015).

4.2.3.3. Registered reports

A second major development regarding the timing of peer review in the publication process, has been the establishment of the registered reports system, first introduced by the journal *Cortex* in 2013 (Chambers, 2013; Mellor, 2016). In this form of peer review, which is still restricted mainly to medical fields and psychology, manuscripts are usually reviewed in two stages. The initial and most important review stage takes place after the study has been designed, but prior to data collection. At this stage, only the rationale for undertaking the research, the research questions and the research methodology are reviewed. On the basis of these criteria a study is either accepted or rejected, before any data has been collected. In the subsequent stage, after data collection and analysis have taken place, authors compose their manuscript by adding their results and conclusions to the registered report. The final manuscript can then be reviewed on the basis of consistency and adequately having drawn conclusions from the data. Taking this further, *BioMed Central (BMC) Psychology* recently published the first articles that had been through a completely 'results-free review', in which the second phase of peer review was entirely omitted (Clark, 2017).

The main reason for introducing registered reports lies in the alleged 'replication crisis' in several areas of science. Registered reports are a means of making the execution of replication studies more attractive: "Peer review prior to data collection lowered the barrier to conduct replications because authors received editorial feedback about publication likelihood before much of the work was done" (Nosek & Lakens, 2014). Generally, many journals are reluctant to publish replication studies, which potentially deters scientists from performing them: "If journals will not publish replications, why would researchers bother doing them?" (Nosek & Lakens, 2014). Prior clarity about publication chances based on research design, and not on the novelty of results, could encourage replication studies. In addition, registered reports can alter incentives for authors and reviewers to act with more integrity, in the sense that methodological accuracy and transparency become more important than pleasing possible readers: "Because the study is accepted in advance, the incentives for authors change from producing the most beautiful story to the most accurate one" (Chambers, Feredoes, Muthukumaraswamy, & Etchells, 2014) and "review prior to data collection focused researchers and reviewers to evaluate the methodological quality of the research, rather than the results" (Nosek & Lakens, 2014). Hence, contrary to innovations that are mainly designed to allow additional scrutiny of the reviewer, registered reports address the integrity of the author and promise to reduce researchers' rewards for dubious behaviour.

4.2.4. The changing peer review criteria

Besides yielding the system of pre-print archives, the advent of the internet and large databases further enabled journals to publish nearly unlimited numbers of articles. Novel publishing strategies and related peer review models became possible. A major development in this respect came with the launch of the open access journal *PLoS ONE,* by the *Public Library of Science* (PLoS), in 2006. In this journal's review process and business model, reviewers are asked to base their recommendation for acceptance or rejection purely on the soundness and validity of the research, comprising the methodology, soundness of results and reporting. According to the journals' philosophy, reviewers should not judge the novelty, relevance or importance of research, which should be left to the reader and wider community (Hames, 2014). By focussing on rigour and (ethical) soundness of research, the journal aims to ensure that useful results will all be published, and to prevent subjective assessment of a study's importance or relevance.

Since its launch, *PLoS ONE* has been one of the most rapidly growing publication venues. In 2013, it published over 30,000 articles (Davis, 2017; Graham, 2014), turning itself into the largest open access publisher and one of the largest scientific journals worldwide.

Subsequently, other journals and publishers, such as *BMJ Open* and *SAGE Open*, have adopted the same non-restrictive review model (Hames, 2014).

These changes in review criteria content and in how they select have their roots in discussions on scientific integrity. Several motives have prompted PLOS and other outlets to focus on rigour and soundness of research (BMJ Open, 2018; PLOS, 2018; Sage Open, 2018). First, it ensures the publication of all 'valid' research, irrespective of the study's perceived importance by reviewers. This, among other things, facilitates the publication of replication studies and negative results (BMJ Open, 2018). In addition, the journals aim to deter authors from overstating results or otherwise engaging in questionable research practices in order to meet reviewer standards of importance. This review format was therefore partly set up to promote scientific integrity, not so much by increasing the detectability of fraudulent research or misconduct, as by stimulating scientific integrity from the outset (Hames, 2014). However, this system could unintentionally also create new concerns regarding the literature's integrity, for instance by overloading it with research of little relevance, or by creating incentives and opportunities to publish (irresponsibly) high numbers of articles.

Partly due to the less restrictive review process, the number of papers published in outlets employing this non-restrictive review model has grown rapidly. As a result, new challenges have emerged in the publication process. One of them is finding enough qualified reviewers to handle all submissions. For example, by 2014, *PLoS ONE* used more than 70,000 reviewers to process all submissions and the average review time drastically increased since *PLoS'*s launch in 2006 (Davis, 2017; Graham, 2014). In addition, the high number of published articles generates a growing concern about the scientific literature becoming unmanageably large, resulting from an abundance of articles many of which add little to the stock of knowledge. At the least, this creates a growing need for further filtering to ensure researchers can cope with the enormous number of potentially interesting papers. Novel systems will need to be established to draw readers' attention to articles that are most likely to be useful to them.

4.2.4.1. Introduction of software tools to the review process

In addition to the possibilities of preprints and virtually unlimited numbers of publications, the advances of the internet and new digital technologies also offered dedicated technical support to assess whether papers are publishable. Technical assistance in various formats has by now become standard practice and most certainly will be extended in the (near) future (BioMed Central, 2017). The first major technical assistance to be implemented in peer review was plagiarism detection software. Copying text from various sources became easier than before once electronic publishing was introduced, and with internet assistance added concerns about plagiarism spread throughout academia, regarding student papers as well as research articles (Auer & Krupar, 2001). However, the first versions of plagiarism detection tools originated in the context not of textual plagiarism, but the copying of parts of programming code (Faidhi & Robinson, 1987). Only in later phases did this evolve into plagiarism detection tools for journals to recognise unwarranted copying in research articles (Ercegovac & Richardson, 2004). Currently, the vast majority of journals and publishers use some form of plagiarism detection tool to assist in peer review (Elizondo, De Rijcke, & Van Leeuwen, 2017), the *CrossCheck* system being the most common (Zhang, 2010).

Besides assisting with plagiarism detection, online tools have recently come to assist reviewers in several other ways. Most notably, some automatic analysis that checks for the correct use of statistics in manuscripts has been introduced (Epskamp & Nuijten, 2014). Aided by artificial intelligence technologies, software protocols have been developed to assess completeness, consistency and validity of statistical tests in academic writing, thereby specifically targeting the (intentional) misuse of statistics in research, which some believe to be a major factor in the alleged integrity and reproducibility crisis (Munafò et al., 2017). Additionally, the assistance of software in detecting image manipulation, which is considered an increasing form of fraud in various research areas, has successfully been implemented by several journals (Scheman & Bennett, 2017). However, we should note that the use of image and statistics scanners is still rare and limited to specific research areas, most notably the medical sciences, physics and psychology.

In the future, automated computer software could well play an even more substantive role in the review process. Aided by machine-learning techniques, it has already become possible to check for bad reporting (failing to report key information or inconsistencies in reporting), data fabrication, and image manipulation. In addition, Chedwich deVoss, the director of *StatReviewer,* even claims: “In the not-too-distant future, these budding technologies will blossom into extremely powerful tools that will make many of the things we struggle with today seem trivial. In the future, software will be able to complete subject-oriented review of manuscripts. […] this would enable a fully automated publishing process – including the decision to publish.” (BioMed Central, 2017) Although one should have some reservations on such predictions of a technological future, they do reveal some of the current expectations for peer review.

The implementation of software-aided detection mechanisms requires us to increasingly distinguish the ‘peer review process’ from ‘peer review’. Due to digital technologies and software tools normally not being imposed on the reviewer, but handled by the journal’s staff or editorial team, the review process now entails much more than individual reviewers merely doing quality assessment. Therefore, the use of these tools should be considered an additional step in the review process, rather than an integral part of the actual review by a ‘peer’.

In sum, digital technologies and software tools based on machine learning and artificial intelligence have been incorporated in some parts of the peer review process. Their primary use currently is to detect plagiarism, text recycling and duplicate publication; to analyse and review of statistics and statistical analysis in specific fields; and to a lesser extent to detect figure or data manipulation (BioMed Central, 2017; Epskamp & Nuijten, 2014; Fyfe et al., 2017; Tennant et al., 2017). All of these clearly target the integrity of research and authors under review, and specifically target those practices that have traditionally been labelled as outright fraud, namely falsification, fabrication and plagiarism. Hence, these digital technologies are a primary example of innovations in peer review specifically targeted to increase the detectability of fraudulent or erroneous research.

4.2.5. Novel actors and cooperation in the review process

Over the past decades, new actors have joined the review process, thereby compelling peer review itself to become more specialised. This applies to its content, for example introducing specialised statistical reviewers, as well as to the process, with commercial parties specialising in the reviewing process.

4.2.5.1. Statistical review

During the second half of the 20th century, the use of statistics in research articles has drastically increased, especially in medical and psychological research (Altman, 1998). The use of ever more complex, statistical models raised concerns about the validity of some statistical methods. In response to the publication of reviews demonstrating that published articles often report statistically unsound analyses, journals and publishers set out to dedicate more attention to statistical analyses in their review processes. From the 1960s onwards, several journals included specialist statistical reviewers to judge the soundness and quality of methodology and statistics in submitted manuscripts, again mainly in medicine and psychology (Altman, 1998; Schor & Karten, 1966).

Despite repeated demonstration of widespread statistical and methodological errors in (medical) research, increasing the use of specialist reviewers to check for such errors has been slow. A 1985 survey of journals and publishers showed that only a very small proportion of journals paid specific attention to those factors in their review process (George, 1985). Fuelled by current issues regarding research reproducibility and replicability (Ioannidis, 2005; Munafò et al., 2017), many still agitate for intensifying the scrutiny of statistics. One consequence was the formation of a project called SMARTA, which brings together members of international statistical societies to assess the use of statistics in biomedical literature (Goodman, 2017). Such developments may well lead to statistics being given more attention in review, and even to further specialisation of reviewers.

4.2.5.2. Commercial review platforms

Besides the introduction of specialist statisticians to the review process, a new set of refereeing bodies has recently emerged (Tennant et al., 2017). In these new initiatives, review is dissociated from the journal in which the article is published. Several formats have emerged, of which one arranges the reviewing of articles prior to publication by independent third parties. Platforms such as *Peerage of Science, RUBRIQ* and *Axios Review* (Peerage of Science, 2017; Research Square, 2017b) provide tools and services to conduct reviews and forward submitted manuscripts along with referee reports to a journal. In this way, reviews can be done faster and more efficiently, also by reducing the likelihood of a manuscript going through multiple reviews for various journals.

Notably, one of the commercial services providing independent review, *Research Square*, specifically focuses on the promotion of scientific integrity with the assistance of software tools. The platform attaches badges to manuscripts that pass various tests addressing specific "aspects of a research manuscript that [are] critical for ensuring the integrity and utility of the scholarly record" (Research Square, 2017a). It awards such badges after an 'integrity precheck', 'statistical check', 'figcheck' and 'sound science check', to name just a few. Thereby the platform explicitly claims that such assessments can indeed be made as part of the peer review process. In a pilot study on submissions to two medical journals, *Research Square* actually reports frequently detecting integrity issues much more frequently than would be expected considering current estimates on the extent of misconduct in science (Pattinson & Prater, 2017).

In addition to the systems providing pre-publication review, other independent platforms have emerged, such as *PubPeer* (PubPeer Foundation, 2017), in which any reader can comment on any published manuscript. These systems constitute examples of post-publication review independent of journals and publishers. These new trends have increasingly widened the definition of a *peer*, so that the term now refers not only to a small cluster of editor-selected experts, but to anyone who feels capable of understanding and evaluating a given piece of research. This emergence of an 'extended peer community' gives rise to novel challenges concerning the role of expertise in peer review, as well as to questions regarding who has the right and competence to judge the quality, soundness and relevance of scientific research (Funtowicz & Ravetz, 2001). In addition, some scholars have expressed concern about the role of public forums in signalling cases of problematic research, as this can lead to stigmatising researchers without them having due opportunity to defend themselves.

4.2.5.3. Cooperation in review

Another way of reducing the burden on peer review lies in the concept of 'cascading peer review'. This model, which was first consistently used at the beginning of the 21st century, became common practice in the *BMJ* journals in 2010 (Davis, 2010) and is now widely used, especially by larger publishing houses. The system aims to avoid final rejection of a manuscript after peer review by redirecting critically reviewed manuscripts to potentially more suitable journals. In practice, larger publishing houses often use this system of redirecting manuscripts that are rejected for publication in top-tier journals to lower-tier journals within their portfolio. However, currently peer review consortiums are formed to facilitate the practice of cascading review in smaller publishing houses as well (Barroga, 2013). The system of cascading reviews responds to the growing expectation of the review system to not necessarily act as a gatekeeper, but rather serve as a mechanism to direct relevant research to the right audience. As the system of cascading reviews is designed to avoid final rejection, it potentially focuses on the relevance of a manuscript, rather than its soundness, quality or integrity. This could have major implications for the scientific publishing system. Low rejection rates can raise questions about the veracity of knowledge, tolerance for 'alternative facts' (Sismondo, 2017), and rating the value of publications in research career assessment.

Both of these peer review models, cascading review and review by third parties, are designed to assure that one single manuscript does not have to go through multiple rounds of peer review. Sharing review reports, either from a commercial party or from a rejecting journal, with a potentially interested journal, decreases the number of reviewers assessing a single manuscript (Barroga, 2013; Tennant et al., 2017). This answers to a concern of the past few decades, that the peer review system is getting overloaded (Kovanis, Porcher, Ravaud, & Trinquart, 2016). In addition, automatically (re-)directing manuscripts to the most suitable journal after review could reduce perverse incentives for authors, such as rewarding work in which conclusions are overstated to get the study published. On the other hand, it could also work in the opposite direction in that relaxing review standards might tempt authors to neglect nuances in the confidence that their work will eventually get published somewhere anyway.

4.2.5.4. New openness: discussion during review

Finally, the advent of digital technologies has paved the way for new levels of openness in the review process. Some journals, most notably journals at EMBO (*European Molecular Biology Organization)* and the *elife* journal, have attempted to improve editorial decision making by introducing interactive stages in the review process, during which reviewers and editors can share or discuss their reports and opinions on a manuscript before communicating a final decision to the author (EMBO Press, 2017; Schekman, Watt, & Weigel, 2013). In 2011, the *elife* journal pioneered this new model, referring to movements concerning transparency and accountability in peer review as rationale (Schekman et al., 2013). Later, other journals followed suit, partly related to the open science movements in which review reports are not only shared among reviewers, but also with the general readership.

The *Frontiers* journals launched in 2013 later established a more radical variant of this peer review model, labelled the 'collaborative peer review'. This process set up a review forum for interaction between authors and reviewers. Such forums serve as an interactive stage in the review process, during which authors and reviewers discuss the paper online until they reach agreement on the most effective way to improve its quality (Frontiers, 2014; Hames, 2014).

4.3. Diversity of forms

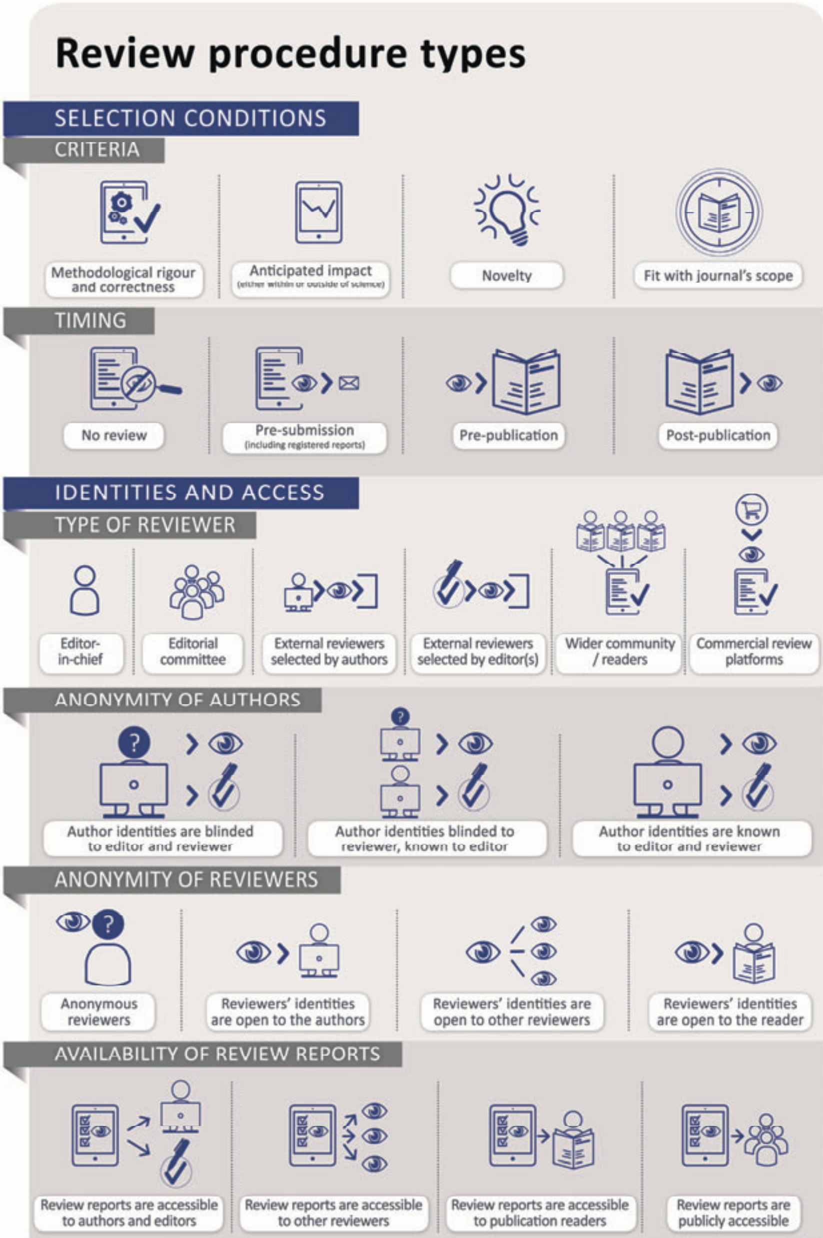
Concluding from the overview in the previous subsections, the diversity of peer review forms has clearly increased significantly over the past few decades, thereby also diversifying the practice of quality control in research.

Structuring the discussion in the preceding subsections, the distinguishing attributes of various review forms can be classified along four dimensions, namely the selection conditions, the identity and access among actors involved, the level of specialisation in the review process, and the extent to which technological tools have been introduced. Each of the attributes has a range of possibilities, as presented in Table 4.2. The typology discloses a clear ordering of the current variety in peer review, providing a solid foundation for further research on, e.g., how often various forms are used, or how various peer review forms relate to other properties of the publication system.

4.4. Diversity of expectations

4.4.1. What is the publication system for?

The overwhelming variety of current forms reflects the substantial variation in what is expected of peer review. Some of these expectations relate closely to diverging purposes of scientific publishing, which have also shifted over time and are more disparate than one might expect. At first, the main purpose of scientific journals was to settle priority claims, as a social device to establish and maintain intellectual recognition. Specifically using journals for the publication of essentially new knowledge is a relatively recent phenomenon (Fyfe et al., 2017; Mahoney, 1985). The main motivation for the prototype of the modern scientific manuscript was "the establishment and maintenance of intellectual property. It was the need which scientists felt to lay claim to newly won knowledge as their own, the never-gentle art of establishing priority claims" (de Solla Price, 1963). This original purpose of journals became even more apparent in the system of *pli cacheté* that was in place in many journals during the 18th, 19th and even 20th century (Erren, 2009).


Review procedure types
SELECTION CONDITIONS
CRITERIA
Methodological rigour and correctness
Anticipated impact (either within or outside of science)
Novelty
Fit with journal's scope
TIMING
No review
Pre-submission (including registered reports)
Pre-publication
Post-publication
IDENTITIES AND ACCESS
TYPE OF REVIEWER
Editor-in-chief
Editorial committee
External reviewers selected by authors
External reviewers selected by editor(s)
Wider community / readers
Commercial review platforms
ANONYMITY OF AUTHORS
Author identities are blinded to editor and reviewer
Author identities blinded to reviewer, known to editor
Author identities are known to editor and reviewer
ANONYMITY OF REVIEWERS
Anonymous reviewers
Reviewers' identities are open to the authors
Reviewers' identities are open to other reviewers
Reviewers' identities are open to the reader
AVAILABILITY OF REVIEW REPORTS
Review reports are accessible to authors and editors
Review reports are accessible to other reviewers
Review reports are accessible to publication readers
Review reports are publicly accessible

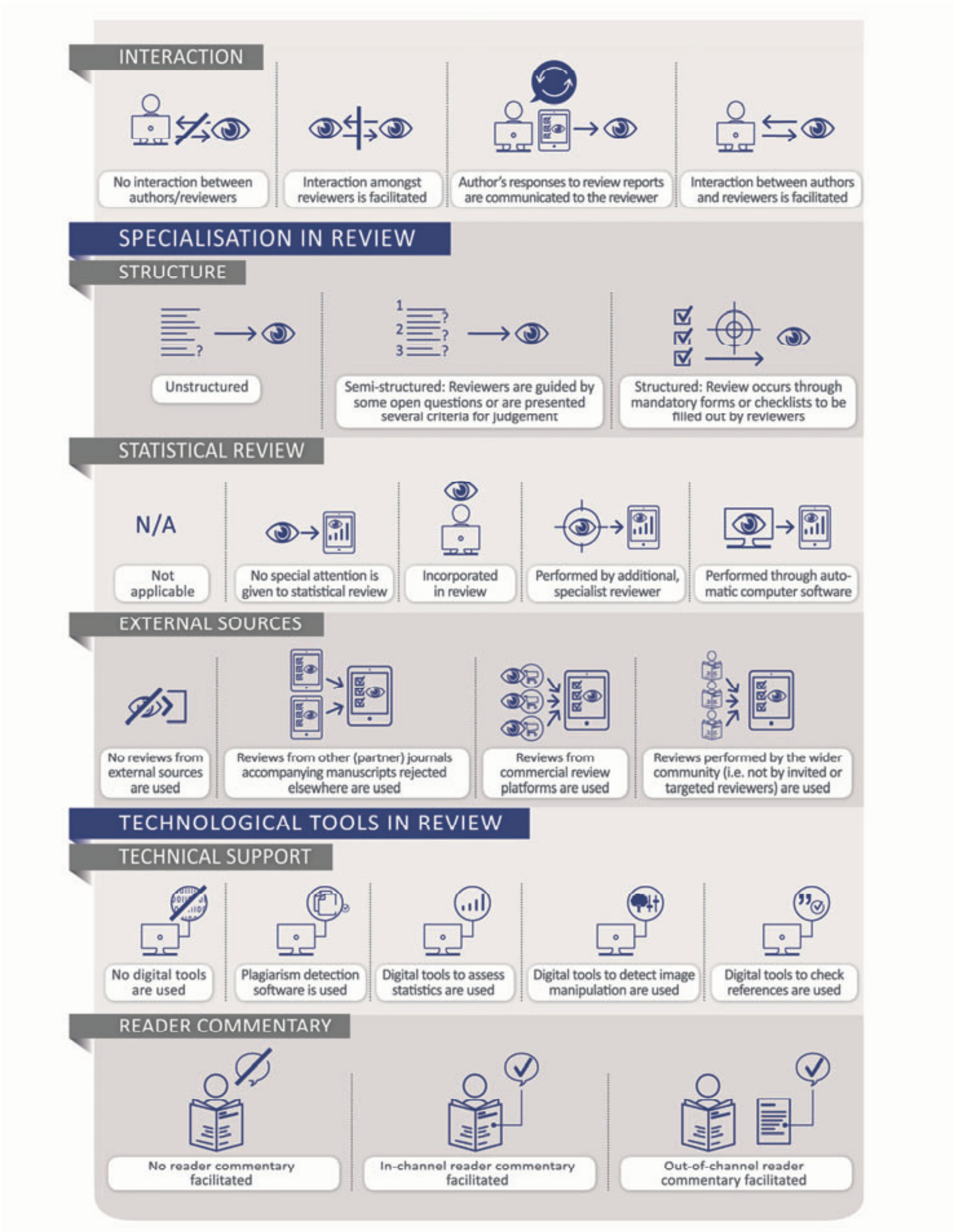


Table 4.2 Forms of peer review categorized by dimension and attributes. Illustration by Paul de Groot (Dikke Punt).

In this system, authors sent their manuscripts to journals in sealed envelopes, to be opened only at the author's request. This allowed researchers to submit discoveries about which they were uncertain, while allowing them to claim priority in case other researchers wanted to publish the same or very similar results (Erren, 2009).

Besides settling priority issues and providing due credit to authors, scientific publishing has given rise to three other major expectations. The first is to facilitate the exchange of knowledge and ideas among scholars working in the same narrow field, providing the specialised communication on which research progress depends. The second is to form a constantly evolving historical archive of scholarly thought (Solomon, 2002). The third is to provide a hierarchy of published results based on peer-defined excellence (Biagioli, 2002; Cronin, 2005; Solomon, 2002; Weller, 2001). Or, more briefly stated: "In their ideal, journals do not just transmit information; they filter, evaluate, [store] and unify it" (LaFollette, 1992).

Peer review plays a major role in two of these functions, namely in facilitating the exchange of ideas among scholars and providing a hierarchy of published results. Firstly, regarding the exchange of knowledge there "slowly developed the practice of having the substance of manuscripts legitimated, principally before publication although sometimes after, through evaluation by institutionally assigned and ostensibly competent reviewers" (Zuckerman & Merton, 1971). As such, peer review is "the instrument for ensuring trustworthiness" in science (Cronin, 2005). Kassirer and Campion explained that the review process "is probably best described as an intellectual exercise to detect flaws in experimental design, presentation, interpretation, and the overall importance of a study; at a certain point a manuscript reaches the rejection threshold, which tips the editorial scale toward its rejection" (Kassirer & Campion, 1994). That peer review plays a pivotal role in validating research is widely accepted (BioMed Central, 2017; Bornmann, 2011; Rennie, 2003; Ware, 2008; Weller, 2001). This could be the most important aspect of scientific publishing. "Ensuring the accuracy and quality of the information contained in a manuscript as well as the clarity of the writing and quality of the presentation is far more important and in some cases crucial" (Solomon, 2002). The role of quality assurance is attributed to all involved in the review process, not only to reviewers, but specifically also to editors (Garfield, 1983).

Secondly, academic publishing provides a hierarchy of published results. Peer review is particularly instrumental in sustaining this hierarchy, by establishing a continuum ranging from top-tier journals to outlets of lower status. An interesting example, in which this expectation of peer review becomes particularly visible, is the mathematics '*arXiv* overlay' journal *SIGMA (Symmetry, Integrability and Geometry: Methods and Applications)*. This electronic journal, does not 'publish' or archive its own articles, but merely adds a signature to articles on *arXiv*, after having reviewed them (SIGMA, 2005). As such, the journal does not facilitate the spread or storage of knowledge, but rather assesses articles' quality and classifies them as sound science. Such classification distinguishes reviewed articles from other manuscripts on arXiv, thereby raising them in the hierarchy of published results. This is not merely an epistemological exercise, but also a quest for recognition of published manuscripts. 'Peer reviewed publications' increasingly serve as the basis of research evaluation, be it in grant applications, organisational audits, job interviews or tenure decisions (e.g. Hicks, Wouters, Waltman, De Rijcke, & Rafols, 2015). Therefore, elevating manuscripts from the status of preprints to peer reviewed articles serves as a mechanism that not only warrants quality, but also establishes a form of recognition and credit.

Given this hierarchical allocation of recognition, the content of review criteria has become increasingly contentious. Questions arise regarding whether journals merely judge adequacy, consistency and methodological accuracy (e.g. the *PLoS* format), or whether they also account for relevance, perceived impact or usefulness to future research. As a result, tensions have arisen regarding the expectations of what peer review can establish.

Thirdly, the academic publishing system is expected to provide equal and fair opportunities to all participants. As was indicated in section 4.2, due to the central role peer review has played in its development, this major expectation evolved more gradually (Godlee, 2002; Ross-Hellauer, 2017). Equal assessment opportunities required submitted manuscripts to be judged on content only, without attention to circumstantial information such as the authors' affiliation, gender or background. Here, referring to peer-reviewed articles in research career assessment is crucial.

A fourth major expectation of the academic publishing system, and of peer review in particular, emerged in a debate regarding the system's effectiveness in tracing misconduct. Despite the recognition of peer review's crucial role in ensuring the accuracy and quality of scientific work, since the late 1980s its capacity to detect fraud has been a growing concern (Robin & Burke, 1987). The discussion was fuelled by reports on major scandals in science, followed by substantial public outcry, including on the Darsee and Baltimore cases (LaFollette, 1992, 1994; Stewart & Feder, 1987). Under the threat of intensified congressional involvement in the USA, the scientific community used the peer review system as one of their main defence arguments. Former National Academy of Sciences (NAS) president Philip Handler called the problem 'grossly exaggerated' and expressed complete confidence in the existing system "that operates in an effective, democratic and self-correcting mode" (Guston, 2007). Similarly, National Institutes of Health (NIH) director Donald S. Fredrickson testified "misconduct was not and would never be a problem because of scientific self-regulation" (Guston, 2007). In this context, the late 1980s started to exhibit the first major signs of peer review being put forward as a means of safeguarding the scientific enterprise from fraud and misconduct.

However, this argument received criticism from the outset (Garfield & Welljams-Dorof, 1990; Kochan & Budd, 1992; LaFollette, 1992). In the founding days of scientific societies and scientific journals in the 17th century, general consensus maintained that the responsibility to guarantee the credibility and soundness of the research record did not lie with the professional society or the publisher (Kronick, 1976; Manske, 1997). Editors and publishers who still agree that “the peer review system was never designed to detect fraud” (LaFollette, 1992), implicitly rely on other institutions and whistle-blowers to detect fraudulent data or plagiarised material (Yankauer, 1990).

Regarding journals’ responsibility to act against misconduct, several actors arrived at different opinions. Even though many journals introduced some measures to address misconduct, for example by issuing retractions and corrections, many believed that more should be done, especially in journals taking a gatekeeper role. In the same period, mainly driven by considerable increases in subscription and submission fees, librarians and authors became more demanding regarding the validity and integrity of published research. At the 1989 annual meeting of the Society for Scholarly Publishing, Hendrik Edelman of Rutgers University declared to generous support of fellow librarians that “given the high costs of subscriptions, publishers should guarantee 'fraud-free' products" (LaFollette, 1992). The dramatic price increases resulted in heightened agitation for quality control, which was later reinforced by other scholars and librarians (Russel, 2008).

4.4.2. Tensions regarding peer review and research integrity

The expectation that publishers should be responsible for ensuring the integrity of the scientific literature comes from two sides. Firstly, politicians and funding agencies demand their money be put to good use and thus insist on quality control for the work they finance. From this perspective, peer review plays a role in public accountability. Secondly, authors and librarians increasingly demand value for money, given the high submission and subscription fees of academic journals. Peer review then becomes a matter of product quality.

Despite this two-fold call for editors and publishers to take responsibility, many actors, primarily editors and publishers themselves, express disquiet about peer review’s ability to detect fraudulent research. This became strikingly clear in Weller’s 2001 seminal work on the peer review system in which she argues that “the underlying strength of editorial peer review is the concerted effort by large numbers of researchers and scholars who work to assure that valid and valuable works are published, and conversely, to assure that invalid or non-valuable works are not published." At the same time, just a few paragraphs later, she asserts: "Fraudulent behavio[u]r on the part of a researcher has not been discussed, primarily because of the limited ability of reviewers or editors to identify fraudulent activities or fabricated data"

(Weller, 2001). This clearly points to the tension between actors' desires and expectations regarding the peer review system and the abilities that can reasonably be attributed to it.

In spite of such diverging expectations, some of the current innovations clearly move towards peer review as a factor in improved research integrity. The novel pilot by *Research Square,* providing badges for 'research with integrity', arguably indicates that peer review *can* detect fraudulent behaviour if it is specifically designed to do so (Pattinson & Prater, 2017; Scheman & Bennett, 2017). In addition, different forms of fraudulent behaviour should be properly differentiated. As has been noted before, it is notoriously difficult for peer reviewers to detect cases of intentional data manipulation or fabrication. However, one can expect several kinds of questionable research practices that are thought to be much more common (John, Loewenstein, & Prelec, 2012; Martinson, Anderson, & de Vries, 2005) to be detected by reviewers, as in cases of spin, inappropriate use of statistical analysis or data cooking. In addition, the use of software tools to detect (self-)plagiarism (Horbach & Halffman, 2017a), image manipulation and poor statistical analyses has recently increased the detectability of outright misconduct. Detecting these forms of misbehaviour might not reasonably be expected of a single peer reviewer, but can increasingly be expected from *the peer review process*.

4.5. Conclusions

Our review demonstrates the remarkable diversity in contemporary models of peer review. Ever since its establishment, peer review has developed into a wide and expanding variety of forms. The development of review forms can be systematised along four dimensions: (i) the selection conditions, including the timing of the review and its selectiveness, (ii) the identity of and interaction between the actors involved, (iii) the levels of specialisation within the review process, and (iv) the extent to which technological assistance has been implemented in the review system. These four dimensions cover an array of peer review processes than can map both the historic and current forms of peer review, and suggests some axes of possible future development. In addition, this classification can serve as the basis for future empirical research assessing the quality, effectiveness or feasibility of the diverse peer review forms.

Many of the recent innovations have come about as a response to shifting expectations of what peer review can or should achieve. Whereas the post-war dissemination of the system was presented as a form of quality-guarantee, it later responded to concerns regarding inequality in science, the efficiency of the publication system, and a perceived increase in scientific misconduct. Currently, four major expectations of the peer review system can be distinguished: (i) assuring quality and accuracy of research, (ii) establishing a hierarchy of published work, (iii) providing fair and equal opportunities to all actors, and (iv) assuring a fraud-free research record. Different peer review formats will be preferred, depending on which of these expectations take precedence, as not all of these expectations can be easily combined. For example, a hierarchy of published work through a review process that favours highly relevant, high-impact research can jeopardise equal opportunity, and potentially even accuracy or integrity, as authors go to extreme lengths competing for attention at the top.

To date, very little systematic research has investigated whether peer review can live up to these differing expectations. There is limited evidence on peer review's capacity to guarantee accurate and high-quality research. Additionally, the potential of peer review to distinguish between possibly relevant and seemingly irrelevant research, or between fraudulent and non-fraudulent research, has not been adequately studied. This leaves a clear knowledge gap to be addressed in future empirical research. Our classification of review forms can constitute a useful tool to set up such comparisons between review practices.

The existing discrepancy between what some expect of the system and what others believe it is capable of has led to several current tensions. Most notably, the expectation that the peer review system should be used in gatekeeping to prevent erroneous or fraudulent research, is problematic. Many have blamed peer review for not properly detecting erroneous research; however, simultaneously others claim it was never designed to do so. Recent new developments and tools in peer review suggest that it is increasingly possible to detect and filter erroneous or fraudulent research in the peer review process. However, more research is needed to investigate the extent to which these innovations can live up to the expectations.

Meanwhile, some of the fraud detection innovations in peer review seem to shift the modalities of knowledge validation. Whereas peer review used to rely on the inter-subjectivity of colleagues to check the objectivity of research, currently statistics scanners or image-checkers permit more automated judgement in peer review, which aims to reduce human judgement. From inter-subjective checking, the focus is shifting towards more mechanical forms of objectivity, with automated discovery as an uncomfortable asymptote (Daston & Galison, 2007).

These tensions about peer review's expectations and abilities, point to more fundamental shifts in ambitions for the scientific publication system. At first, the scientific literature was primarily perceived as a large (public) library containing reports on scientific research, review papers, discussion papers, and the like. While this view still prevails, we would argue that an additional frame has appeared, which presents the scientific literature as a database of accurate knowledge or 'facts'. This new frame, which seems specifically attractive to those holding realist and positivist views of knowledge, is witnessed, for example, in the belief that 'inaccurate knowledge' should be retracted from the literature. In the library frame, questioned research was addressed through further publications, referencing and commenting on earlier publications, without removing them. Propositions and knowledge claims, as well as their denials, co-existed in an inter-textual universe of scientific knowledge claims – some more, some less veracious. The publication system as a database insists on removing erroneous records and replacing them with newer, corrected versions through innovative technologies such as corrections, retractions, statistics-checks, or post-publication reviews, facilitated by the digital revolution in publishing. The publication system as database creates new expectations about a body of reliable knowledge, including the possibility of meta-studies or systematic reviews, which are in turn used as arguments to shift further towards a database model. Seemingly technical innovations in the peer review system could therefore be signs of far more fundamental shifts in notions of objectivity or the status of the knowledge contained in 'the scientific literature'.

4.6. References

Altman, D. G. (1998). Statistical reviewing for medical journals. *Statistics in Medicine, 17*(23), 2661-2674. Retrieved from <Go to ISI>://WOS:000077335600003. doi:10.1002/(sici)1097-0258(19981215)17:23<2661::aid-sim33>3.0.co;2-b

American Sociological Review. (1955). Notice to contributors. *American Sociological Review, 20*(3), 341.

Amsen, E. (2014). What is open peer review. Retrieved from https://blog.f1000.com/2014/05/21/what-is-open-peer-review/

Auer, N. J., & Krupar, E. M. (2001). Mouse click plagiarism: The role of technology in plagiarism and the librarian's role in combating it. *Library Trends, 49*(3), 415.

Baldwin, M. (2013). 'Keeping in the race': physics, publication speed and national publishing strategies in Nature, 1895–1939. *The British Journal for the History of Science, 47*(2), 257-279. Retrieved from https://www.cambridge.org/core/article/keeping-in-the-race-physics-publication-speed-and-national-publishing-strategies-in-nature-18951939/1DD54105D8A0B87F2CB44316B46241F5. doi:10.1017/S0007087413000381

Baldwin, M. (2015). Credibility, peer review, and Nature, 1945–1990. *Notes and Records: the Royal Society journal of the history of science, 69*(3), 337-352. Retrieved from http://rsnr.royalsocietypublishing.org/content/roynotesrec/69/3/337.full.pdf. doi:10.1098/rsnr.2015.0029

Baldwin, M. (2017). In referees we trust? *Physics Today, 70*(2), 44.

Barroga, E. F. (2013). Cascading peer review for open-access publishing. *Eur Sci Ed, 39*, 90-91.

Begley, C. G., & Ioannidis, J. P. (2015). Reproducibility in science. *Circulation research, 116*(1), 116-126.

Benedek, E. P. (1976). Editorial practices of psychiatric and related journals: implications for women. *American Journal of Psychiatry, 133*(1), 89-92.

Biagioli, M. (2002). From book censorship to academic peer review. *Emergences: Journal for the Study of Media & Composite Cultures, 12*(1), 11-45.

BioMed Central. (2017). *What might peer review look like in 2030?* Retrieved from figshare: https://doi.org/10.6084/m9.figshare.4884878.v1

BMJ Open. (2018). Reviewer guidelines. Retrieved from http://bmjopen.bmj.com/pages/reviewerguidelines/

Bohlin, I. (2004). Communication Regimes in Competition. *Social Studies of Science, 34*(3), 365-391. Retrieved from http://journals.sagepub.com/doi/abs/10.1177/0306312704041522. doi:doi:10.1177/0306312704041522

Bornmann, L. (2011). Scientific peer review. *Annual Review of Information Science and Technology, 45*(1), 197-245. Retrieved from http://dx.doi.org/10.1002/aris.2011.1440450112. doi:10.1002/aris.2011.1440450112

Burnham, J. C. (1990). The evolution of editorial peer review. *Jama, 263*(10), 1323-1329. Retrieved from http://dx.doi.org/10.1001/jama.1990.03440100023003. doi:10.1001/jama.1990.03440100023003

Chambers, C. D. (2013). Registered reports: a new publishing initiative at Cortex. *Cortex, 49*(3), 609-610.

Chambers, C. D., Feredoes, E., Muthukumaraswamy, S. D., & Etchells, P. (2014). Instead of" playing the game" it is time to change the rules: Registered Reports at AIMS Neuroscience and beyond. *AIMS Neuroscience, 1*(1), 4-17.

Clark, A. (2017). Results-free review: impressions from the first published articles. Retrieved from http://blogs.biomedcentral.com/bmcseriesblog/2017/06/12/results-free-review-impressions-from-the-first-published-article/

Cronin, B. (2005). *The hand of science: Academic writing and its rewards*: Scarecrow Press.

Csiszar, A. (2016). Peer Review: Troubled from the start. *Nature, 532,* 306-308. Retrieved from http://www.nature.com/news/peer-review-troubled-from-the-start-1.19763

Daston, L., & Galison, P. (2007). *Objectivity*. Cambridge: MIT Press.

Davis, P. (2010). Cascading Peer-Review: The future of open access publishing? Retrieved from https://scholarlykitchen.sspnet.org/2010/10/12/cascading-peer-review-future-of-open-access/

Davis, P. (2017). PLOS ONE Output Drops Again In 2016. Retrieved from https://scholarlykitchen.sspnet.org/2017/01/05/plos-one-output-drops-again-in-2016/

de Oliveira, D. E., Marques, C. S., & Losi, V. C. (2016). "Cell identity" crisis: Another call for immediate action. *Cancer Letters, 381*(1), 122-123. Retrieved from <Go to ISI>://WOS:000383935600014. doi:10.1016/j.canlet.2016.07.031

de Solla Price, D. J. (1963). *Little science, big science*: Columbia University Press New York.

Dickersin, K. (1990). The existence of publication bias and risk factors for its occurrence. *Jama, 263*(10), 1385-1389.

Easterbrook, P. J., Gopalan, R., Berlin, J., & Matthews, D. R. (1991). Publication bias in clinical research. *The Lancet, 337*(8746), 867-872.

Eckberg, D. L. (1991). When nonreliability of reviews indicates solid science. *Behavioral and Brain Sciences, 14*(1), 145-146. Retrieved from https://www.cambridge.org/core/article/when-nonreliability-of-reviews-indicates-solid-science/D92D8E802927B4D135F3E1FFCD6EBCE0. doi:10.1017/S0140525X00065791

Elizondo, A. R., De Rijcke, S., & Van Leeuwen, T. (2017). *Handling publishing misconduct: tools used by publishing houses and editors*. Retrieved from Leiden: http://printeger.eu/wp-content/uploads/2017/09/D3.5.pdf

EMBO Press. (2017). About: Cross Peer Review. *EMBO Press.* Retrieved from http://emboj.embopress.org/about#Cross_peer_review

Epskamp, S., & Nuijten, M. (2014). statcheck: Extract statistics from articles and recompute p values (R package version 1.0. 0.).
Ercegovac, Z., & Richardson, J. V. (2004). Academic dishonesty, plagiarism included, in the digital age: A literature review. *College & Research Libraries, 65*(4), 301-318.
Erren, T. C. (2009). On establishing priority of ideas: Revisiting the "pli cacheté" (deposition of a sealed envelope). *Medical Hypotheses, 72*(1), 8-10. Retrieved from http://www.sciencedirect.com/science/article/pii/S0306987708004192. doi:http://dx.doi.org/10.1016/j.mehy.2008.08.013
Faidhi, J. A., & Robinson, S. K. (1987). An empirical approach for detecting program similarity and plagiarism within a university programming environment. *Computers & Education, 11*(1), 11-19.
Fitzpatrick, K. (2009). The future of peer review. In K. Fitzpatrick (Ed.), *Planned Obsolescene: Publishing, Technology, and the Future of the Academy*. New York: NYU Press.
Fitzpatrick, K. (2010). Peer-to-peer Review and the Future of Scholarly Authority. *Social Epistemology, 24*(3), 161-179. Retrieved from http://dx.doi.org/10.1080/02691728.2010.498929. doi:10.1080/02691728.2010.498929
Franke, R. H., & Kaul, J. D. (1978). The Hawthorne Experiments: First Statistical Interpretation. *American Sociological Review, 43*(5), 623-643. Retrieved from http://www.jstor.org/stable/2094540. doi:10.2307/2094540
Frontiers. (2014). Frontiers is a community-rooted, open-access academic publisher. *About Frontiers.* Retrieved from https://www.frontiersin.org/about/reviewsystem
Funtowicz, S., & Ravetz, J. R. (2001). Peer review and quality control. *International Encyclopaedia of the Social and Behavioural Sciences, Elsevier*, 11179-11183.
Fyfe, A., Coate, K., Curry, S., Lawson, S., Moxham, N., & Rostvik, C. M. (2017). Untangling academic publishing: a history of the relationship between commercial interests, academic prestige and the circulation of research.
Fyfe, A., McDougall-Waters, J., & Moxham, N. (2015). 350 years of scientific periodicals. *Notes and Records: the Royal Society journal of the history of science, 69*(3), 227-239. Retrieved from http://rsnr.royalsocietypublishing.org/content/roynotesrec/69/3/227.full.pdf. doi:10.1098/rsnr.2015.0036
Garfield, E. (1983). More on the ethics of scientific publication: abuses of authorship attribution and citation amnesia undermine the reward system of science. *Essays of an Information Scientist, 5*, 1981-1982.
Garfield, E., & Welljams-Dorof, A. (1990). The impact of fraudulent research on the scientific literature: The Stephen E. Breuning case. *Jama, 263*(10), 1424-1426.
George, S. L. (1985). STATISTICS IN MEDICAL JOURNALS - A SURVEY OF CURRENT POLICIES AND PROPOSALS FOR EDITORS. *Medical and Pediatric Oncology, 13*(2), 109-112. Retrieved from <Go to ISI>://WOS:A1985AFV8400012. doi:10.1002/mpo.2950130215
Godlee, F. (2002). Making reviewers visible: openness, accountability, and credit. *Jama, 287*(21), 2762-2765.
Goodman, S. N. (2017). *Statistical Review in Biomedical Journals: Can We Get SMARTA?* Paper presented at the Eight International Congress on Peer Review and Scientific Publication, Chicago.

Graham, K. (2014). Thanking Our Peer Reviewers. Retrieved from http://blogs.plos.org/everyone/2014/01/06/thanking-peer-reviewers/
Guédon, J.-C. (2001). *In Oldenburg's long shadow: Librarians, research scientists, publishers, and the control of scientific publishing*: Association of Research Libr.
Gunnarsdottir, K. (2005). Scientific journal publications: On the role of electronic preprint exchange in the distribution of scientific literature. *Social Studies of Science, 35*(4), 549-579.
Guston, D. H. (2007). *Between politics and science: Assuring the integrity and productivity of reseach*: Cambridge University Press.
Hames, I. (2014). The changing face of peer review. *Sci Ed, 1*(1), 9-12. Retrieved from https://doi.org/10.6087/kcse.2014.1.9

http://www.escienceediting.org/journal/view.php?number=4. doi:10.6087/kcse.2014.1.9
Hicks, D., Wouters, P., Waltman, L., De Rijcke, S., & Rafols, I. (2015). Bibliometrics: the Leiden Manifesto for research metrics. *Nature, 520*, 429-431.
Hiney, M. (2015). *What it Means, Why it Is Important and How we Might Protect it*. Retrieved from Brussels:
Horbach, S. P. J. M., & Halffman, W. (2017a). The extent and causes of academic text recycling or 'self-plagiarism'. *Research Policy*. Retrieved from http://www.sciencedirect.com/science/article/pii/S0048733317301543. doi:https://doi.org/10.1016/j.respol.2017.09.004
Horbach, S. P. J. M., & Halffman, W. (2017b). Promoting Virtue or Punishing Fraud: Mapping Contrasts in the Language of 'Scientific Integrity'. *Science and Engineering Ethics, 23*(6), 1461-1485. Retrieved from https://doi.org/10.1007/s11948-016-9858-y. doi:10.1007/s11948-016-9858-y
Horner, J., & Minifie, F. D. (2011). Research Ethics II: Mentoring, Collaboration, Peer Review, and Data Management and Ownership. *Journal of Speech Language and Hearing Research, 54*(1), S330-S345. Retrieved from <Go to ISI>://WOS:000286804500003. doi:10.1044/1092-4388(2010/09-0264)
Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Med, 2*(8), e124.
John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling. *Psychological Science, 23*(5), 524-532. Retrieved from <Go to ISI>://WOS:000314464500014. doi:10.1177/0956797611430953
Kennefick, D. (2005). Einstein Versus the Physical Review. *Physics Today, 58*(9), 43-48. Retrieved from http://physicstoday.scitation.org/doi/abs/10.1063/1.2117822. doi:10.1063/1.2117822
Knoepfler, P. (2015). Reviewing post-publication peer review. *Trends in Genetics, 31*(5), 221-223.
Kochan, C. A., & Budd, J. M. (1992). The persistence of fraud in the literature: the Darsee case. *Journal of the American Society for Information Science, 43*(7), 488.
Koshland, D. E. (1987). Fraud in Science. *Science, 235*(4785), 141-141. Retrieved from <Go to ISI>://WOS:A1987F534400001. doi:10.1126/science.3798097

Kovanis, M., Porcher, R., Ravaud, P., & Trinquart, L. (2016). The Global Burden of Journal Peer Review in the Biomedical Literature: Strong Imbalance in the Collective Enterprise. *Plos One, 11*(11), e0166387. doi:10.1371/journal.pone.0166387

Kronick, D. A. (1976). A history of scientific and technical periodicals: the origins and development of the scientific and technological press, 1665-1790.

LaFollette, M. C. (1992). *Stealing into print: fraud, plagiarism, and misconduct in scientific publishing*: Univ of California Press.

LaFollette, M. C. (1994). The Politics of Research Misconduct: Congressional Oversight, Universities, and Science. *The Journal of Higher Education, 65*(3), 261-285. Retrieved from http://www.jstor.org/stable/2943967. doi:10.2307/2943967

Larivière, V., Haustein, S., & Mongeon, P. (2015). The Oligopoly of Academic Publishers in the Digital Era. *Plos One, 10*(6), e0127502. Retrieved from https://doi.org/10.1371/journal.pone.0127502. doi:10.1371/journal.pone.0127502

Larivière, V., Sugimoto, C. R., Macaluso, B., Milojević, S., Cronin, B., & Thelwall, M. (2014). arXiv E-prints and the journal of record: An analysis of roles and relationships. *Journal of the Association for Information Science and Technology, 65*(6), 1157-1169. Retrieved from http://dx.doi.org/10.1002/asi.23044. doi:10.1002/asi.23044

Lee, C. J., Sugimoto, C. R., Zhang, G., & Cronin, B. (2013). Bias in peer review. *Journal of the Association for Information Science and Technology, 64*(1), 2-17.

Lock, S. (1985). *A difficult balance: editorial peer review in medicine*: Nuffield Provincial Hospitals Trust.

Mahoney, M. J. (1985). Open exchange and epistemic progress. *American Psychologist, 40*(1), 29.

Manske, P. R. (1997). A review of peer review. *The Journal of Hand Surgery, 22*(5), 767-771. Retrieved from http://www.sciencedirect.com/science/article/pii/S0363502397800676. doi:http://dx.doi.org/10.1016/S0363-5023(97)80067-6

Martinson, B. C., Anderson, M. S., & de Vries, R. (2005). Scientists behaving badly. *Nature, 435*(7043), 737-738. Retrieved from <Go to ISI>://WOS:000229638700024. doi:10.1038/435737a

Mellor, D. (2016). Registered Reports: Peer review before results are known to align scientific values and practices. Retrieved from https://cos.io/rr/

Moxham, N., & Fyfe, A. (2017). The Royal Society and the prehistory of peer review, 1665–1965. *The Historical Journal*, 1-27. Retrieved from https://www.cambridge.org/core/article/royal-society-and-the-prehistory-of-peer-review-16651965/93B903FD4D6561AA7224C62EE57B0C18. doi:10.1017/S0018246X17000334

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., . . . Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour, 1*, 0021. Retrieved from http://dx.doi.org/10.1038/s41562-016-0021. doi:10.1038/s41562-016-0021

Nosek, B. A., & Lakens, D. (2014). Registered reports: A method to increase the credibility of published results. *Social Psychology, 45*(3), 137-141. doi:10.1027/1864-9335/a000192

Okike, K., Hug, K. T., Kocher, M. S., & Leopold, S. S. (2016). Single-blind vs double-blind peer review in the setting of author prestige. *Jama, 316*(12), 1315-1316. Retrieved from http://dx.doi.org/10.1001/jama.2016.11014. doi:10.1001/jama.2016.11014

Pattinson, D., & Prater, C. (2017). *Assessment of the Prevalence of Integrity Issues in Submitted Manuscripts*. Paper presented at the Eigth International Congress on Peer Review and Scientific Publication, Chicago.

Peerage of Science. (2017). A Free service for scientific peer review and publishing. Retrieved from https://www.peerageofscience.org/

Peters, D. P., & Ceci, S. J. (1982). Peer-review practices of psychological journals: The fate of published articles, submitted again. *Behavioral and Brain Sciences, 5*(2), 187-195. Retrieved from https://www.cambridge.org/core/article/peerreview-practices-of-psychological-journals-the-fate-of-published-articles-submitted-again/AFE650EB49A6B17992493DE5E49E4431. doi:10.1017/S0140525X00011183

PLOS. (2018). Guidelines for Reviewers. Retrieved from http://journals.plos.org/plosone/s/reviewer-guidelines

Pontille, D., & Torny, D. (2014). The Blind Shall See! The Question of Anonymity in Journal Peer Review. *Ada: A Journal of Gender, New Media, and Technology, 4*. doi:doi:10.7264/N3542KVW

Pontille, D., & Torny, D. (2015). From Manuscript Evaluation to Article Valuation: The Changing Technologies of Journal Peer Review. *Human Studies, 38*(1), 57-79. Retrieved from http://dx.doi.org/10.1007/s10746-014-9335-z. doi:10.1007/s10746-014-9335-z

Pöschl, U. (2012). Multi-stage open peer review: scientific evaluation integrating the strengths of traditional peer review with the virtues of transparency and self-regulation. *Frontiers in computational neuroscience, 6*.

Prechelt, L., Graziotin, D., & Fernández, D. M. (2017). On the status and future of peer review in software engineering. *arXiv*.

PubPeer Foundation. (2017). PubPeer. Retrieved from https://pubpeer.com/

Rennie, D. (2003). Editorial peer review: its development and rationale. *Peer review in health sciences, 2*, 1-13.

Research Square. (2017a). Editorial Checks & Badges. Retrieved from https://www.researchsquare.com/publishers/badges

Research Square. (2017b). Rubriq. Retrieved from https://www.rubriq.com/

Robin, E. D., & Burke, C. M. (1987). Peer review in medical journals. *CHEST Journal, 91*(2), 252-257.

Rojas, F. (2007). A modest proposal: triple blind review. *Orgtheory.net.* Retrieved from https://orgtheory.wordpress.com/2007/01/23/a-modest-proposal-triple-blind-review/

Ross-Hellauer, T. (2017). *What is open peer review? A systematic review [version 2; referees: 1 approved, 3 approved with reservations]* (Vol. 6).

Ross, J. S., Gross, C. P., Desai, M. M., & et al. (2006). Effect of blinded peer review on abstract acceptance. *Jama, 295*(14), 1675-1680. Retrieved from http://dx.doi.org/10.1001/jama.295.14.1675. doi:10.1001/jama.295.14.1675

Russel, R. (2008). The business of academic publishing: A strategic analysis of the academic journal publishing industry and its impact on the future of scholarly publishing. *Electron J Acad Spec Librarianship. http://southernlibrarianship. icaap. org/content*.

Sage Open. (2018). More information: Description. Retrieved from https://uk.sagepub.com/en-gb/eur/journal/sage-open#description
Schekman, R., Watt, F., & Weigel, D. (2013). The eLife approach to peer review. *Elife, 2*. Retrieved from <Go to ISI>://WOS:000328614700007. doi:10.7554/eLife.00799
Scheman, R., & Bennett, C. N. (2017). *Assessing the Outcomes of Introducing a Digital Image Quality Control Review Into the Publication Process for Research Articles in Physiology Journals.* Paper presented at the International Congress on Peer Review and Scientific Publication, Chicago.
Schor, S., & Karten, I. (1966). Statistical evaluation of medical journal manuscripts. *Jama, 195*(13), 1123-1128.
SIGMA. (2005). SIGMA is arXiv overlay. Retrieved from http://www.emis.de/journals/SIGMA/about.html#overlay
Sismondo, S. (2017). Post-truth? *Social Studies of Science, 47*(1), 3-6. Retrieved from <Go to ISI>://WOS:000394789900001. doi:10.1177/0306312717692076
Smith, R. (1999). Opening up BMJ peer review. *A beginning that should lead to complete transparency, 318*(7175), 4-5. Retrieved from http://www.bmj.com/content/bmj/318/7175/4.full.pdf. doi:10.1136/bmj.318.7175.4
Smith, R. (2006). Peer review: a flawed process at the heart of science and journals. *Journal of the Royal Society of Medicine, 99*(4), 178-182.
Solomon, D. (2002). Talking past each other: making sense of the debate over electronic publication. *First monday, 7*(8).
Stewart, W. W., & Feder, N. (1987). THE INTEGRITY OF THE SCIENTIFIC LITERATURE. *Nature, 325*(6101), 207-214. Retrieved from <Go to ISI>://WOS:A1987F616900044. doi:10.1038/325207a0
Stroebe, W., Postmes, T., & Spears, R. (2012). Scientific Misconduct and the Myth of Self-Correction in Science. *Perspectives on Psychological Science, 7*(6), 670-688. Retrieved from <Go to ISI>://WOS:000310852500022. doi:10.1177/1745691612460687
Südhof, T. C. (2016). Truth in Science Publishing: A Personal Perspective. *PLOS Biology, 14*(8), e1002547. Retrieved from https://doi.org/10.1371/journal.pbio.1002547. doi:10.1371/journal.pbio.1002547
Tennant, J. P., Dugan, J. M., Graziotin, D., Jacques, D. C., Waldner, F., Mietchen, D., . . . Crick, T. (2017). A multi-disciplinary perspective on emergent and future innovations in peer review. *F1000Research, 6*(1151).
van Assen, M. A. L. M., van Aert, R. C. M., Nuijten, M. B., & Wicherts, J. M. (2014). Why Publishing Everything Is More Effective than Selective Publishing of Statistically Significant Results. *Plos One, 9*(1), e84896. Retrieved from https://doi.org/10.1371/journal.pone.0084896. doi:10.1371/journal.pone.0084896
Walker, R., & Rocha da Silva, P. (2015). Emerging trends in peer review—a survey. *Frontiers in Neuroscience, 9*(169). Retrieved from http://journal.frontiersin.org/article/10.3389/fnins.2015.00169. doi:10.3389/fnins.2015.00169
Ware, M. (2008). Peer review: benefits, perceptions and alternatives. *Publishing Research Consortium, 4*, 1-20.
Weller, A. C. (2001). *Editorial peer review: Its strengths and weaknesses*: Information Today, Inc.

Whelehan, I. (1995). *Modern Feminist Thought: From the Second Wave to\Post-Feminism*: NYU Press.
Yankauer, A. (1990). Editor's report: scientific misconduct and the responsibility of journal editors. *American journal of public health, 80*(4), 399-400.
Zhang, H. Y. (2010). CrossCheck: an effective tool for detecting plagiarism. *Learned publishing, 23*(1), 9-14.
Zuckerman, H. (1984). Norms and deviant behavior in science. *Science, Technology, & Human Values, 9*(1), 7-13.
Zuckerman, H., & Merton, R. K. (1971). Patterns of evaluation in science: Institutionalisation, structure and functions of the referee system. *Minerva, 9*(1), 66-100. Retrieved from http://dx.doi.org/10.1007/BF01553188. doi:10.1007/bf01553188

Chapter 5 - The ability of different peer review procedures to flag problematic publications

This chapter was published as:
Horbach, S. P. J. M., & Halffman, W. (2019). The ability of different peer review procedures to flag problematic publications. Scientometrics, 118(1), 339-373. doi: 10.1007/s11192-018-2969-2

5.1. Introduction

There is a growing concern about erroneous or fraudulent research that gets published in the scientific literature. Mainly originating in the biomedical sciences, scholars have demonstrated that a large proportion of published articles contains flaws (Ioannidis 2005), are not reproducible (Open Science Collaboration 2015), involve questionable research practices, or even outright misconduct (e.g. Horbach and Halffman 2017b; Nuijten et al. 2016). The potential consequences of such problematic publications include sending research on unfruitful avenues, wasting valuable research time and funds, skewing meta-analyses or systematic reviews (Tramer et al. 1997), or building policy recommendations and medical treatments on shaky grounds. Others have expressed worries over the potential reputational damage to science (Drenth 2006).

Retractions are one of research journals' tools to correct the scientific record or redress fraudulent publication credits. The transition to electronic publishing has made it relatively easy to retroactively flag problematic publications. With considerable hesitation over the possible reputational damage for both authors and journals, editors have nonetheless increased the use of this instrument to rectify problematic publications (Cokol et al. 2008; He 2013). As a result, the number of retractions has grown sharply over the last decades, which has led some scholars and journalists to use retractions as a window to study problematic research practices (e.g. Fanelli 2009; Fanelli et al. 2015).

Besides attempts to retroactively redress problematic publications, there have also been several calls and initiatives to try and improve journals' peer review systems to prevent problematic publications in the first place. However, journals' use of peer review to identify fraudulent research is highly contentious. Some argue that peer review was never intended to track fraud and cannot be expected to do so (Biagioli 2002; Smith 2006). Nevertheless, concerns about tracing data manipulation, plagiarism, sloppy statistics, inappropriate referencing, or similar improper behaviour have explicitly motivated several recent peer review innovations (e.g. Scheman and Bennett 2017; Epskamp and Nuijten 2014; Kharasch and Houle 2018; Horbach and Halffman 2018). Such initiatives include the use of various software tools, such as text similarity software or 'plagiarism scanners' (Zhang 2010), but also modifications in peer review procedures, such as the use of checklists or specialised statistics reviewers (Goodman 2017).

These contradictory expectations raise the question to what extent peer review innovations are able to catch problematic research reports before publication and thereby prevent the need for retractions further down the line. In fact, while various actors have been calling for 'evidence-based' improvements of the peer review system, very little is known about the performance of different review models (Rennie 2016). We will address this knowledge-gap in this study. More specifically, in this chapter we investigate whether and how different peer review procedures (e.g. blind, double blind, or 'open') and instruments are related to retraction rates. Using survey data on peer review procedures in a wide range of journals, we relate journal articles to the review procedure they went through. Subsequently, we analyse the relative number of retractions for each review procedure, taking a closer look at the research discipline in which the article was published and, in case of retracted articles, the reason for retraction. Thereby we analyse the effectiveness of different peer review procedures to detect various types of errors and questionable or fraudulent research practices. This leads to informed recommendations for journal editors about strengths and weaknesses of peer review procedures, allowing them to select review procedures that address issues relevant to their field.

In this chapter, we first discuss the contentious expectations for journal peer review and the motivation behind its recent innovations, resulting in a taxonomy of peer review procedures. Second, we discuss retractions and their ambivalent nature as both indicator *of* problematic research and measure *against* problematic research, leading to important caveats about the interpretation of our findings. Third, we describe the methods used, with a survey among editors and the use of the *Retraction Watch* database. Next, we present and discuss our results per peer review procedure, along with a discussion of the motivation behind them and a discussion of the findings. In the final section, we provide an overview of the statistically significant relations and discuss the limitations of our findings, the consequences and recommendations for journal editors, as well as some questions for further research.

5.2. Theoretical framework

5.2.1. Diversity and expectations in peer review

Self-regulating mechanisms are considered an important means of ensuring the quality of the published literature (Stroebe et al. 2012; Hiney 2015). Among them, the peer review system holds a central position (Horner and Minifie 2011). Especially after WWII, peer review of publications gradually came to be seen as the best quality guarantee for the research record, spreading from the natural sciences to other disciplines (Cintas 2016; Baldwin 2017; Fyfe et al. 2017).

Even though the expectation that peer review can detect erroneous research has historically been criticised, it is currently expressed with increasing intensity (LaFollette 1992; Stroebe et al. 2012). Mainly editors and publishers have long asserted that peer review was never designed, nor meant to detect errors or fraud in submitted manuscripts. However, various other actors have increasingly come to expect peer review to help assure a fraud-free published literature. This trend mainly emerged as a response to high subscription costs for journals, leading users to demand better quality assurance, as well as to novel technologies and techniques that promise to help editors and journals to detect errors in research (Fyfe et al. 2017; Larivière et al. 2015).

Peer review procedures are highly diverse, with innovations appearing at an increasing pace (Horbach and Halffman 2018). Whereas the use of external reviewers did not become common practice till well after WWII (Baldwin 2017, 2015), subsequent innovations in review procedures have emerged quickly. These include changes in the relative timing of review in the publication process (Chambers 2013; Nosek and Lakens 2014; Knoepfler 2015), the range and anonymity of actors involved in the review process and the interaction between them (Pontille and Torny 2014; Okike et al. 2016; Godlee 2002), the level of cooperation and specialisation in review (Barroga 2013; Goodman 2017), and the use of digital tools to assist review.

However, very little is known about the effectiveness of various peer review procedures to detect erroneous or fraudulent research. Several studies suggest that peer review is currently under severe threat and falling below standards. Faulty and even fraudulent research slips through peer review at alarming rates (Stroebe et al. 2012; Bohannon 2013; Lee et al. 2013; van der Heyden et al. 2009; Claxton 2005). The fact that only very few of the widely reported misconduct cases were detected through peer review (Stroebe et al. 2012) also raises questions about its fraud detection potential. However, even though peer review in general seems to fail to detect problematic research, little is known about the relative effectiveness of its different procedures.

To assess the effectiveness of various review procedures, we use the taxonomy presented in Table 5.1, based on the peer review inventory in (Horbach and Halffman 2018). The peer review procedures are characterised by twelve key attributes, grouped in four dimensions.

Dimension	Attribute	Range
Selection conditions	Timing	a. No review
		b. Pre-submission (including registered reports)
		c. Pre-publication
		d. Post-publication
	Criteria	a. Methodological rigour and correctness
		b. Anticipated impact (either within or outside of science)
		c. Novelty
		d. Fit with journal's scope
		e. Other
Identities and access	Type of reviewer	a. Editor-in-chief
		b. Editorial committee
		c. External reviewers selected by authors
		d. External reviewers selected by editor(s)
		e. Wider community / readers

		f. Commercial review platforms
	Anonymity of authors	a. Author identities are blinded to editor and reviewer
		b. Author identities are blinded to reviewer but known to editor
		c. Author identities are known to editor and reviewer
	Anonymity of reviewers	a. Anonymous reviewers
		b. Reviewers' identities are open to the authors
		c. Reviewers' identities are open to other reviewers
		d. Reviewers' identities are open to the reader
	Availability of review reports	a. Review reports are accessible to authors and editors
		b. Review reports are accessible to other reviewers
		c. Review reports are accessible to readers of the published manuscript
		d. Review reports are publicly accessible
	Interaction	a. No interaction between authors/reviewers
		b. Interaction amongst reviewers is facilitated
		c. Author's responses to review reports are communicated to the reviewer
		d. Interaction between authors and reviewers is facilitated
Specialisation in review	Structure	a. Unstructured
		b. Semi-structured: Reviewers are guided by some open questions or are presented several criteria for judgement
		c. Structured: Review occurs through mandatory forms or checklists to be filled out by reviewers
	Statistical review	a. Not applicable
		b. No special attention is given to statistical review
		c. Incorporated in review
		d. Performed by additional, specialist reviewer
		e. Performed through automatic computer software
	External sources	a. No reviews from external sources are used
		b. Reviews from other (partner) journals accompanying manuscripts rejected elsewhere are used
		c. Reviews from commercial review platforms are used
		d. Reviews performed by the wider community (i.e. not by invited or targeted reviewers) are used
Technological tools in review	Technical support	a. No digital tools are used
		b. Plagiarism detection software is used
		c. Digital tools to assess validity or consistency of statistics are used
		d. Digital tools to detect image manipulation are used

		e. Digital tools to check references are used
		f. Other Technical support (e.g. machine learning techniques to assess consistency and completeness)
	Reader commentary	a. No reader commentary facilitated
		b. In-channel reader commentary facilitated
		c. Out-of-channel reader commentary facilitated

Table 5.1 Procedures of peer review categorized by dimension and attributes

5.2.2. What are retractions?

Retractions are a measure taken by journal editors to remove publications from the official published scientific record (even though the original text remains available, marked as 'retracted'). A request for retraction may be made by publishers, (co-)authors, research organisations, funders, or any other actor in the research process, but the decision to retract remains an editorial one. Retractions occur for a wide variety of reasons, ranging from honest error to severe cases of research fraud (Hesselmann et al. 2017; Fang et al. 2012).

Retractions are announced through published retraction notices, with a reference to the original publication, and normally also through a warning on the electronic version of the publication. Retraction notices are indexed in databases such as Web of Science or PubMed, but their unstandardized format makes them hard to collect systematically (Hesselmann et al. 2017; Schmidt 2018; Van Leeuwen et al. 2017). In fact, the explanation offered in retraction notices is often obtuse and cryptic. Since 2010, the NGO *Retraction Watch* has been documenting retractions, originally with a journalistic interest in specific misconduct cases. More recently, *Retraction Watch* has developed an online database that provides the most complete overview of retractions currently available, going back to the 20th century (Retraction Watch 2018).

Retractions are a relatively new phenomenon for many research fields. Although the oldest known retraction dates from 1927, retractions occurred only sporadically before 2000 (He 2013). The more traditional response would have been to publish another article, in the form of a rebuttal or challenge to the original publication. The practice became more meaningful as electronic publishing made it possible to add a warning to the original publication, alerting readers accessing the publication online to its problems. Theoretically, this makes it possible to recognise problematic publications without having to know the complete subsequent literature on the matter.

The effects of retractions are complex. Retractions are generally considered a serious reprimand and have detectable negative effects on careers (Azoulay et al. 2017). Some editors may also fear retractions will put a blemish on their journal's reputation, as retractions could be perceived as sign of a failing editorial policy. In contrast, retractions' effect on removal of error or fraud from the scientific record is more modest than expected, as many retracted articles continue to be cited, both by their authors and by others (Van Noorden 2011; Madlock-Brown and Eichmann 2015).

In research on scientific integrity, retractions are sometimes used as indicators of misconduct or questionable research practices (Fanelli et al. 2015; Karabag and Berggren 2016; Hesselmann et al. 2017; Montgomery and Oliver 2017). However, this requires extreme caution. A quarter of retractions do not involve misconduct, but honest error (Fang et al. 2012), sometimes even by journals or publishers. Retractions may signal an offence, but also the social reaction to the offence: an editor has decided (or was pressured) to take the relatively severe measure of retraction to repair the scientific record. Retraction rates are therefore both an indicator of a research community having a problem, but also of this research community taking action to redress this problem. Retractions signal trouble, but also the awareness and resolve to address trouble. Inversely, the absence of retractions does not necessarily signal unproblematic research, but might also indicate an unwillingness to act against it.

In this respect, retractions as indicators of misconduct or problematic research suffer from the same problem as crime statistics: crime rates registered by police or the justice system indicate crime as well as crime fighting (Biderman and Reiss 1967). In criminology, this is known as the 'dark number' problem (Skogan 1977). The issue with dark numbers in estimating misconduct rates have lead scientists, in analogy with criminologists (Van Buggenhout and Christiaens 2016), to adopt various other ways of collecting data on misconduct and errors in research, including (self-reported) misconduct surveys (Martinson et al. 2005), sometimes using incentives for truth-telling (John et al. 2012); or digital tools for detecting problematic research (Horbach and Halffman 2017a), in addition to retraction rates. The latter arguably suffers most from the dark number problem: retraction rates are lower than other misconduct indicators.

A further complication is that measures to trace misconduct may also define or articulate particular behaviour *as* misconduct. Misconduct and scientific error are not just pre-existing, objectively defined phenomena, but may be re-categorised as such by punitive social reaction (Martin 1992). This is an inherently social process in which power structures play a major role (Martin 1992; Callahan 2017). For example, plagiarism scanners have made it possible to systematically trace text recycling, which has raised awareness and also spurred codification of which forms of text recycling are and are not to be considered acceptable (KNAW 2014). Similarly, statistics scanners may also redefine practices that were common and considered normal in particular research fields as problematic or even fraudulent.

These complications imply that retraction rates are ambivalent indicators of integrity problems and have to be interpreted with caution. Nevertheless, we will show that some interpretable and statistically significant relations can be found between some peer review procedures and retractions, including specific effects such as the prevention of plagiarism, or the effect of post-publication reader commentary.

5.3. Methods

5.3.1. Data collection

5.3.1.1. Retracted journal articles

Retracted publications were gathered from the *Retraction Watch* online database (Retraction Watch 2018) on December 11th 2017, at which point it contained 9476 retractions for journal papers (omitting roughly 7500 retracted conference proceedings from the Institute for Electrical and Electronics Engineering [IEEE], which were irrelevant for our focus on journal review). We collected information on the title of the retracted article, the Digital Object Identifier (doi), the Pubmed ID, and the reason for retraction. Using the doi and Pubmed ID, the records were subsequently matched with the Web of Science (WoS) database. This allowed us to gather additional data on the retracted articles available in WoS, such as date of publication, publishing journal, research discipline, and citations to the retracted article. This yielded a list of 7861 retractions. Figure 5.1 displays an overview of the selected retractions.

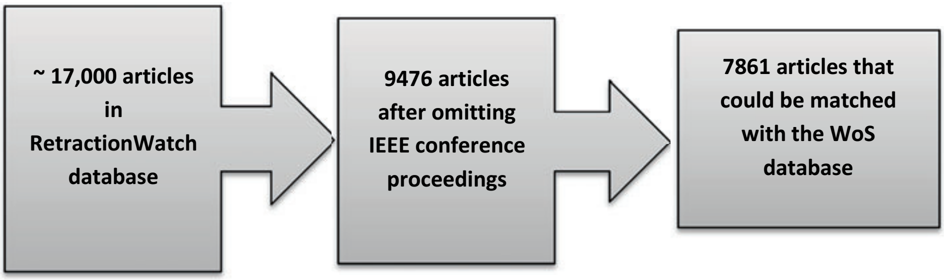


Figure 5.1 Overview of the retractions data. Starting with the RetractionWatch database, we omitted a large number of retracted conference proceedings by IEEE. The final set consists of the resulting retracted articles that are contained in the WoS database.

5.3.1.2. Peer review procedure questionnaire

Unfortunately, journal web pages present surprisingly incomplete information about their peer review procedures, even for procedures currently in use. This information had to be gathered through a short questionnaire among journal editors. The questionnaire consisted of 12 questions, each on a specific attribute of the journal's peer review (Table 5.). The questionnaire can be found in the supplementary material. In addition, respondents were asked to indicate whether, when and how any of these attributes was modified since 2000. Hence the questionnaire allowed us to identify precisely which peer review procedure was used by a specific journal in any specific year since 2000. Subsequently, this allowed us to match every published paper to the specific peer review procedure it went through. Although very simple, we nevertheless pre-tested the questionnaire with two editors to avoid ambiguities, leading to minor modifications.

5.3.1.3. Mailing

Journal editors were contacted via email. Email addresses were gathered in two ways. First, we sampled all 2017 articles indexed in WoS as 'editorial material' and extracted the email address of the corresponding author, on the assumption that authors of 'editorial material' would very likely be editors. This yielded a list of 58763 unique email addresses, covering a total of 6245 different journals. We subsequently collected one random email address for each of these journals. Second, because of our specific interest in journals with a substantive amount of retractions, we also manually collected editorial email addresses of journals with at least 10 retractions, in case they had been omitted via our first sampling strategy. Combined, this yielded editorial email addresses for a total of 6331 journals.

The questionnaire was initially sent on February 23rd 2018 and reminders were sent on March 12th 2018 and March 19th 2018. For the second reminder, we used alternative editorial email addresses from these journals, if available in our database. We received 326 automatic response messages of emails that could, due to various reasons, not be delivered. In addition, 113 people responded that they were not able to fill in the questionnaire, for example because they were not or no longer an editor. Hence, a total of 5892 (=6331-326-113) journals were effectively reached. After sending out the questionnaire, several respondents offered to further disseminate the questionnaire among their networks, which we gratefully accepted. Hence the questionnaire was subsequently distributed among the European Association of Science Editors (EASE), the International Society of Managing and Technical Editors (ISMTE) and through the newsletter of the Committee on Publication Ethics (COPE). We stopped collecting responses on April 5th 2018.

After the reminders, we eventually obtained a total of 361 useful responses. The final response rate of 6.12% is low, but comparable to, or even higher, than response rates of similar online surveys among journal editors or authors regarding issues related to academic integrity (Hopp and Hoover 2017; Stitzel et al. 2018). Nevertheless, our sample covers a wide range of research fields and reflects the distribution of journals over research fields.

5.3.1.4. Data analysis

Using the database of retracted articles and the Web of Science database of published articles, we identified the number of published and the number of retracted articles per journal per year. We limited the analysis to publications indexed in WoS as research articles, rather than, for instance, editorials or book reviews. In total, the journals responding to our survey published 833172 articles since 2000, of which 670 were retracted. This constitutes the eventual sample. Each article serves as a record in our dataset, thereby taking individual articles, rather than journals, as our unit of analysis. Taking journals as unit of analysis would create considerable complications. First, this would imply using retraction rates of single journals as our measure, but the numbers of publications vary substantially between journals. Hence retraction rates are of unequal accuracy. Second, as journals changed their review procedures since 2000, such as with the introduction of plagiarism scanners, a journal is not a constant unit of analysis. Because we asked editors when and how their journal's peer review procedures changed, we know what procedure articles went through based on their publication date, assuming editors report procedures accurately and procedures are applied consistently. As this may not always be the case, we acknowledge that we actually test the relation between peer review *procedures* and retractions, rather than peer review *practices* and retractions: the actual review may differ from the formal procedure.

We attributed to each record: the peer review procedure that the article went through, for each of the 12 attributes identified in the survey; the research area that the article belongs to (Social Science & Humanities, Biomedical & Health Science, Physical Sciences & Engineering, Life & Earth Science, or Mathematics & Computer Sciences), based on the classification in research areas of the Leiden Ranking (Waltman and van Eck 2012; CWTS 2018); and, in case the article was retracted, the reason for retraction, based on the data from *Retraction Watch*. Table 5. presents an overview of the article distribution over research areas. The retractions in our sample reflect the distribution of articles and retractions over fields. The high number of biomedical retractions corresponds to the more frequent use of retractions in this field. Note that some journals responding to our survey are not indexed by the Leiden Ranking and hence are not classified in a research field, causing the numbers per research area to add up to (slightly) less than the total number of articles or retractions.

Research area	Social Science & Humanities	Biomedical & Health Sciences	Physical Sciences & Engineering	Life & Earth Sciences	Mathematics & Computer Sciences	Total
Total articles published	1660394	6325415	5717615	2948177	1605903	18257504
Articles in our sample	53922	382950	196845	160126	38225	833172
% of total articles in our sample	3.25%	6.05%	3.44%	5.43%	2.38%	4.56%
Total retractions	380	4394	1693	800	281	7547
Retractions in our sample	30	368	183	67	21	670
% of retracted articles in our sample	7.90%	8.38%	10.81%	8.38%	7.48%	8.86%

Table 5.2 The number of articles and retracted articles since 2000 in our sample, according to research area, as defined in the Leiden Ranking.

Table 5.2 shows that our sample contains a relatively high share of retractions. Indeed, relatively many journals with a high number of retractions have responded to our survey. The same observation can be made from figure 5.2 showing the number of journals with a specific number of retractions. It shows that our sample contains a good representation of both journals with large and small numbers of retractions, but with a slight over-representation of journals with large numbers of retractions. For example, Figure 5.2 demonstrates that while in general 90% of all journals has fewer than 8 retractions, in our sample this is only 70% of all journals. In contrast, while in the general population of journals the top 5% of journals with most retractions has at least 12 retractions, the top 5% in our sample only starts at 24 retractions.

We suspect that editors with previous experience with retractions were more likely to answer the survey, but we also note that our sampling strategy included more journals with a relatively high number of retractions among the recipients of our survey.

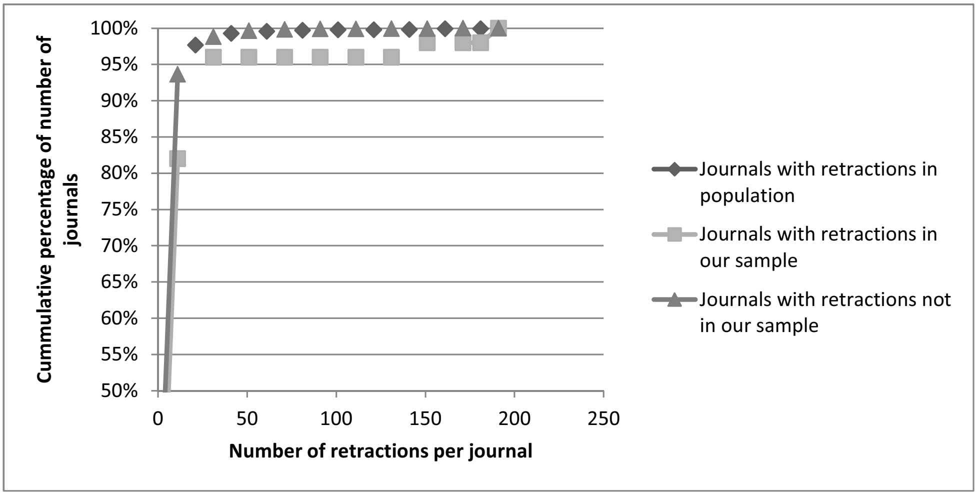


Figure 5.2 Cumulative distribution of retractions over journals, either in the entire population, in our sample, or outside of our sample.

The data from *Retraction Watch* was classified in nine categories of reasons for retraction: Plagiarism/Duplication of text, Falsification/Fabrication, Ethical concerns, Authorship issues, Issues with References, Image manipulation/Data issues, (Honest) Errors, Fake Review, and Misconduct (grouped). Misconduct (grouped) is a category encompassing all other categories except (honest) error and further includes the general labels referring to misconduct as given by *Retraction Watch*. A complete classification of the reasons for retraction as presented by *Retraction Watch* into the categories specified above, can be found in the supplementary material. Note that a single retraction usually occurs due to various reasons, classifying individual retractions in multiple categories. In addition we note that *Retraction Watch* also uses the 'error' category in cases where other reasons for retraction, such as misconduct, cannot be proven (even though strong suggestion to the contrary might exist).

For each of the twelve characteristics of peer review we studied, we analysed whether the review procedure had a (significant) association with the chance of a paper being retracted. Significance was tested using a log likelihood ratio-test on the ratios of retracted and non-retracted papers. In addition, subgroup-specific effects for the five research areas were tested using binary logistic regression as interaction test between the peer review procedure and the research discipline. Here we use the Wald-statistic on the logistic model with the interaction term, compared to the model without the interaction term. Last, specific effects for the various retraction reasons were analysed, using an ANOVA test to analyse differences in mean retraction numbers, and multinomial logistic regression to analyse the effect size of various peer review procedures on retraction rates. Analyses were performed using the SPSS 25.0 statistics software.

5.4. Results

This section presents the results of our analyses, grouped by the various attributes of peer review presented in Table 5.. For each of the twelve attributes included in our survey, we briefly outline some of the differences between the various procedures, the original rationale behind their development and some effects on rates of problematic publications or retractions expected in the literature.[1] We should note that many of these procedures may be combined in the review process of journals, for instance using both registered reports as well as pre-publication review, or involving multiple actors in the review process. Subsequently, we present the results of our analyses according to whether and how these peer review procedures are associated with difference in the rate of retraction, whether research area is a (significantly) mediating factor and whether differences in retraction rate for different reasons of retraction are visible. We conclude with a short discussion of the results.

5.4.1. Timing

Traditionally, peer review occurs between the submission and publication of a manuscript. However, over the past decades, new peer review procedures have been proposed for different phases of the publication process. Most notably these include: pre-submission review (e.g. through registered reports) (Chambers 2013; Nosek and Lakens 2014; Mellor 2016), in which articles are reviewed prior to data collection based on their rationale, research question and proposed method; and post-publication review (Knoepfler 2015; Pöschl 2012), in which articles are reviewed only after publication, potentially involving a wider community rather than merely invited reviewers. The latter procedure was mainly introduced in order to speed up publication and enhance fast knowledge exchange, whereas the pre-submission procedure was primarily introduced to foster publication of negative or null-results and deter researchers from hunting for spectacular outcomes (Chambers et al. 2014; Nosek and Lakens 2014).

Our results suggest that the pre-submission system is indeed related to fewer retracted articles (Table 5.2): in total, 7.6% of all articles went through pre-submission review, whereas only 4.8% of retractions went through this review procedure ($\Lambda(3) = 18.899$, $p<0.001$). Due to the ambivalent nature of retractions as both indicator of undetected errors or of the willingness to repair errors, this could mean that these journals are less prone to take action after publication. However, since this system is used with the explicit intention to prevent tweaking of data or

[1] We realise this mixes theoretical expectation, data, and interpretation, but found it easier to describe what is interesting about our results than if separated in different sections.

statistics, it seems highly unlikely that this lower retraction rate is due to lax editorial attitudes towards problematic research.

The rates for traditional pre-publication review (97.6% vs. 96.3%) and post-publication review (0.8% vs. 0.0%) did not show significant differences. However, the fact that no retractions were reported in journals using post-publication review is interesting. It might suggest that potential issues are dealt with in review and commentaries, rather than using retractions as a mechanism to correct the literature, but the number of publications reviewed in this way is still relatively low. No significant interactions were found with respect to research area (WALD = 5.445, df = 5, p = 0.364), nor reasons for retraction (F(1,1266) = 5.409, p = 0.020).

At what stage of the publication process does review take place?	Retracted	95% CI	Non-retracted	95% CI
No review takes place	0 (0.0%)	0.0% - 0.0%	648 (0.1%)	0.1% -0.1%
Pre-submission review (including registered reports)	32 (4.8%)	3.2% - 6.4%	63262 (7.6%)	7.5% - 7.7%
Pre-publication review	645 (96.3%)	94.8% - 97.7%	812362 (97.6%)	97.5% - 97.6%
Post-publication review	0 (0.0%)	0.0% - 0.0%	7008 (0.8%)	0.8% - 0.9%

Table 5.2 Timing of peer review relative to the publication process related to number of non-retracted and retracted articles in our sample.

5.4.2. Review criteria

Journals use a variety of review criteria. Commonly, methodological rigour and correctness, conceptual soundness, and fit with the journal's scope are used as assessment criteria. However, also scientific novelty and anticipated impact (either within or outside of science) are used to assess manuscripts. Some journals have deliberately decided *not* to take factors like novelty nor anticipated impact into account when judging manuscripts (BMJ Open 2018; PLOS 2018; Sage Open 2018). Their rationale for doing so is to allow all valid research (i.e. methodologically and conceptually sound research) to be published, irrespective of whether results are positive or negative, and irrespective of novelty or impact. Thereby they facilitate the publication of replication studies and do not incentivise authors to obtain spectacular, new or (significantly) positive results. This arguably takes away incentives for questionable research practices and may hence foster research integrity.

What quality criteria does your journal use for peer review?	Retracted	95% CI	Non-retracted	95% CI
Methodological rigour and correctness	666 (99.4%)	99.1% - 99.7%	810932 (97.4%)	97.4% - 97.4%
Anticipated impact (either within or outside of science)	549 (81.9%)	80.4% - 83.5%	523629 (62.9%)	62.8% - 63.0%
Novelty	610 (91.0%)	89.9% - 92.2%	697846 (83.8%)	83.8% - 83.9%
Fit with journal's scope	641 (95.7%)	94.9% - 96.5%	733670 (88.1%)	88.1% - 88.2%

Table 5.3 Review criteria related to number of non- retracted and retracted articles in our sample.

The results of our analysis (Table 5.3) suggest that journals taking novelty and anticipated impact into account when assessing manuscripts are indeed associated with more retractions. The criteria used for assessing articles demonstrate a significant impact on the number of retractions ($\Lambda(3) = 18.779$, $p < 0.001$), with significantly more retractions for journals using novelty and anticipated impact as assessment criteria. No significant interactions were observed regarding research area (WALD = 16.171, df = 12, $p = 0.161$), nor reason for retraction ($F(3,4665) = 1.220$, $p = 0.301$), suggesting that the effect is homogeneous with respect to research discipline and type of problematic research.

The higher retraction levels among journals aiming to publish highly relevant and novel research, usually journals with high impact factors, has also been established in previous research on retractions (Steen 2011; He 2013). As such, focussing on high-impact and novel research might be a deliberate high-risk/high-gain strategy for journals, potentially leading to high impact factors and citation scores, but also to a higher risk of having to retract articles. Here too, the lower retraction rate seems more plausibly associated with prevention of problematic publications, rather than with the willingness to rectify it. In fact, journals that use anticipated impact as a selection criterion have a significantly higher rejection rate (70 % vs 63%, $t(280) = -3.043$, $p = 0.016$). Apparently, they have 'more to choose from' than journals that do not use impact as criterion, and/or have a tighter limit on the number of articles they can publish (e.g. printed versus exclusively electronic journals). However, the higher retraction rates suggest that these journals either attract more problematic submissions or are less capable of filtering them.

In addition, it could be expected that the strategy to select articles with the highest anticipated impact would be expressed in a higher journal impact factor (JIF). However, the journals that use anticipated impact as a selection criterion, on average, do not have a higher journal impact factor. On the contrary, journals in our sample using impact as a selection criterion have a slightly *lower* 2016 JIF than those which do not (2,51 vs. 2,86). The precise relation between impact as a selection criterion, JIF, and retraction rates would have to be analysed in a larger, multivariate analysis, but our findings suggest the impact criterion provokes more retractions *and* fails to increase the JIF rating.

5.4.3. Type of reviewers

The use of external reviewers, i.e. researchers not directly affiliated with the journal, did not become standard practice until well after WWII (Baldwin 2017). Still today, the actors performing reviews range from the editor-in-chief, editorial committee members, external reviewers (either suggested by authors or merely selected by editors), to the wider community (usually in post-publication review), or even independent commercial review platforms (Research Square 2017; Tennant et al. 2017). The latter have recently emerged as organisations to which authors may submit their manuscript for review, after which the manuscript together with review reports (or certain assigned 'badges') are sent to a suitable journal (PubPeer Foundation 2017; Research Square 2017). This has mainly been introduced to prevent manuscripts from going through several rounds of review after rejection at an initial journal, thereby decreasing the burden on the peer review system.

Type of reviewers	Retracted	95% CI	Non-retracted	95% CI
Editor-in-chief	197 (29.4%)	27.2% - 31.6%	347035 (41.7%)	41.6% - 41.8%
Editorial committee	376 (56.1%)	53.7% - 58.5%	439754 (52.8%)	52.8% - 52.9%
External reviewers suggested by authors	372 (55.5%)	53.1% - 57.9%	445636 (53.5%)	53.5% - 53.6%
External reviewers suggested and selected by editor(s)	653 (97.5%)	96.7% - 98.2%	805787 (96.8%)	96.8% - 96.8%
Wider community / readers	34 (5.1%)	4.0% - 6.1%	146502 (17.6%)	17.5% - 17.6%
Commercial review platforms	0 (0.0%)	0.0% - 0.0%	9192 (1.1%)	1.1% - 1.1%

Table 5.4 Identity of reviewer related to number of non- retracted and retracted articles in our sample.

Our analysis shows a significant impact of the actor type performing the review ($\Lambda(5) = 116.527$, $p < 0.0001$), with relatively few retractions occurring when editors-in-chief or the wider community are involved in review (Table 5.4). In addition, a significant difference was found regarding the reason for retraction ($F(4,2782) = 10.538$, $p < 0.001$): when the editor-in-chief, the editorial committee or author-suggested reviewers are involved, relatively few retractions appear for fake review reports or issues with references, while relatively more retractions occur for authorship or ethical issues. Finding relatively few retractions for fake peer review when author-suggested reviewers are used, is somewhat puzzling, as these types of actors seem most vulnerable to fraud with review reports. More research will be needed to elucidate the mechanism underlying this association.

The finding that involvement of the wider community is related to fewer retractions is in line with expectations expressed in the literature, which suggest that wider involvement would lead to higher levels of scrutiny and hence a higher quality review, as well as a stronger deterring effect diverting fraudulent papers away from these journals (Harris et al. 2015). Our result that involvement of the editor-in-chief would lead to higher quality review also raises some further questions. Future research could look at this in more detail, for instance specifying the role of the editor-in-chief in the review process or distinguishing between editors for whom editorial work is their main occupation and those doing it more or less voluntarily in their free time. In any case, again, involvement of such actors seems unlikely to be related to poor willingness to address problematic research. Hence, in this case also, low retraction rates are more likely explained by more effectiveness to detect such research in an early phase.

5.4.4. Author anonymity

In the early days of peer review, editors and reviewers were (nearly) always aware of authors' identities, whereas authors knew the identity of the editor-in-chief, but not necessarily of the editorial committee or invited outside reviewers (single-blind review). Responding to issues of equality and fairness (Zuckerman and Merton 1971; Peters and Ceci 1982), the systems of double-blind and triple-blind review were introduced, in which author identities were blinded to reviewers and editors respectively (Pontille and Torny 2014). The ambition of these innovations was to judge manuscripts on content rather than extraneous factors such as authors' gender, affiliation or nationality.

We analysed the impact of blinding author identities to editors and/or reviewers (Table 5.5Table 5.). The results demonstrate a significantly lower rate of retractions in case author identities are blinded to the reviewer ($\Lambda(2) = 106.042$, $p < 0.0001$). The effect can be witnessed in all research areas, but is especially strong in the social sciences and humanities. In this research area, 79% of all articles went through double-blind review, whereas only 13% of all retracted articles went through this review procedure. In contrast, only 19% of the articles were reviewed in a procedure allowing reviewers to see authors' identities, whereas 87% of all retractions went through such review. The figures for biomedical and health sciences show a similar, but weaker relation (83% of articles did not have author identities blinded during review, but 95% of retractions occur in this procedure). For the other research areas similar trends were found, but no significant differences occurred. In addition, significant differences occurred when comparing the various reasons for retraction (F(1,1260) = 10.630, p = 0.001), with the strongest effects for the category of retractions due to fake review, ethical violations, and misconduct.

Level of author anonymity		Author identities are known to editor and reviewer	Author identities are blinded to reviewer but known to editor	Author identities are blinded to editor and reviewer
All articles (percentage of all articles)		692920 (83.2%)	135019 (16.2%)	7011 (0.8%)
Non-retracted (percentage of non-retracted papers)	Total	692280 (83.2%)	134989 (16.2%)	7011 (0.8%)
	Social Science & Humanities	10206 (18.9%)	42819 (79.5%)	1192 (2.2%)
	Biomedical & Health sciences	317182 (82.9%)	65329 (17.1%)	1327 (0.3%)
	Physical sciences & Engineering	181201 (92.1%)	12000 (6.1%)	3461 (1.8%)
	Life & Earth sciences	146040 (91.2%)	13186 (8.2%)	1032 (0.6%)
	Mathematics & computer sciences	36849 (96.5%)	1355 (3.5%)	0 (0.0%)
Retracted (percentage of retracted papers)	Total	640 (95.5%)	30 (4.5%)	0 (0.0%)
	Social Science & Humanities	26 (86.7%)	4 (13.3%)	0 (0.0%)
	Biomedical & Health sciences	348 (94.6%)	20 (5.4%)	0 (0.0%)
	Physical sciences & Engineering	180 (98.4%)	3 (1.6%)	0 (0.0%)
	Life & Earth sciences	65 (97.0%)	2 (3.0%)	0 (0.0%)
	Mathematics & computer sciences	20 (95.4%)	1 (4.8%)	0 (0.0%)

Table 5.5 Level of author anonymity during peer review related to the number of non- retracted and retracted articles per research area.

Studies in psychology and economics have previously suggested that people are more strict when reviewing or judging the unknown rather than the known or the familiar (Cao et al. 2009). Our results suggest the same to be true in academic peer review. In addition, one could argue that, especially in Social Sciences & Humanities, adopting a single-blind review format is a sign of innovation and commitment to act on problematic research. Hence the higher retraction rates might here indicate a higher willingness to address issues, rather than a poorer capability to detect them.

However, what specific mechanism accounts for the difference in retraction rate between single- and double-blind reviewed papers remains to be studied. This is especially so regarding the current discussion about the effectiveness of blinding in the digital age, in which authors are easily identified with a simple Google-search.

5.4.5. Reviewer anonymity

Similar to the anonymity of the author, some discussions regarding peer review procedures have centred around the anonymity of the reviewer (Amsen 2014; Ross-Hellauer 2017). Contrary to the system of double- or triple-blind review, open review has been proposed as a way to tackle reviewer bias by rendering the review process more transparent (Smith 1999; Godlee 2002). The expectation is that by disclosing the identity of the reviewer to either the authors of the submitted manuscript, other reviewers of the same manuscript, the readers of the published manuscript, or even the general public, reviewers are held accountable for their choices while they do receive credit for their work. The combination of both incentives is argued to facilitate more rigorous review, thereby augmenting the likelihood of detecting erroneous or fraudulent research (Walker and Rocha da Silva 2015; Ross-Hellauer 2017).

Our data (Table 5.6) does not seem to uphold the claim that known reviewer identities increase the likelihood of retracted papers ($\Lambda(3) = 5.964$, $p = 0.0494$). Neither do we find significant differences when correcting for research fields (WALD = 15.717, df = 7, $p = 0.028$) nor reasons for retraction ($F(3,1262) = 2.839$, $p = 0.784$). This might mainly be due to the fact that an overwhelming majority of the articles, as well as the retractions, goes through the same review procedure: a system in which reviewer identities are blinded to all relevant actors. Hence, to properly study the influence of this review attribute other research strategies such as randomised trails or other intervention studies could be employed.

Extent to which reviewers are anonymised	Retracted	95% CI	Non-retracted	95% CI
Reviewers are anonymous (both to authors and other reviewers as well as to readers of the published manuscript)	654 (97.6%)	96.5% - 98.8%	816784 (98.1%)	98.1% - 98.1%
Reviewer identities are known to other reviewers of the same manuscript	6 (0.9%)	0.2% - 1.6%	4004 (0.5%)	0.5% - 0.5%
Reviewer identities are known to the authors	9 (1.3%)	0.5% - 2.2%	5412 (0.7%)	0.6% - 0.7%
Reviewer identities are known to the readers of the published manuscript	5 (0.7%)	0.1% - 1.4%	4782 (0.6%)	0.6% - 0.6%

Table 5.6 Level of reviewer anonymity related to number of non- retracted and retracted articles in our sample.

5.4.6. Review reports

In addition to disclosing reviewer identities, open review frameworks have proposed to also make the review reports accessible. We distinguish four levels of accessibility: review reports accessible (i) to authors and editors, (ii) to other reviewers of the same manuscript, (iii) to readers of the published manuscript, and (iv) to the wider public, i.e. without restrictions (Walker and Rocha da Silva 2015; Ross-Hellauer 2017). Making review reports widely accessible has been proposed with the same rationale as disclosing reviewer identities: it provides a transparent and hence supposedly more thorough review process.

In our data (Table 5.7) we found no significant influence of the accessibility of review reports on the number of retractions ($\Lambda(3) = 9.081$, $p = 0.0128$). However, we did find some specific influences when regarding research area (WALD = 47.551, df = 5, $p < 0.0001$) and the reason for retraction ($F(3,1821) = 6.897$, $p < 0.001$). Making review reports accessible not only to authors and editors, but also to other reviewers of the same manuscript was associated with fewer retraction due to fake reviews and issues with references, while in this case we see an increase in the rate of retractions due to plagiarism, falsification, image and/or data issues and ethical violations. The fact that no significant effects were measured for the other two review procedures, those in which reports are shared with the manuscript's readers or the wider public, might again be due to the low number of articles and retractions going through these review procedures. Again, other research set-ups could be employed to study the effect of making review reports more or less widely accessible on the quality of review.

Accessibility of review reports	**Retracted**	**95% CI**	**Non-retracted**	**95% CI**
Review reports are accessible to authors and editors	619 (92.4%)	90.8% - 94.0%	811087 (97.4%)	97.4% - 97.5%
Review reports are accessible to other reviewers	448 (66.9%)	64.1% - 69.7%	487933 (58.6%)	58.5% - 58.7%
Review reports are accessible to readers of the published manuscript	5 (0.7%)	0.2% - 1.3%	4790 (0.6%)	0.6% - 0.6%
Review reports are publicly accessible	2 (0.3%)	0.0% - 0.6%	2108 (0.3%)	0.2% - 0.3%

Table 5.7 Accessibility of review reports related to number of non- retracted and retracted articles in our sample.

5.4.7. Interaction between actors

Besides sharing review reports or disclosing identities, some journals have introduced review procedures in which interaction between various actors in the review process is facilitated. This includes modest levels of interaction by allowing reviewers to read author responses to their review report, but also goes further by facilitating interaction between reviewers of the same manuscript (Schekman et al. 2013; EMBO Press 2017), or even facilitating direct communication between authors and reviewers of a manuscript (on top of formal communication by means of review reports and responses to them) (Amsen 2014; Frontiers 2014). Again, a quest for transparency and accountability in review were the main motivators for introducing these review procedures. In addition, they are claimed to improve the quality of reviews by allowing actors to discuss and respond efficiently to reviewers' questions or comments.

Interaction between authors and reviewers	Retracted	95% CI	Non-retracted	95% CI
No interaction between authors or reviewers is facilitated	130 (19.4%)	16.6% - 22.2%	339782 (40.8%)	40.7% - 40.9%
Author's responses to review reports are communicated to the reviewer	562 (83.9%)	81.3% - 86.5%	573308 (68.9%)	68.8% - 69.0%
Interaction between reviewers is facilitated	48 (9.0%)	5.3% - 9.0%	75100 (9.0%)	9.0% - 9.1%
Interaction between authors and reviewers is facilitated (on top of formal review reports and formal responses to review reports)	18 (2.7%)	1.5% - 3.8%	52088 (6.3%)	6.2% - 6.3%

Table 5.8 Level of interaction between authors and reviewers related to number of non- retracted and retracted articles in our sample.

The data from our study (Table 5.8) actually rather suggest that the opposite is true, finding significantly fewer retractions when no interaction between authors and reviewers is facilitated and relatively more retractions when authors are allowed to respond to review reports ($\Lambda(3) = 126.4$, $p < 0.0001$). More specifically, allowing no interaction reduces the likelihood of retractions for fake review, ethical issues and misconduct in general. Contrarily, allowing authors to respond to review reports increases the likelihood of retractions occurring for fake review, ethical concerns or issues with references ($F(3,1405) = 21.269$, $p < 0.001$). Research area was also found to be a significantly mediating factor (WALD = 85.710, df = 12, $p < 0.0001$) with stronger effects in the biomedical and health sciences as well as the physical sciences and engineering.

In particular, it might be deemed surprising that interaction between reviewers is not associated with lower retraction rates, as more interaction is expected to lead to higher scrutiny during review and hence to fewer retractions. Indeed, in other settings, such as detecting medication errors, it has been suggested that higher levels of cooperation and interaction would be beneficial for effective error detection (Kaushal et al. 2001). Similar relations might be expected from editorial peer review. The specific effect (or lack thereof) of interaction and communication between reviewers is open to future research.

5.4.8. Checklists: level of structure in review criteria

Another salient difference distinguishing review procedures is the level of structure that editors require from their reviewers. We distinguish three levels of structure: structured, when reviewers are asked to fill out a form or checklist listing specific (closed) questions or to rate specific aspects of the manuscript; semi-structured, when reviewers are presented a list of guiding questions or criteria that might assist them in writing their review; and unstructured, when reviewers receive a manuscript for review without further guidance about review criteria.

Our data suggests (Table 5.) that the level of structure plays a significant role in the relative number of retractions appearing after peer review ($\Lambda(2) = 58.907$, $p < 0.0001$), with fewer retractions appearing in either structured and unstructured review, but more retractions appearing after semi-structured review. Specifically, semi-structured review is related to significantly more retractions for fake review, authorship and ethical issues and concerns over references ($F(2,1382) = 12.538$, $p < 0.001$). In addition, subject area turned out to be a significant mediating factor (WALD = 145.578, df = 8, $p < 0.0001$), with particularly strong effects in Social Science & Humanities and Mathematics & Computer Science, and relatively weak effects in Life & Earth Sciences.

Level of structure		**Unstructured**	**Semi-structured**	**Structured**
All articles (percentage of all articles)		126282 (15.2%)	627559 (75.3%)	158274 (19.0%)
Non-retracted (percentage of non-retracted papers)	Total	126233 (15.2%)	626971 (75.3%)	158184 (19.0%)
	Social Science & Humanities	17275 (32.1%)	33574 (62.3%)	14652 (27.2%)
	Biomedical & Health Sciences	45414 (11.9%)	270776 (70.8%)	82097 (21.5%)
	Physical Sciences & Engineering	19946 (10.1%)	171251 (87.1%)	51172 (26.0%)
	Life & Earth Sciences	35010 (21.9%)	122726 (76.7%)	7783 (4.9%)
	Mathematics & Computer Sciences	8290 (21.7%)	27910 (73.1%)	2113 (5.5%)
Retracted (percentage of retracted papers)	Total	49 (7.3%)	588 (87.8%)	90 (13.4%)
	Social Science & Humanities	1 (3.3%)	26 (86.7%)	3 (10.0%)
	Biomedical & Health sciences	21 (5.7%)	303 (82.3%)	47 (12.8%)
	Physical sciences & Engineering	1 (0.5%)	182 (99.5%)	10 (5.5%)
	Life & Earth sciences	26 (38.8%)	56 (83.6%)	29 (43.3%)
	Mathematics & computer sciences	0 (0.0%)	21 (100.0%)	0 (0.0%)

Table 5.9 Level of structure in review related to the number of non- retracted and retracted articles per research area.

Interestingly, both extremes of the spectrum appear related to fewest retractions. This suggests that either guiding reviewers very specifically through the review process or leaving them to decide on appropriate ways of reviewing themselves is most effective in detecting problematic publications. Alternatively, partly guiding reviewers seems to be least effective. We could speculate that reviewers in this case would only consider those aspects referred to in their checklist, while editors might expect them to take more aspects of the manuscript into account. However, other mechanisms might also be at play. To obtain a better understanding of this phenomenon, future research could compare specific guidelines for reviewers with the retraction rates on a more qualitative level.

Since especially highly structured review procedures were introduced expressly with the intent to address problematic research, it seems improbable that lower retraction rates are to be seen as an indication of unwillingness to address problematic research. However, low retraction rates for the other side of the spectrum, i.e. unstructured review criteria, are harder to interpret in this way.

5.4.9. Statistics review

Statistical analyses are increasingly recognised as a source of error, questionable research practices, or outright fraud in quantitative scientific papers (Altman 1998; Goodman 2017; Carlisle 2017). Hence, statistics has come under close scrutiny in some journals' review process. This led several journals to assign specialist statistical reviewers to their review pool already in the 1980s (George 1985). In addition, more recently, several digital tools were developed to assist in the review of statistical analyses (Bakker and Wicherts 2011; Nuijten et al. 2016). These all aim to increase the detection likelihood of statistical errors and misrepresentations.

Our data indeed show (Table 5.10Table 5.5) a significant influence of how statistics is included in the review process ($\Lambda(4) = 138.858$, $p < 0.0001$). However, the results do not provide evidence for the effectiveness of assigning specialist statics reviewers or employing digital tools to assist in statistical review. Specifically, we witness more retractions appearing in journals that state that statistics is not relevant for their journal, while less retractions appear in journals either paying 'no special attention' to review, incorporate review in the standard tasks of reviewers, or use specialist statistical reviewers. A significant difference between research areas was witnessed (WALD = 164.869, df = 13, $p < 0.0001$), suggesting stronger effects in the physical sciences and engineering as well as the life and earth sciences.

Level and type of statistical review	Retracted	95% CI	Non-retracted	95% CI
Not applicable (statistics does not play a role in my journal's research area)	193 (28.8%)	25.4% - 32.2%	117760 (14.1%)	14.1% - 14.2%
No special attention is given to statistical review	70 (10.4%)	8.2% - 12.7%	165637 (19.9%)	19.8% - 20.0%
Incorporated in review (assessing statistics is part of reviewer's and editor's tasks)	348 (51.9%)	48.2% - 55.7%	497935 (59.8%)	59.7% - 59.9%
Statistical review is performed by an additional, specialist reviewer	81 (12.1%)	9.7% - 14.5%	152040 (18.3%)	18.2% - 18.3%
Statistics review is performed through automatic computer software	0 (0.0%)	0.0% - 0.0%	4129 (0.5%)	0.5% - 0.5%

Table 5.50 Level and type of statistical review related to number of non- retracted and retracted articles in our sample.

When focussing on the different reasons for retraction, the data show that incorporated statistical review is associated with a significantly lower number of retractions due to fake review, authorship- and ethical issues ($F(3,1303) = 63.503$, $p < 0.001$). On the contrary, we do not see any substantial influence on the number of retractions due to errors or issues related to data, which arguably are more related to statistics. The effect of specialist, incorporated or IT-assisted statistics review on aspects of the manuscript directly related to data analysis remains open for further study.

The fact that retraction rates are particularly high in journals classifying statistics as 'irrelevant' to their research, while similar effects on retraction rates are measured for journals paying either no special attention or use specialist reviewers for statistics, would suggest that many retractions appear which are unrelated to statistics. However, additional statistics review is associated with a lower retraction rate for precisely the categories of retractions where an effect could be expected, raising additional questions. Do specialist statistics reviewers only review statistics or do they in practice consider the entire manuscript? This would, for example, explain why additional, specialist reviewers reduce the retraction rate for fake reviews. In general, higher attention for statistics is used with the intention to prevent tweaking of data or statistics. Thus it seems highly unlikely that this lower retraction rate is due to lax editorial attitudes towards problematic research. A better capability of early detection of such research seems to be a more plausible explanation even though our data cannot provide a definitive answer to this question.

5.4.10. External sources

Partly due to the increasing burden on the peer review system, new procedures have emerged to reduce the number of times a single manuscript potentially needs to be reviewed through cooperation between various parties. One procedure designed to achieve this goal, is that of 'cascading peer review'. In this procedure, (partner) journals redirect a rejected manuscript to another (potentially more suitable) journal, along with the review reports, allowing the new journal to quickly decide on the manuscript's quality, without having to perform another round of reviews (Barroga 2013; Davis 2010). Other procedures for sharing review reports are those in which commercial review platforms assist in review (Pattinson and Prater 2017; Research Square 2017), or in which the wider community (usually in a post-publication prcedure) is invited to review a manuscript. In addition to reducing the burden on the review system, automatically (re-)directing manuscripts to the most suitable journal after review might reduce perverse incentives for authors, such as rewarding overstated conclusions to get work published. This would reduce the risk of retraction, since an incentive to overstate conclusions may provoke questionable research practices. On the other hand, it might also work in the opposite direction by relaxing review standards and allowing authors to neglect nuances, in the confidence that their work will eventually get published somewhere anyway (Horbach and Halffman 2018).

Our data (table 5.11), suggest no difference in retraction rates due to the usage of review reports from external sources ($\Lambda(3) = 42.270$, $p < 0.0001$). No differences were observed between research areas (WALD = 1.052, df = 5, $p = 0.958$), nor between reasons for retraction ($F(2, 813) = 4.166$, $p = 0.016$), hence suggesting similar effects in all research areas and for all types of problematic research.

Reviews from external sources	Retracted	95% CI	Non-retracted	95% CI
No reviews from external sources are used	422 (63.0%)	58.8% - 67.2%	545443 (65.5%)	65.4% - 65.6%
Reviews from other (partner) journals are used	90 (13.4%)	10.5% - 16.4%	244385 (29.4%)	29.3% - 29.5%
Reviews from commercial review platforms are used	0 (0.0%)	0.0% - 0.0%	3971 (0.5%)	0.5% - 0.5%
Reviews performed by the wider community are used	1 (0.1%)	0.0% - 0.5%	3159 (0.4%)	0.4% - 0.4%

Table 5.11 Extent to which reviews from external sources are used related to number of non- retracted and retracted articles in our sample.

The fact that no significant differences were found for review reports from commercial platforms or the wider community might be attributed to the low number of articles going through these kinds of review. Hence the effect of those review procedures remains to be studied. The positive effect of sharing review reports with partner journals on the number of retractions is promising, in the sense that sharing review reports potentially not only lowers the burden on the review system, but also improves the quality of the published literature. Here too, since external sources are typically used by journals trying to improve peer review, lower retraction rates are unlikely to be a sign of low willingness to act against problematic research, but rather of a high capability to detect it.

5.4.11. Digital tools

One of the most promising innovations in peer review's error and fraud detection is probably the introduction of digital tools such as plagiarism detection software, image manipulation software, software to check references (for instance for references to retracted articles), or software to assist in statistical review. Such digital tools have been implemented in a wide variety of journals with specific detection objectives (Elizondo et al. 2017; BioMed Central 2017; Scheman and Bennett 2017), and the expectation of reduced retraction rates.

Our data (Table 5.12) indeed suggest a significant relation between the usage of digital tools as assistance in peer review and retraction rates ($\Lambda(4) = 42.270$, $p < 0.0001$). In particular, more retractions occur when articles were reviewed without the assistance of digital tools and when (only) software to scan references was used. In this, subject area was a significant mediating factor (WALD = 69.496, df = 15, $p < 0.0001$), with stronger effects in the Social Sciences and Humanities. In addition, the usage of various digital tools has a specific effect on different reasons for retraction ($F(4, 1880) = 27.990$, $p < 0.001$). When no tools are used, we witness more retractions for plagiarism and falsification, while those retractions are sparse when plagiarism detection software is used. Similar to previous attributes, these lower retraction rates seem unlikely to be due to lax editorial attitudes towards problematic research.

In contrast, when software to check references is used, we witness more retractions for fake review and for issues with references. The latter is clearly contrary to what should be expected, but might be explained by the sensitivity of these journals to issues with references, making them more willing to file retractions for such reasons. Here, higher retraction rates might hence be a sign of a more pro-active policy in using retractions to address issues with problematic research.

Level of structure		No digital tools are used	Plagiarism detection software	Digital tools to check references	Digital tools to detect image manipulation	Digital tools to assess statistics
All articles (percentage of all articles)		259601	482414	206083	69533	23567
Non-retracted (percentage of non-retracted papers)	Total	259321 (31.1%)	482037 (57.9%)	205861 (24.7%)	69485 (8.3%)	23561 (2.8%)
	Social Science & Humanities	25389 (47.1%)	28096 (52.1%)	8089 (15.0%)	5105 (9.5%)	2155 (4.0%)
	Biomedical & Health Sciences	161264 (42.2%)	192353 (50.3%)	50651 (13.2%)	30135 (7.9%)	10622 (2.8%)
	Physical Sciences & Engineering	29085 (14.8%)	156573 (79.6%)	94567 (48.1%)	7688 (3.9%)	1014 (0.5%)
	Life & Earth Sciences	36911 (23.1%)	84049 (51.9%)	48009 (30.0%)	26557 (16.6%)	8238 (5.1%)
	Mathematics & Computer Sciences	6231 (16.3%)	19882 (52.0%)	4177 (10.9%)	0 (0.0%)	1533 (4.0%)
Retracted (percentage of retracted papers)	Total	280 (41.8%)	377 (56.3%)	222 (33.1%)	48 (7.2%)	6 (0.9%)
	Social Science & Humanities	9 (30.0%)	21 (70.0%)	7 (23.3%)	0 (0.0%)	0 (0.0%)
	Biomedical & Health sciences	242 (65.8%)	120 (32.6%)	29 (7.9%)	39 (10.6%)	6 (1.6%)
	Physical sciences & Engineering	2 (1.1%)	180 (98.4%)	163 (89.1%)	1 (0.5%)	0 (0.0%)
	Life & Earth sciences	26 (38.8%)	35 (52.2%)	16 (23.9%)	8 (11.9%)	0 (0.0%)
	Mathematics & computer sciences	1 (4.8%)	20 (95.2%)	6 (28.6%)	0 (0.0%)	0 (0.0%)

Table 5.12 Usage of digital tools in peer review related to number of non- retracted and retracted articles in our sample.

Another way of testing the effectiveness of digital tools is by comparing submissions to journals prior to and after the installation of digital tools. Because the number of changes in review procedures is relatively small, we can only meaningfully perform such an analysis specifically for plagiarism scanning tools. For this case, our results show that journals installing plagiarism software published 70097 articles prior to the introduction of the software, leading to 38 retractions, 11 of them for plagiarism or duplication. These same journals published 41043 articles after the introduction of the software, leading to 19 retractions, of which only 1 for plagiarism or duplication. Even though these numbers are still relatively small, it does suggest that the introduction of plagiarism software is an effective way of preventing retractions, specifically for reasons of plagiarism or duplication.

5.4.12. Reader commentary

A last peer review characteristic analysed in our study concerns the extent to which journals facilitate reader commentary after the review process. Even if reader commentary is not used as a formal review mechanism, it may provide effective ways to assess manuscript quality and point out potential strengths or weaknesses to future readers. Digital technologies allow journals to provide in-channel facilities for direct reader commentary on their website, for instance in the form of blogs or forums, as well as directing readers to out-of-channel platforms that facilitate reader commentary such as *PubPeer* (PubPeer Foundation 2017). Reader commentary, and thereby heightened scrutiny on published manuscripts, may deter authors from engaging in dubious publication practices, leading to fewer retractions. At the same time, the increased detection likelihood could also increase retraction rates.

Our analyses (Table 5.13) demonstrate significantly higher levels of retractions with greater facilities for reader commentary, especially when in-channel reader commentary is facilitated ($\Lambda(2) = 108.759$, $p < 0.0001$). This suggests that higher scrutiny by readers does indeed increase the detection likelihood of problematic research reports that slipped through review, thereby leading to more retractions. In addition, we find significant differences between research fields (WALD = 20.967, df = 7, $p = 0.004$) and various reasons for retraction ($F(2, 1306) = 26.607$, $p < 0.001$). In particular we find strong effects in the Biomedical & Health Sciences as well as in the Life & Earth Sciences. Regarding reasons for retraction, peer review procedures with in-channel reader commentary is associated with fewer retractions due to fake review and issues with references, while there are more retractions for falsification and image/data issues compared to review procedures without direct reader commentary.

The fact that more retractions appear when readers are able to comment on articles, suggests that reader commentary is a way to flag issues and put the retraction mechanism in motion. This might hence be a specific effect of how journals deal with errors in the literature. Whereas issues might be addressed internally, or in closed communication with the authors, this becomes more difficult when errors have been publicly announced and reported in reader comments. Again, higher retraction rates might hence be a sign of heightened willingness to address issues with problematic research by means of retractions. However, the extent to which reader commentary leads to retractions should be researched in more detail.

Level of reader commentary		No direct reader commentary is facilitated	Reader commentary is facilitated on the journal's website	Out-of-channel reader commentary is facilitated
All articles (percentage of all articles)		627267 (75.3%)	203692 (24.4%)	3735 (0.4%)
Non-retracted (percentage of non-retracted papers per research area)	Total	626849 (75.3%)	203441 (24.4%)	3712 (0.4%)
	Social Science & Humanities	45185 (83.8%)	7057 (13.1%)	1691 (3.1%)
	Biomedical & Health sciences	215921 (56.4%)	166460 (43.5%)	201 (0.1%)
	Physical sciences & Engineering	185461 (94.3%)	10841 (5.5%)	360 (0.2%)
	Life & Earth sciences	141586 (88.5%)	18473 (11.5%)	1460 (0.9%)
	Mathematics & computer sciences	37781 (98.9%)	423 (1.1%)	0 (0.0%)
Retracted (percentage of retracted papers per research area)	Total	418 (62.4%)	251 (37.5%)	23 (3.4%)
	Social Science & Humanities	27 (90.0%)	3 (10.0%)	0 (0.0%)
	Biomedical & Health sciences	122 (33.2%)	246 (66.8%)	0 (0.0%)
	Physical sciences & Engineering	182 (99.5%)	0 (0.0%)	1 (0.5%)
	Life & Earth sciences	65 (97.0%)	2 (3.0%)	22 (32.8%)
	Mathematics & computer sciences	21 (100.0%)	0 (0.0%)	0 (0.0%)

Table 5.13 Level of reader commentary related to the number of non- retracted and retracted articles per research area.

5.4.13. Summary results

Combining the results from the previous sections, Table 5.14 presents an overview of our results. The table lists the significant correlations between retractions and peer review procedures, as well as significant interaction terms with either research area or reasons for retraction.

Review attribute	Significant effects on retractions	Significant interaction with research area	Significant variation between reasons for retraction
1 – Timing	Pre-submission review is related to fewer retractions.	No	No
2 – Criteria	Focussing on anticipated impact and novelty is related to more retractions.	No	No
3 –Type reviewer	Involvement of editor-in-chief and wider community is related to fewer retractions.	Yes*	Yes*
4 – Author anonymity	Blinding author identities is related to fewer retractions.	No	Yes
5 – Reviewer anonymity	No significant effects.	No	No
6 – Review reports	No significant effects in general, some effect for specific reasons for retraction and difference in effect for different research areas.	Yes*	Yes*
7 – Interaction	No interaction is related to fewer retractions, having authors respond to review reports is related to more retractions.	Yes*	Yes*
8 – Structure	Unstructured and structured review is related to fewer retractions, semi-structured to more.	Yes*	Yes*
9 – Statistics	More retractions in journals deeming statistics ‘not relevant’, fewer retractions in journals paying no specific attention to statistics, incorporating statistics in review, or employing specialist statistics reviewers.	Yes*	Yes*
10 – External sources	Using review reports from partner journals is related to fewer retractions.	No	No
11 – IT-tools	Not using digital tools and using tools to check references is related to more retractions.	Yes	Yes*
12 – Reader commentary	Not facilitating reader commentary is related to fewer retractions, facilitating in-channel commentary is related to more retractions.	Yes	Yes*

Table 5.14 Overview of review procedures’ significant effects on retraction rates. Significance is measured at the level of $\alpha = 0.01$. $*p < 0.001$.

5.5. Limitations

Our research project may have suffered from various limitations. First, some selection and response bias may have been introduced in journal review data collection. When sampling editors’ email addresses, we searched the Web of Science for editorials written in 2017. Hence only journals indexed in the Web of Science were included. This may have caused some young, non-English, or smaller, niche journals to be excluded from our sample. In terms of publications numbers or retractions, this arguably has little effect on our general results. However, some of these excluded journals may have particularly innovative review procedures, potentially underestimating the spread of peer review innovations.

On the other hand, from the journals sampled our survey, those paying specific attention to their review process, and those particularly keen on innovating their review procedures, were arguably most likely to respond to our survey. Hence, from the journals indexed by Web of Science, we expect to have obtained responses exactly from the most innovative journals with respect to peer review procedures. Indeed, our data suggests this, with substantially more innovations reported in journals that responded very quickly to our survey request, compared to journals responding later, after one or two reminders.

To assess the effectiveness of different peer review procedures to detect fraudulent or erroneous research we used retracted journals articles as a proxy of problematic publications. We did not consider the number of published errata or corrections because these are not collected as systematically and reliably as the number of retracted articles. This approach has some limitations. First, it allows us only to trace problematic research articles that have been identified as such. Doubtless, many articles containing errors have not been retracted, either because the errors have not yet been identified, because editors are hesitant about retraction measures, authors are successfully fighting retraction measures, because of ambivalences or disagreement about what constitutes error, or a host of other reasons. Hence, our control group of non-retracted articles contains (potentially many) papers that should have been among the group of problematic articles. Nevertheless, we expect this number to be relatively small compared to the size of the control group, hence only moderately affecting our results.

Second, by using retracted journal articles, we employ problematic articles that slipped through peer review. With this method we were not able to identify problematic manuscripts that did *not* pass through peer review, i.e. those manuscripts in which errors were identified during the review process. The validity of our findings hence rests on the assumption that problematic articles were submitted via a uniform distribution to journals using different peer review models, i.e. that erroneous research was not submitted with higher probability to journals holding a specific review procedure.

A final limitation of our study rests in the survey approach to collect data about peer review procedures. Even though we tested our survey before distributing it, this type of data collection is prone to misunderstanding of the wording or confusion about definitions, as well as incomplete knowledge of the editors of (past) review procedures. This might have led editors to classify their journal's review procedures differently from how we intended it and hence have influenced our analyses. Procedures reported by editors may not always reflect review practice.

With these limitations, we believe our research nevertheless provides valuable recommendations for journal editors to effectively design or reconsider their review procedures.

5.6. Conclusion

In spite of various calls for more research on the effectiveness of various peer review procedures, actual evidence is rare. This study addresses precisely this knowledge gap. Our analyses reveal major differences between various ways of organising peer review and the number of retracted journal articles going through these review procedures. Thereby, they provide an indication of the effectiveness of various peer review models in detecting erroneous or fraudulent research. Even though hard causal connections cannot be made, our data suggest that some review procedures are significantly more effective in preventing retractions. In particular, author blinding seems more effective than reviewer blinding; involving the wider community in review seems beneficial; using digital tools to assist review is related to fewer retractions, as is constraining interaction between authors and reviewers, and using pre-submission review procedures such as registered reports. In addition, our data suggest differences in the effectiveness of various review procedures between scientific disciplines, as well as between specific reasons for retraction. Thereby we present a systematic comparison of review procedures' effectiveness in detecting problematic research publications.

Our results provide specific recommendations and guidance for journal editors and publishers on how to improve the ability of their review processes to detect forms of problematic research that are particularly relevant in their subject area (with the understanding that preferences for particular peer review procedures are informed by many other considerations besides the prevention of retractions). For example, image manipulation or issues related to authorship may be of particular concern to journals in specific research areas. Our findings provide suggestions for editors to organise peer review in order to address exactly those issues, thereby allowing them to tweak review models to their specific demands.

In addition, our results suggest directions for future research in order to identify and assess the specific mechanisms underlying the effects of different review procedures. Future research should elucidate the causal connections behind the identified strong correlations. What specific mechanism, for instance, makes double-blind review better capable of detecting errors in research records than single-blind review? And what causes augmented interaction between authors and reviewers to be related to more retracted journal articles? A closer look at the practice of peer review, and particularly in those cases that led to retractions, could clarify this further. (For some very specific questions, we refer to the discussion of results per peer review procedure above.) To allow for such future initiatives, we call on journals to be more transparent about their editorial policies and review procedures. Providing such information on journals' webpages would allow for more inclusive analyses, strengthen the power of the analyses and thereby lead to more detailed results.

Much uncertainty exists about which peer review innovations actually work. Our assessment of peer review procedures addresses this knowledge gap and may provide valuable assistance to journal editors, publishing houses, and even funding agencies. We hope this will also contribute to improved research integrity.

5.7. Postscript

After the publication of this chapter in *Scientometrics*, we have been granted a ZonMw implementation grant to establish a *Platform for Responsible Editorial Practices (PREP).* Within this project, starting autumn 2019, we will create an online platform allowing journal editors to report their editorial procedures in an open access database. This will both facilitate transparency of editorial policies as well as enrich the dataset used for the analyses presented in this chapter. We hence plan to perform follow-up analyses once the new database is online and information about additional journals is available.

5.8. References

Altman, D. G. (1998). Statistical reviewing for medical journals. *Statistics in Medicine, 17*(23), 2661-2674, doi:10.1002/(sici)1097-0258(19981215)17:23<2661::aid-sim33>3.0.co;2-b.

Amsen, E. (2014). What is open peer review. In F. Research (Ed.), *F1000 Researchblog* (Vol. 2017).

Azoulay, P., Bonatti, A., & Krieger, J. L. (2017). The career effects of scandal: Evidence from scientific retractions. *Research Policy, 46*(9), 1552-1569, doi:https://doi.org/10.1016/j.respol.2017.07.003.

Bakker, M., & Wicherts, J. M. (2011). The (mis)reporting of statistical results in psychology journals. *Behavior Research Methods, 43*(3), 666-678, doi:10.3758/s13428-011-0089-5.

Baldwin, M. (2015). Credibility, peer review, and Nature, 1945–1990. *Notes and Records: the Royal Society journal of the history of science, 69*(3), 337-352, doi:10.1098/rsnr.2015.0029.

Baldwin, M. (2017). In referees we trust? *Physics Today, 70*(2), 44.

Barroga, E. F. (2013). Cascading peer review for open-access publishing. *Eur Sci Ed, 39*, 90-91.

Biagioli, M. (2002). From book censorship to academic peer review. *Emergences: Journal for the Study of Media & Composite Cultures, 12*(1), 11-45.

Biderman, A. D., & Reiss, A. J. (1967). On Exploring the "Dark Figure" of Crime. *The ANNALS of the American Academy of Political and Social Science, 374*(1), 1-15, doi:10.1177/000271626737400102.

BioMed Central (2017). What might peer review look like in 2030? In B. Central (Ed.). figshare.

BMJ Open (2018). Reviewer guidelines. http://bmjopen.bmj.com/pages/reviewerguidelines/.

Bohannon, J. (2013). Who's Afraid of Peer Review? [Article]. *Science, 342*(6154), 60-65.

Callahan, J. L. (2017). The retrospective (im)moralization of self-plagiarism: Power interests in the social construction of new norms for publishing. *Organization, 25*(3), 305-319, doi:10.1177/1350508417734926.
Cao, H. H., Han, B., Hirshleifer, D., & Zhang, H. H. (2009). Fear of the unknown: Familiarity and economic decisions. *Review of Finance, 15*(1), 173-206.
Carlisle, J. B. (2017). Data fabrication and other reasons for non-random sampling in 5087 randomised, controlled trials in anaesthetic and general medical journals. *Anaesthesia, 72*(8), 944-952, doi:10.1111/anae.13938.
Chambers, C. D. (2013). Registered reports: a new publishing initiative at Cortex. *Cortex, 49*(3), 609-610.
Chambers, C. D., Feredoes, E., Muthukumaraswamy, S. D., & Etchells, P. (2014). Instead of" playing the game" it is time to change the rules: Registered Reports at AIMS Neuroscience and beyond. *AIMS Neuroscience, 1*(1), 4-17.
Cintas, P. (2016). Peer review: from recognition to improved practices. *Fems Microbiology Letters, 363*(12), doi:10.1093/femsle/fnw115.
Claxton, L. D. (2005). Scientific authorship: Part 1. A window into scientific fraud? *Mutation Research/Reviews in Mutation Research, 589*(1), 17-30, doi:http://dx.doi.org/10.1016/j.mrrev.2004.07.003.
Cokol, M., Ozbay, F., & Rodriguez-Esteban, R. (2008). Retraction rates are on the rise. [Letter]. *Embo Reports, 9*(1), 1, doi:10.1038/sj.embor.7401143.
CWTS (2018). CWTS Leiden Ranking - Fields. http://www.leidenranking.com/information/fields.
Davis, P. (2010). Cascading Peer-Review: The future of open access publishing? *The Scholarly Kitchen*.
Drenth, P. J. D. (2006). Responsible conduct in research. [Article; Proceedings Paper]. *Science and Engineering Ethics, 12*(1), 13-21, doi:10.1007/pl00022265.
Elizondo, A. R., De Rijcke, S., & Van Leeuwen, T. (2017). Handling publishing misconduct: tools used by publishing houses and editors. In PRINTEGER (Ed.). Leiden: Leiden University.
EMBO Press (2017). About: Cross Peer Review. http://emboj.embopress.org/about#Cross_peer_review.
Epskamp, S., & Nuijten, M. (2014). statcheck: Extract statistics from articles and recompute p values (R package version 1.0. 0.).
Fanelli, D. (2009). How Many Scientists Fabricate and Falsify Research? A Systematic Review and Meta-Analysis of Survey Data. [Article]. *Plos One, 4*(5), 11, doi:10.1371/journal.pone.0005738.
Fanelli, D., Costas, R., & Lariviere, V. (2015). Misconduct Policies, Academic Culture and Career Stage, Not Gender or Pressures to Publish, Affect Scientific Integrity. [Article]. *Plos One, 10*(6), 18, doi:10.1371/journal.pone.0127556.
Fang, F. C., Steen, R. G., & Casadevall, A. (2012). Misconduct accounts for the majority of retracted scientific publications. [Article]. *Proceedings of the National Academy of Sciences of the United States of America, 109*(42), 17028-17033, doi:10.1073/pnas.1212247109.
Frontiers (2014). Frontiers is a community-rooted, open-access academic publisher. https://www.frontiersin.org/about/reviewsystem.

Fyfe, A., Coate, K., Curry, S., Lawson, S., Moxham, N., & Rostvik, C. M. (2017). Untangling academic publishing: a history of the relationship between commercial interests, academic prestige and the circulation of research.
George, S. L. (1985). STATISTICS IN MEDICAL JOURNALS - A SURVEY OF CURRENT POLICIES AND PROPOSALS FOR EDITORS. *Medical and Pediatric Oncology, 13*(2), 109-112, doi:10.1002/mpo.2950130215.
Godlee, F. (2002). Making reviewers visible: openness, accountability, and credit. *Jama, 287*(21), 2762-2765.
Goodman, S. N. Statistical Review in Biomedical Journals: Can We Get SMARTA? In *Eight International Congress on Peer Review and Scientific Publication, Chicago, 2017*
Harris, J., Graue, M., Dunning, T., Haltbakk, J., Austrheim, G., Skille, N., et al. (2015). Involving people with diabetes and the wider community in diabetes research: a realist review protocol. [journal article]. *Systematic Reviews, 4*(1), 146, doi:10.1186/s13643-015-0127-y.
He, T. W. (2013). Retraction of global scientific publications from 2001 to 2010. [Article]. *Scientometrics, 96*(2), 555-561, doi:10.1007/s11192-012-0906-3.
Hesselmann, F., Graf, V., Schmidt, M., & Reinhart, M. (2017). The visibility of scientific misconduct: A review of the literature on retracted journal articles. *Current Sociology, 65*(6), 814-845, doi:10.1177/0011392116663807.
Hiney, M. (2015). What it Means, Why it Is Important and How we Might Protect it. Brussels: Science Europe, working group research integrity.
Hopp, C., & Hoover, G. A. (2017). How prevalent is academic misconduct in management research? *Journal of Business Research, 80*, 73-81, doi:https://doi.org/10.1016/j.jbusres.2017.07.003.
Horbach, S. P. J. M., & Halffman, W. (2017a). The extent and causes of academic text recycling or 'self-plagiarism'. *Research Policy*, doi:https://doi.org/10.1016/j.respol.2017.09.004.
Horbach, S. P. J. M., & Halffman, W. (2017b). The ghosts of HeLa: How cell line misidentification contaminates the scientific literature. *Plos One, 12*(10), e0186281, doi:10.1371/journal.pone.0186281.
Horbach, S. P. J. M., & Halffman, W. (2018). The changing forms and expectations of peer review. [journal article]. *Research Integrity and Peer Review, 3*(1), 8, doi:10.1186/s41073-018-0051-5.
Horner, J., & Minifie, F. D. (2011). Research Ethics II: Mentoring, Collaboration, Peer Review, and Data Management and Ownership. *Journal of Speech Language and Hearing Research, 54*(1), S330-S345, doi:10.1044/1092-4388(2010/09-0264).
Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Med, 2*(8), e124.
John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling. [Article]. *Psychological Science, 23*(5), 524-532, doi:10.1177/0956797611430953.
Karabag, S. F., & Berggren, C. (2016). Misconduct, Marginality and Editorial Practices in Management, Business and Economics Journals. *Plos One, 11*(7), e0159492, doi:10.1371/journal.pone.0159492.

Kaushal, R., Bates, D. W., Landrigan, C., McKenna, K. J., Clapp, M. D., Federico, F., et al. (2001). Medication errors and adverse drug events in pediatric inpatients. *Jama, 285*(16), 2114-2120.

Kharasch, E. D., & Houle, T. T. (2018). Seeking and reporting apparent research misconduct: errors and integrity. *Anaesthesia, 73*(1), 125-126, doi:10.1111/anae.14147.

KNAW (2014). CORRECT CITATION PRACTICE, Academy Advisory Memorandum. KNAW, Committee on Citation Practice.

Knoepfler, P. (2015). Reviewing post-publication peer review. *Trends in Genetics, 31*(5), 221-223.

LaFollette, M. C. (1992). *Stealing into print: fraud, plagiarism, and misconduct in scientific publishing*: Univ of California Press.

Larivière, V., Haustein, S., & Mongeon, P. (2015). The Oligopoly of Academic Publishers in the Digital Era. *Plos One, 10*(6), e0127502, doi:10.1371/journal.pone.0127502.

Lee, C. J., Sugimoto, C. R., Zhang, G., & Cronin, B. (2013). Bias in peer review. [Review]. *Journal of the American Society for Information Science and Technology, 64*(1), 2-17, doi:10.1002/asi.22784.

Madlock-Brown, C. R., & Eichmann, D. (2015). The (lack of) Impact of Retraction on Citation Networks. [Article]. *Science and Engineering Ethics, 21*(1), 127-137, doi:10.1007/s11948-014-9532-1.

Martin, B. (1992). Scientific fraud and the power structure of science. *Prometheus, 10*(1), 83-98.

Martinson, B. C., Anderson, M. S., & de Vries, R. (2005). Scientists behaving badly. [Editorial Material]. *Nature, 435*(7043), 737-738, doi:10.1038/435737a.

Mellor, D. (2016). Registered Reports: Peer review before results are known to align scientific values and practices. https://cos.io/rr/.

Montgomery, K., & Oliver, A. L. (2017). Conceptualizing Fraudulent Studies as Viruses: New Models for Handling Retractions. [journal article]. *Minerva, 55*(1), 49-64, doi:10.1007/s11024-016-9311-z.

Nosek, B. A., & Lakens, D. (2014). Registered reports: A method to increase the credibility of published results. *Social Psychology, 45*(3), 137-141, doi:10.1027/1864-9335/a000192.

Nuijten, M. B., Hartgerink, C. H. J., van Assen, M. A. L. M., Epskamp, S., & Wicherts, J. M. (2016). The prevalence of statistical reporting errors in psychology (1985–2013). [journal article]. *Behavior Research Methods, 48*(4), 1205-1226, doi:10.3758/s13428-015-0664-2.

Okike, K., Hug, K. T., Kocher, M. S., & Leopold, S. S. (2016). Single-blind vs double-blind peer review in the setting of author prestige. *Jama, 316*(12), 1315-1316, doi:10.1001/jama.2016.11014.

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), aac4716.

Pattinson, D., & Prater, C. (2017). *Assessment of the Prevalence of Integrity Issues in Submitted Manuscripts.* Paper presented at the Eigth International Congress on Peer Review and Scientific Publication, Chicago,

Peters, D. P., & Ceci, S. J. (1982). Peer-review practices of psychological journals: The fate of published articles, submitted again. *Behavioral and Brain Sciences, 5*(2), 187-195, doi:10.1017/S0140525X00011183.

PLOS (2018). Guidelines for Reviewers. http://journals.plos.org/plosone/s/reviewer-guidelines2018.
Pontille, D., & Torny, D. (2014). The Blind Shall See! The Question of Anonymity in Journal Peer Review. *Ada: A Journal of Gender, New Media, and Technology, 4*, doi:doi:10.7264/N3542KVW. .
Pöschl, U. (2012). Multi-stage open peer review: scientific evaluation integrating the strengths of traditional peer review with the virtues of transparency and self-regulation. *Frontiers in computational neuroscience, 6*.
PubPeer Foundation (2017). PubPeer. https://pubpeer.com/.
Rennie, D. (2016). Make peer review scientific. [Editorial Material]. *Nature, 535*(7610), 31-33.
Research Square (2017). Editorial Checks & Badges. https://www.researchsquare.com/publishers/badges.
Retraction Watch (2018). The Retraction Watch Retraction Database. http://retractiondatabase.org/RetractionSearch.aspx?
Ross-Hellauer, T. (2017). *What is open peer review? A systematic review [version 2; referees: 1 approved, 3 approved with reservations]* (Vol. 6, Vol. 588).
Sage Open (2018). More information: Description. https://uk.sagepub.com/en-gb/eur/journal/sage-open#description2018.
Schekman, R., Watt, F., & Weigel, D. (2013). The eLife approach to peer review. *Elife, 2*, doi:10.7554/eLife.00799.
Scheman, R., & Bennett, C. N. Assessing the Outcomes of Introducing a Digital Image Quality Control Review Into the Publication Process for Research Articles in Physiology Journals. In *International Congress on Peer Review and Scientific Publication, Chicago, 2017*
Schmidt, M. (2018). An Analysis of the Validity of Retraction Annotation in PubMed and the Web of Science. [Article]. *Journal of the Association for Information Science and Technology, 69*(2), 318-328, doi:10.1002/asi.23913.
Skogan, W. G. (1977). Dimensions of Dark Figure of Unreported Crime. *Crime & Delinquency, 23*(1), 41-50, doi:10.1177/001112877702300104.
Smith, R. (1999). Opening up BMJ peer review. *A beginning that should lead to complete transparency, 318*(7175), 4-5, doi:10.1136/bmj.318.7175.4.
Smith, R. (2006). Peer review: a flawed process at the heart of science and journals. *Journal of the Royal Society of Medicine, 99*(4), 178-182.
Steen, R. G. (2011). Retractions in the medical literature: how many patients are put at risk by flawed research? [Article]. *Journal of Medical Ethics, 37*(11), 688-692, doi:10.1136/jme.2011.043133.
Stitzel, B., Hoover, G. A., & Clark, W. (2018). More on Plagiarism in the Social Sciences*. *Social Science Quarterly*, n/a-n/a, doi:10.1111/ssqu.12481.
Stroebe, W., Postmes, T., & Spears, R. (2012). Scientific Misconduct and the Myth of Self-Correction in Science. *Perspectives on Psychological Science, 7*(6), 670-688, doi:10.1177/1745691612460687.
Tennant, J. P., Dugan, J. M., Graziotin, D., Jacques, D. C., Waldner, F., Mietchen, D., et al. (2017). A multi-disciplinary perspective on emergent and future innovations in peer review. *F1000Research, 6*(1151).

Tramer, M. R., Reynolds, D. J. M., Moore, R. A., & McQuay, H. J. (1997). Impact of covert duplicate publication on meta-analysis: a case study. [Article]. *British Medical Journal, 315*(7109), 635-640.
Van Buggenhout, M., & Christiaens, J. (2016). Deviance in science: a criminological analysis. In PRINTEGER (Ed.). Brussels.
van der Heyden, M. A. G., van de Ven, T. D., & Opthof, T. (2009). Fraud and misconduct in science: the stem cell seduction. *Netherlands Heart Journal, 17*(1), 25-29.
Van Leeuwen, T., Reyes Elizondo, A., & De Rijcke, S. (2017). Analysis of papers retracted in the Web of Science. In PRINTEGER (Ed.). Leiden.
Van Noorden, R. (2011). The trouble with retractions. *Nature, 478*(7367), 26-28.
Walker, R., & Rocha da Silva, P. (2015). Emerging trends in peer review—a survey. [Review]. *Frontiers in Neuroscience, 9*(169), doi:10.3389/fnins.2015.00169.
Waltman, L., & van Eck, N. J. (2012). A new methodology for constructing a publication-level classification system of science. *Journal of the American Society for Information Science and Technology, 63*(12), 2378-2392, doi:10.1002/asi.22748.
Zhang, H. Y. (2010). CrossCheck: an effective tool for detecting plagiarism. *Learned publishing, 23*(1), 9-14.
Zuckerman, H., & Merton, R. K. (1971). Patterns of evaluation in science: Institutionalisation, structure and functions of the referee system. [journal article]. *Minerva, 9*(1), 66-100, doi:10.1007/bf01553188.

Chapter 6 - Peer review and editorial evaluation: cautious innovator or sleepy giant?

This chapter is accepted for publication as:
Horbach, S. P. J. M., & Halffman, W. (In Press). Journal Peer Review and Editorial Evaluation: Cautious Innovator or Sleepy Giant? Minerva.

6

Peer review of journal submissions plays a key role in the work of virtually all academics and researchers. Reviewer comments become conditions for publication as papers are "put through peer review". The expression is used as if peer review is a singular system, but in fact journal peer review and editorial evaluation are now more diverse than ever. A host of enthusiastic innovators has proposed and experimented with new procedures and technologies, but little is known about whether these innovations manage to convince other journal editors. The innovations are advocated with a wide range of concerns over peer review and editorial evaluation, ranging from its fairness, its ability to assess the solidity of statistics or text recycling, to its transparency or cost. In response to such concerns, editorial innovations now include procedures such as open peer review, registered reports (reviewing research protocols rather than results), various scanners (including statistics and text similarity scanners), or external review platforms that offer outsourced peer review. In this sense, traditional peer review has become enveloped in editorial procedures that employ software support tools to guarantee the quality or integrity of publications, at least in some parts of the research publication system. Some functions previously attributed to peer review are now covered, or at least supported, by software tools or organisational innovations in editorial management. It is this wider set of editorial assessment practices (in which peer review still holds a central position) that constitutes the topic of our analysis. Table 6.1 outlines some of the major innovations in peer review and editorial procedures over the past decades.

Editorial procedure	First implemented in	First implemented by
Plagiarism detection software	1980s	Various universities and journals to detect similarity in computer programming code
Open review	Late 1990s	*British Medical Journal*
Post-publication review	1997	*Electronic Transactions in Artificial Intelligence*
Commercial review platforms assisting in review	2004	*Research Square*
Image manipulation software	2010	Multiple journals, including journals by the *American Psychological Society*
Interactive review: Discussion among reviewers, editors and authors	2011	*elife*
Registered reports	2013	*Cortex*
Digital tools for statistics review	2014	*StatCheck*

Table 6.1 Some of the major innovations in editorial procedures over the past decades, including their establishment date and initial implementer

While the innovations have been published and are often passionately defended by their proponents, there is little research on the actual distribution of the current editorial procedure diversity. Available research has focused on a single review aspect, for instance single-blind vs.

double-blind (Pontille and Torny 2014); or on journals in a specific research area; or on a single publisher (Taylor & Francis 2015). In addition, these studies commonly focus on perceptions of review procedures. This leaves open questions regarding the implementation of new editorial and review procedures, as well as what motivates editors or publishers to engage or reject novel procedures.

This paper will address these questions based on a survey among journal editors of 361 research journals, covering a wide range of research fields. Using the results of the survey, we aim to answer two main questions. First, based on an inventory of different editorial and peer review models (Horbach and Halffman 2018b), we aim to assess how often these models are employed and how their usage is distributed over different academic disciplines and publishers. Second, we set out to elucidate the process of editorial innovation, asking why some innovations are implemented faster and wider than others. In this chapter, we focus on peer review of journal submissions (as opposed to review of grant applications or other types of review). In this, we will distinguish between the editorial process, encompassing the entire workflow from submission to decisions on acceptance or rejection; and the peer review process, referring to the actual intellectual work done by peers in evaluating a manuscript. For example, the usage of text similarity scanners is part of the editorial process, but complements the actual peer review process.

Our analysis shows that, in spite of enthusiastic innovation, the adoption of new editorial procedures is in fact very slow, with the exception of text similarity scanners. For now, innovations appear to be restricted to specific niches in academic publishing, despite the ardent commitment of their proponents.

After describing peer review and the editorial process from the perspective of innovations as a contentious set of procedures, we describe the method used to gather data. Our findings are organised as follows: after an overview of various editorial procedures' occurrence rate, we analyse their distribution over research fields, publishers, and changes over time. Based on qualitative responses, we can describe the conditions for innovation, as reported by the editors in our sample. After a general overview, we return to specific innovation niches in which editorial procedures are changing more quickly or drastically, in order to analyse the conditions for change. This leads to conclusions about the distribution of editorial procedures, the pace of innovation in the system, as well as some potential factors stimulating or triggering innovation in academic publishing.

6.1. Innovation in a contentious set of procedures

Practices to evaluate the quality of articles submitted to scholarly journals have always been varied and contentious. Even though peer review is nowadays often presented as a universal gold standard guaranteeing the epistemic reliability of published work, the wide-spread use of referees by research journals is in fact relatively new (Fyfe et al. In Press; Fyfe et al. 2015; Baldwin 2015; Csiszar 2018). Systematic use of referees to evaluate submitted work was only introduced in learned societies in the early nineteenth century. Deep into the twentieth century, the use of referees by journal editors was still considered optional (Fyfe et al. In Press; Fyfe et al. 2015; Csiszar 2016). The term "peer review" itself is a neologism that only became common in the 1970s (Baldwin 2018, 2015). The general use of peer review to evaluate knowledge claims became prominent in the specific context of US grant proposal evaluation in the late 70s (Baldwin 2018; LaFollette 1992) and only later became common practice beyond the realm of natural and medical sciences (Moxham and Fyfe 2017). Throughout this history, refereeing has been used in various ways and for various purposes, ranging from state censorship, allocation of printing resources, fraud detection and quality improvement, to the protection of reputations (Fyfe et al. 2017; Biagioli 2002; LaFollette 1992). As Moxham and Fyfe put it:

> "(…) the relative durability of refereeing as a practice should not be mistaken for simple continuity of purpose or of meaning. What it was meant to accomplish, whom it was intended to benefit, and the perception of its virtues and defects varied considerably with time and place." (Moxham and Fyfe 2017, p. 888)

With the gradual uptake of review practices, the expectations of the system have been in constant flux and still remain controversial today (Horbach and Halffman 2018b). Two expectations meet with more or less general agreement. First, peer review is currently generally expected to act as a filter, distinguishing 'good' from 'bad' science. Second, it is widely expected to improve manuscripts: through their comments and feedback, reviewers are expected to assist authors in improving their manuscript's quality (Zuckerman and Merton 1971). However, other expectations are less universal and the understanding of 'quality' may vary considerably. For example, others go further, including expectations about fraud detection, the creation of fair and equal opportunities for all authors, and the establishment of a hierarchy of published results (Bohlin 2004). These expectations are disputed by other actors, mainly publishers, arguing that 'peer review was never designed nor intended to do so.' With the editorial process now encompassing much more than just the act of peer reviewing a manuscript, debates about the role of 'peer review' should now be reformulated as the role of the editorial process in a broader sense, among others due to the introduction of several automated scanners.

In recent years, several additional challenges and concerns surrounding the editorial system have emerged. These include the discussions on open access publishing, potentially drastically impacting on publisher's business models; and other open science initiatives, demanding more transparency from both authors and journals (cOALition S 2018). Also, the rise of alternative publishing formats, such as pre-print servers that no longer require the direct involvement of publishers (Walker and Rocha da Silva 2015), raises further questions about the role of publishers. The involvement of wider communities, outside of invited reviewers, additionally challenges the notion of a ‘peer’ and its relevant expertise: who is sufficiently capable of assessing the quality and merit of a manuscript? And what is the role of metrics, including altmetrics, in assessing manuscripts? (Thelwall et al. 2013)

The differing expectations of the review and editorial system have led to a range of novel review procedures. For instance, the expectation to detect fraud has triggered the development of text and image similarity scanners (Ercegovac and Richardson 2004). An expectation of more extensive reviewing has advocated the involvement of a wider community through post-publication review (Knoepfler 2015), and even of commercial review platforms (Research Square 2017). Furthermore, changing selection criteria, interaction between stakeholders and cooperation between journals have aimed to foster research integrity through review procedures, rather than detecting integrity breaches after they occur (Hames 2014). Open review, including sharing review reports, as well as its radical opposite, the double-blind review system, have emerged out of questions of fairness in review (Rojas 2007; Pontille and Torny 2014; Okike et al. 2016).

These innovations have also been fuelled by concerns over peer review and the publication system in general. An extensive literature of editorials, letters and research articles has emerged claiming that peer review is essentially flawed (R. Smith 2006), that it is inconsistent (Peters and Ceci 1982), slow (Nguyen et al. 2015), and ineffective (R. Smith 2006; Lee et al. 2013). Moreover, it has been characterised as an ‘empty gun’ (R. Smith 2010), ‘with no evidence that it works and no established way to provide [reviewer] training’ (Patel 2014). More recently, claims have emerged about peer review as ‘old fashioned and outdated’ (Tennant et al. 2018), arguing that the peer review system is not keeping pace with rapid changes in (scientific) communication and research practices.

Given these growing concerns about peer review and editorial practices, we might expect these practices to be in considerable flux and novel innovations to distribute quickly throughout the publishing system. We described and systematised the emergence of novel editorial procedures elsewhere (Horbach and Halffman 2018b), but here, we want to investigate the spread and implementation of these innovations among journals. We will understand ‘implementation of innovations’ as ‘changes in the review procedures’, which also allows for the re-implementation of more traditional models of peer review.

The spread and implementation of innovations has previously been widely studied by scholars from sociology, organisation studies, and science and technology studies, to name just a few (e.g. Wejnert 2002; Franke and Shah 2003; Peres et al. 2010). The spread of innovations in most of the traditional literature has been described as the 'diffusion of innovation'. However, this term is slightly misleading: the adaptation or implementation of an innovative technique, process, mechanism or practice requires active agents. Adoption is the result of concrete actions, rather than of some naturally occurring phenomenon, as is the case with diffusing chemicals (MacKenzie and Wajcman 1999; von Hippel 1976; Oudshoorn and Pinch 2007).

Accordingly, several scholars, most notably in the 'social construction of technology' tradition (Bijker et al. 1987), have outlined the various characteristics that describe the process in which innovations travel. First, innovations are picked up by active seekers of knowledge, in an active process of adoption (Greenhalgh et al. 2004). They do not travel driven by their own force and are not taken up by a passive set of recipients. One of the consequences is that strong communication networks in particular foster the spread of innovations. Second, users predominantly search for innovations when they experience a problem or opportunity to improve their practices (Wisdom et al. 2014). Third, users *domesticate* innovations, to make them correspond to their specific conditions and needs (Oudshoorn and Pinch 2007; von Hippel 1976). Fourth, users' expectations of new technologies play a significant role in their willingness to implement such innovations (Brown and Michael 2003; Verbong et al. 2008; Van Lente 1993). Especially given the contentious nature of peer review procedures and the contrasting expectations of the system, these expectations might be prominent factors in the implementation of review innovations. Indeed, what constitutes 'good' review varies among scientists and editors (Taylor & Francis 2015), thereby potentially leading to high diversity in review procedures, but also to obstacles for innovation as it leaves the requirements for novel procedures contentious (Van Lente 1993).

Academic publishing meets at least some of these conditions with increasing and shifting expectations of the system explicitly voiced, pressure from related (social) networks to adopt changes, well-developed communication capacities that facilitate knowledge exchange, and sufficient financial clout (Larivière et al. 2015). In the remainder of this chapter we will present empirical data to demonstrate that, despite these expectations, the editorial process is a fairly stable system, in which change or adoption of novel procedures is actually rare. Even though the system may seem highly innovative in particular niches of academic publishing, it appears rather constant when looking at academic publishing in general.

6.2. Methods

Detailed information on editorial procedures used by journals is surprisingly hard to find. While some journal 'instructions for authors' indicate procedures for blinding author identities or reviewer selection, most journals do not explain the details of their editorial procedures. For our study, we were obliged to gather data through an online survey among editors, which was distributed in the context of a previous study on the effectiveness of peer review forms in preventing article retractions. We will briefly describe our methodology here and refer to the previous publication (Horbach and Halffman 2018a) for further details. Gathering information about peer review through a questionnaire has the advantage of a wide coverage and practicality of data collection, but also a drawback: procedures reported by the editors are not necessarily a perfect reflection of actual practices, which could vary between editors or might present a polished account.

1. Editorial procedure questionnaire

Information about journal editorial procedures was gathered through a short questionnaire among journal editors. The questionnaire consisted of twelve questions, each articulating a specific attribute of editorial practices (Table 5.1), based on a classification of editorial procedures by Horbach and Halffman (2018b). Respondents were asked to match their procedures for each of the editorial attributes with the options outlined in the table. In addition, we asked editors to indicate whether, when and how any of these attributes was modified since 2000, which allows us to trace innovations in editorial procedures since the beginning of this century.

2. The sample

The questionnaire was distributed via email to journal editors. Two strategies were used to gather email addresses. First, we used articles indexed in Web of Science (WoS) as 'editorial material' and extracted the email address of the corresponding author, on the assumption that authors of 'editorial material' would very likely be editors. 58763 unique email addresses of editors from 6245 different journals were collected in this way (about a fifth of all journals listed in WoS). Because of our initial interest in retracted journal articles, we manually amended the list with email addresses from editors of journals issuing at least ten retractions, according to the *RetractionWatch* database of retracted journal articles. This yielded a total of 6331 different journals.

After distributing our questionnaire to these journals and sending out two reminders, we eventually obtained a total of 361 useful responses. The final response rate of 6.12% is low, but comparable to, or even higher, than response rates of similar online surveys among journal editors or authors regarding issues related to academic integrity (Hopp and Hoover 2017; Stitzel et al. 2018). Nevertheless, our sample covers a wide range of research fields and reflects the distribution of journals over fields. For a detailed overview of our sample, we refer to Horbach and Halffman (2018a).

3. Analysis

To analyse the distribution of editorial procedures across journals, we used information about the journals' academic disciplines and publishers. For the information regarding publishers we used the available data in the Web of Science database. Here, we distinguished between the five largest publishers (Elsevier, Springer, Wiley, Taylor & Francis, and Sage) and all other, smaller, publishers (Larivière et al. 2015). When analysing the distribution of editorial procedures over scientific disciplines, we made use of the categorisation of research disciplines in the Leiden Ranking comprising the categories: Social Science & Humanities, Biomedical & Health Science, Physical Sciences & Engineering, Life & Earth Science, or Mathematics & Computer Sciences (Waltman and van Eck 2012; CWTS 2018).

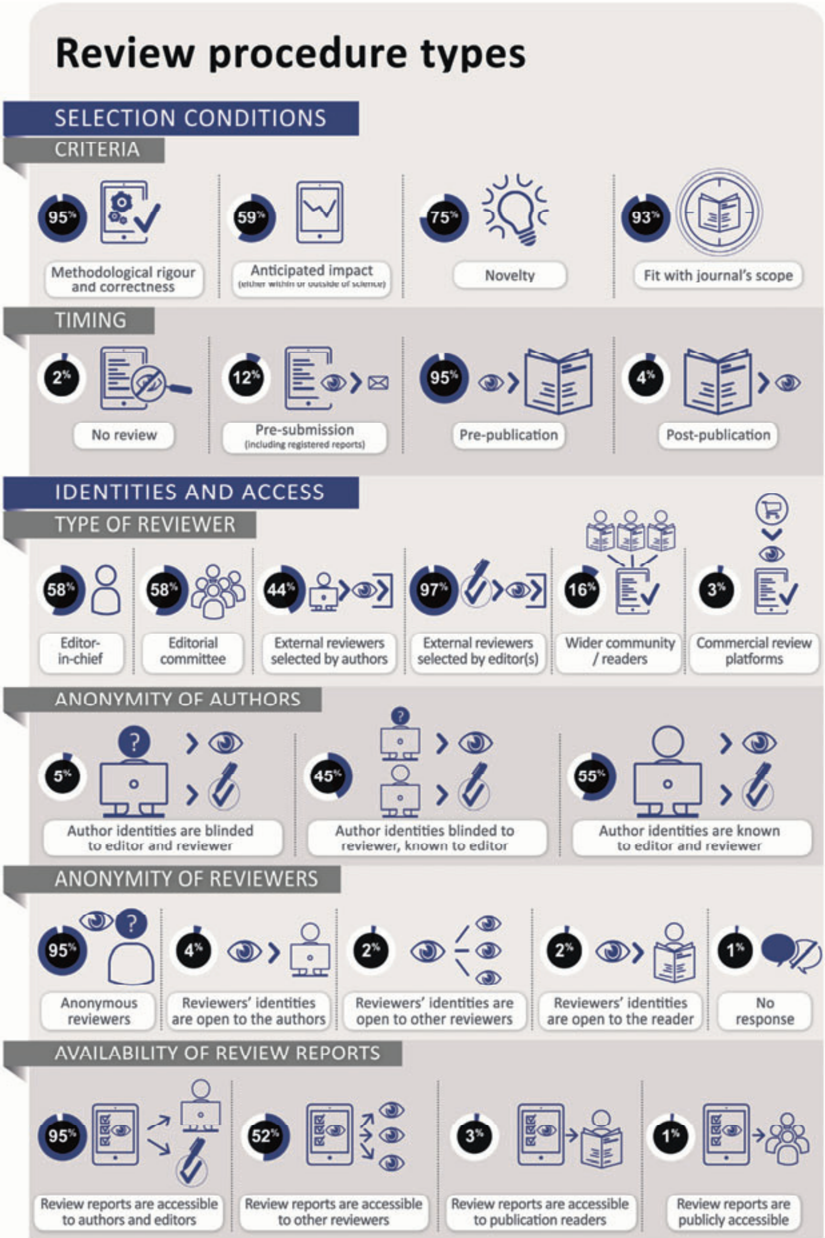
6.3. Results

6.3.1. Current implementation of editorial procedures

Table 6.2 lists all editorial procedures and attributes studied in our research. The table also presents an overview of the current implementation of all twelve editorial attributes according to our data.

The table demonstrates some clear differences in the uptake of editorial procedures, especially indicating that several 'traditional' review procedures are still ubiquitous, such as selection of reviewers by editors (97%), keeping reviewer identities anonymous (94%), or pre-publication review (97%). In contrast, some more recent or innovative procedures are virtually absent, including review in which reviewer identities are made public (2%), review by commercial platforms (1%) and post-publication review (2%).

Even though some editorial procedures are ubiquitous there is no 'standard' model for peer review, nor even a limited set of standard models. The core set of review procedures, used in combination by 75% of all journals, consists of five principles: (i) pre-publication review, (ii) using methodological rigour and correctness as selection criteria, (iii) performed by external reviewers suggested and selected by editor(s), (iv) keeping reviewers anonymous (both to authors and other reviewers, as well as to readers of the published manuscript) and (v) making review reports accessible to authors and editors. However, as soon as we add more characteristics to this set, the commonality between journals quickly drops. Outside of this set, editorial procedures are quite diverse, with journals engaging in review procedures that differ on at least one of the twelve attributes studied. Hence, even though some basic review procedures seem universal, only relatively few journals use all of them and only very few journals perform the editorial process in the exact same way. Given the fact that editorial procedures in a large share of journals are more or less centrally organised through large publishers, this significant heterogeneity in journals' review procedures might be deemed surprising. In the following sections we will look more specifically at the distribution of editorial procedures across scientific disciplines and academic publishers.


Review procedure types
SELECTION CONDITIONS
CRITERIA
95%
Methodological rigour and correctness
59%
Anticipated impact (either within or outside of science)
75%
Novelty
93%
Fit with journal's scope
TIMING
2%
No review
12%
Pre-submission (including registered reports)
95%
Pre-publication
4%
Post-publication
IDENTITIES AND ACCESS
TYPE OF REVIEWER
58%
Editor-in-chief
58%
Editorial committee
44%
External reviewers selected by authors
97%
External reviewers selected by editor(s)
16%
Wider community / readers
3%
Commercial review platforms
ANONYMITY OF AUTHORS
5%
Author identities are blinded to editor and reviewer
45%
Author identities blinded to reviewer, known to editor
55%
Author identities are known to editor and reviewer
ANONYMITY OF REVIEWERS
95%
Anonymous reviewers
4%
Reviewers' identities are open to the authors
2%
Reviewers' identities are open to other reviewers
2%
Reviewers' identities are open to the reader
1%
No response
AVAILABILITY OF REVIEW REPORTS
95%
Review reports are accessible to authors and editors
52%
Review reports are accessible to other reviewers
3%
Review reports are accessible to publication readers
1%
Review reports are publicly accessible

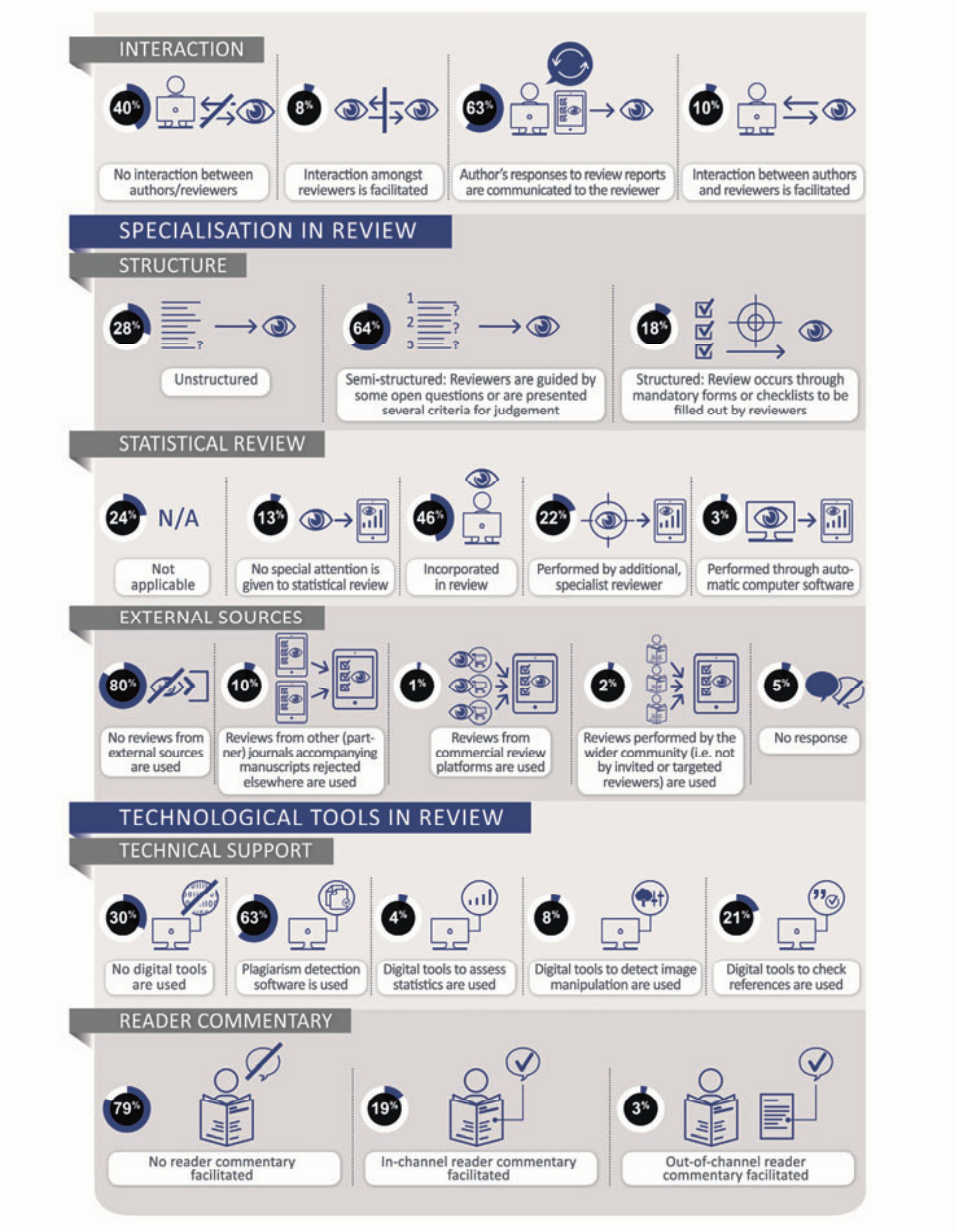


Table 6.2 Distribution of forms of peer review categorized by dimension and attributes. Illustration by Paul de Groot (Dikke Punt).

6.3.2. The distribution of editorial procedures

Research disciplines

Peer review is commonly presented as field-specific, with particular procedures that are common for particular research areas. However, our data suggest rather the opposite. For most of the editorial attributes studied, research fields appear strikingly similar. While journals tend to differ in their editorial procedures in subtle ways, when aggregating over disciplines, these variations dissolve. In fact, only two of the twelve attributes display substantial differences between fields: the level of author anonymity and the form of statistics review.

The former, demonstrated in Figure 6.1, represents a well-known difference between the social sciences and humanities on the one hand, and natural and health sciences on the other. While in SSH journals it is common to blind author identities to reviewers (but not to editors), journals in all other domains more commonly disclose author identities both to editors and reviewers. The biomedical and health sciences demonstrate most diversity with 63% of the journals disclosing author identities, 36% blinding author identities to reviewers only, and 2% blinding author identities both to reviewers and editors. These findings are consistent with a Taylor & Francis survey, in which SSH editors-reported to have used, at some point in time, 86% double-blind and 35% single-blind, while STM editors reported 75% single blind and 42% double blind (Taylor & Francis 2015). Our overall occurrence rate for double-blind procedures resembles that of the Directory of Open Access Journals (48%), but reliable disciplinary break-downs are not provided there (Directory of Open Access Journals 2018).

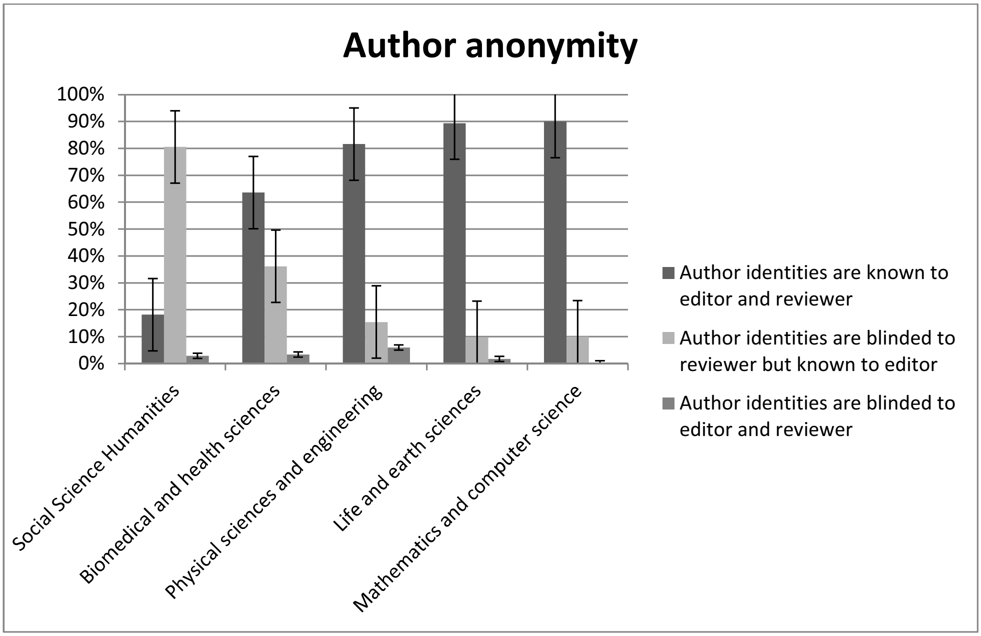


Figure 6.1 Level of author anonymity during review by journals in different research areas

The second major difference between scientific domains consists of how they perform statistics reviews (Figure 6.2). This might be not so surprising given the different importance of statistical analyses in various domains. Most notably, statistical review was deemed to be 'not applicable' to many journals in mathematics and computer sciences, physical sciences, and engineering and social sciences and humanities. In contrast, it is considered relevant for the biomedical and health sciences, as well as for life and earth sciences. In the latter, statistics review is predominantly incorporated in the general review assignment, whereas in the former more than half of the journals report having specialist statistics reviewers to evaluate these aspects of the manuscript.

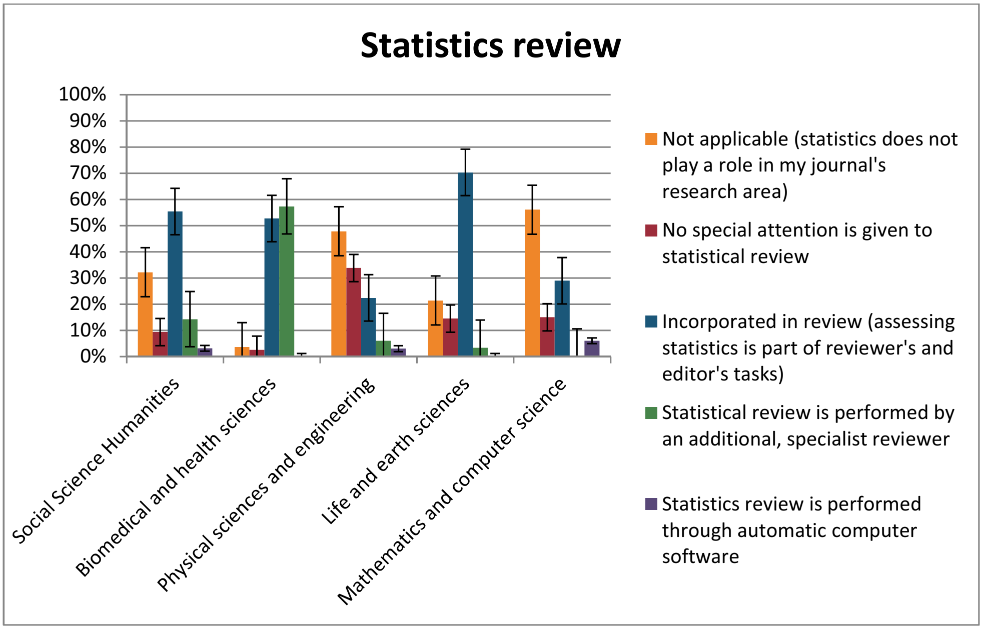


Figure 6.2 Forms of statistics review by journals in different research areas

<u>Publishers</u>

Similar to the distribution of editorial procedures over scientific disciplines, the distribution of procedures over large and small academic publishers is fairly homogeneous. Distinguishing between the five largest publishers (Elsevier, Springer, Wiley, Taylor & Francis, and Sage) and the other, smaller, publishers, we only notice three significant differences in the way their affiliated journals organise their editorial process. Journals affiliated with the large publishers more often communicate author responses to review reports with reviewers (68% vs. 56%), and more often use plagiarism detection software (70% vs. 55%). In contrast, journals affiliated with smaller publishers more often facilitate reader commentary on the journal’s webpage (25% vs. 14%).

Most interestingly, however, the distribution of editorial procedures for all other (forty-seven) attributes does not differ substantially between the largest and the smaller publishers. While some differences are to be expected if only due to chance, this suggests that the heterogeneity in editorial procedures occurs mainly within publishers, rather than across publishers Hence this seems to demonstrate that editors of journals at larger publishers are relatively autonomous in their choice of editorial procedures.. At least no difference can be spotted between the set of most prominent publishers and smaller publishers, including journals run by scientific communities or university presses.

Changes over time

In spite of a constant stream of innovations, editorial procedures in most journals are surprisingly stable. For example, whereas 54.3% of the journals disclosed author identities to reviewers and editors in 2000, 54.6% did so in 2008 and 54.0% in 2018. Similarly, reviewer identities were hidden from authors and other reviewers in 94.7% of the journals in 2000 and in 94.2% of journals in 2018. The vast majority of other procedures studied display very similar patterns.

For all twelve aspects of the editorial process used in our survey, we asked respondents whether any changes had taken place since 2000. Only 169 out of the 361 responding journals (47%) of the journals reported at least one such change and only eleven (3%) reported at least three changes. Hence, the majority of the journals do not report any change, suggesting that their editorial procedures have remained fixed since the beginning of this century. In total, 286 changes were reported, an average of 0.8 changes per journal. The majority of alterations in editorial procedures concerned the introduction of digital tools (most notably text similarity scanners), or changes in review criteria (usually becoming more strict), comprising respectively 39% and 16% of all changes.

Because the number of changes in editorial procedures is so low, hardly any are visible when plotting trends of review procedures for most of the attributes studied. Only for the attribute concerning the usage of digital tools, a clear trend is visible since the year 2000, see Figure 6.3. The figure demonstrates that, especially over the last decade, journals are increasingly adopting text similarity scanners, while the share of journals not using any form of digital support is clearly declining.

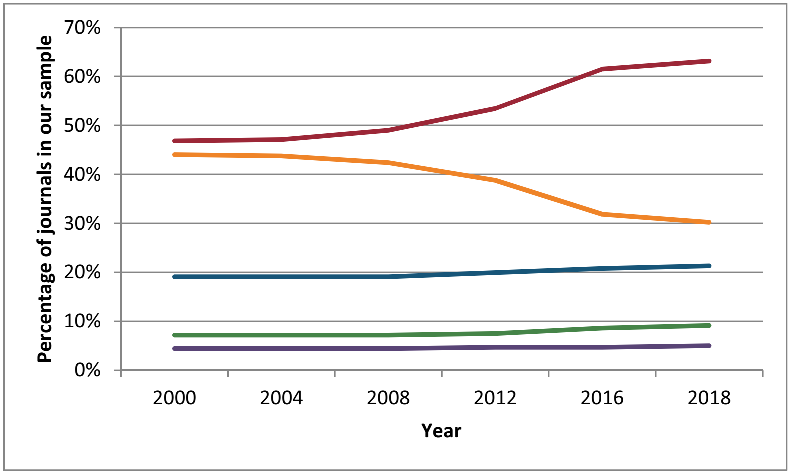


Figure 6.3 The number of journals using a specific form of digital tools: text similarity scanners (red), no tools (orange), tools to check references (blue), tools to detect image manipulation (green), tools to aid in statistics review (purple)

Drawing on the literature about the spread of innovations, certain factors might be expected to drive the implementation of novel editorial procedures. First, we could expect more innovations in journals with high retraction rates, since innovations tend to appear as ways to tackle specific issues, and peer review is increasingly expected to detect fraudulent or erroneous manuscripts. Second, prominent, or highly established journals might be expected to be drivers of change, as they have more resources available, are more centrally positioned in communication networks and their reputation is at stake. Although heavily criticised, the journal impact factor (JIF) remains one of the sole recognised indicators of journal prominence and prestige. Using JIF, one could expect that journals with a higher JIF are more likely to implement novel editorial procedures. However, both factors, retraction rate and JIF, were not significantly correlated with the number of changes in editorial procedures (with r-squared values of 0.003 and 0.00004 respectively). This suggests that both the number of retractions and the JIF neither have a stimulating nor a restricting effect on the implementation of innovative review procedures.

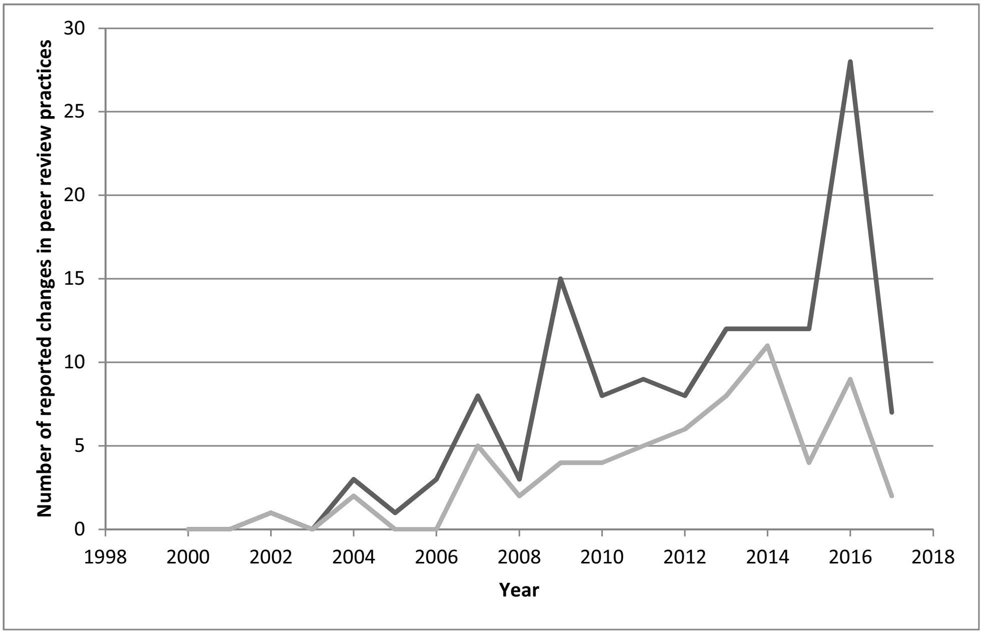


Figure 6.4 The number of reported changes in editorial procedures per year (dark) and the number of changes due to the implementation of plagiarism detection software (light).

Plotting the number of changes for which we have information on the exact date of implementation, we conclude that, even though only few changes occur, the rate of implementation is generally increasing (Figure 6.4). Showing the number of changes due to the implementation of plagiarism detection software, we again conclude that this type of change accounts for the majority of innovations in peer review. It remains to be studied whether the increasing pace of innovation is a general trend, or whether the apparent trend is merely an effect of specific innovations becoming more familiar and more ingrained in communication networks, thereby temporarily lowering the threshold for implementing this specific innovation (Wejnert 2002).

The literature on the spread of innovations distinguishes between smaller or larger collectives of actors implementing an innovation, suggesting that they may be more or less likely to do so in various circumstances. Therefore, we studied the number of changes in editorial procedures in journals of either one of the five largest publishers (Elsevier, Springer, Wiley, Taylor & Francis, and Sage) or any of the other publishers. The results are plotted in Figure 6.5. The figure shows that large publishers contribute slightly more to the number of implemented changes. However, when compensating for the fact that our sample comprises 198 journals from the large publishers and 162 journals from the smaller publishers, this difference in number of implemented changes becomes negligible. In addition, the trends of implementing changes in both larger and smaller publishers are highly similar, suggesting akin underlying mechanisms.

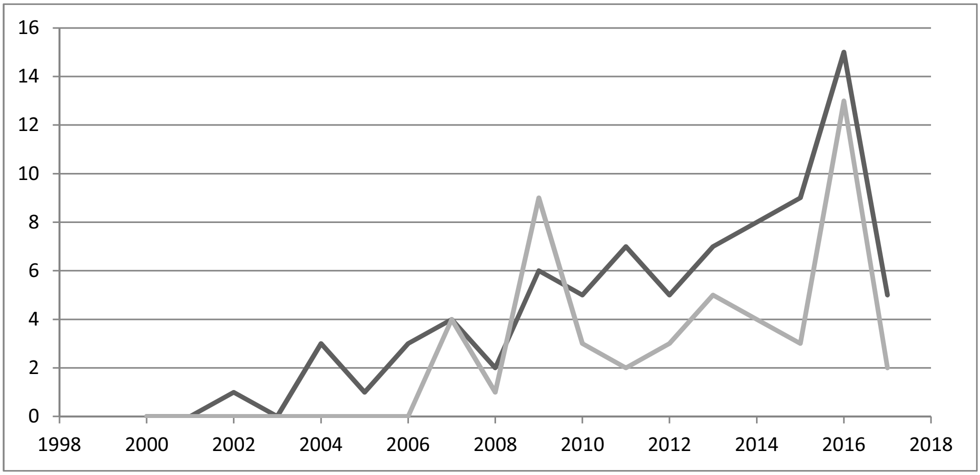


Figure 6.5 The number of reported changes in editorial procedures per year for the five largest publishers (dark) and the other, smaller publishers (light).

6.3.3. Some reflection: reasons for change

Even though it was not a prime aim of our study, our data allows us to get some impression of the reasons why journals alter their editorial procedures. Though not directly invited to do so, a substantial share of the respondents reporting on changes in their editorial procedures, included information about the reason for the change. Out of the 286 reported changes, 61 (21.3%) were provided with information on the reason for change. Even though this data has to be treated with caution, it shows interesting patterns. Most notably, 'the availability of new tools made this possible' was frequently mentioned as a reason to adopt new editorial procedures. It was mentioned in 41% of all cases, not surprisingly, especially when reporting on changes in the use of (text similarity) scanners or support in statistical review. Other reasons frequently presented were the arrival of a new editor-in-chief (15%) or a (new) requirement by the publisher (8%).

Besides these three major reasons, other less frequently occurring motivations for change include 'pressure to increase impact factors', 'increased submission rates' and 'stopped to have access to this service' (e.g. to specialist statistics reviewers). In addition, some journals specifically addressed 'issues with fraud/misconduct' or the intention to 'filter 'bad' science' as reasons to implement different editorial procedures. Notably absent among the list of reported reasons for change was a history of retracted journal articles that 'slipped through' peer review and were later found to be problematic.

This suggests that, by and large, the opportunity to implement editorial innovations (i.e. the availability of and access to new tools, or the new expertise of a novel editor-in-chief) are the main motivators to change. On the contrary, intrinsic arguments to improve peer reviews capabilities or performance are seldom given as motivations for change. Even though our data are to be considered rather exploratory, they do suggest a clear pattern and invoke several questions for future research.

6.3.4. Innovation niches

Our analyses of editorial procedures show a very slow implementation rate. When looking at the editorial process 'from a distance', little seems to be changing. However, despite an apparent stability, some innovations are actually getting a foothold, but only in very specific niches and particular contexts of the publication system, a phenomenon which is extensively described in innovations studies (e.g. A. Smith and Raven 2012). In the following, we will provide short descriptions of four niches in which particular innovations are getting established. This will allow for reflection on the circumstances in which innovations might be more widely implemented.

Text similarity scanners

The only innovation for which we observe substantial implementation are text similarity scanners, with significant increase in usage over the past decade. Combining different pieces of data from our study, a nuanced picture emerges about the reasons for their unique success.

First, text similarity scanners promise a simple fix for the rather uncontested issue of plagiarism and problematic text recycling. Unlike many of the other review procedures, these scanners promise a guaranteed solution to a specific problem, much more so than blinding author or reviewer identities, for instance. Hence, the expectations are clear, allowing for a relatively smooth translation of expectations into requirements for the tools (Van Lente 1993).

Second, journals and publishers have a major (commercial) stake in providing or promising duplication-free manuscripts. It allows them to sell a 'unique' product. Especially the larger, commercial publishers may be interested in this, in line with our finding that the use of text similarity scanners is one of the few examples distinguishing the larger from the smaller publishers.

Third, similarity scanners are not only used in the publishing industry, but also in higher education, scanning student papers for plagiarism. In fact, many of the developers of such scanners consider this their primary market. For editors and publishers, the usage of these scanners in higher education provides a testbed allowing them to see whether the scanners live up to expectations. Since many editors also have a role as lecturer, this allows them to get familiar with these tools via multiple communication networks.

Registered reports in health and psychology journals

A second example of an innovation that finds substantial implementation, though only in a particular niche, are the registered reports, in which research is evaluated only based on its rationale and methodology, usually before data gathering has started. Currently, this review model has been implemented in a substantial amount of psychology journals, as well as some journals in the health sciences (Center for Open Science 2018). Similar to text similarity scanners, registered reports were established with a fairly specific aim. They aim to address the alleged replication crisis, and promise to provide a more or less simple fix by facilitating the publication of negative results (combating publication bias) and making replication studies more attractive (Nosek and Lakens 2014; Horbach and Halffman 2018b). In addition, the registered report model is highly similar to the review model used in grant applications, which is also solely based on a study's a priori rationale and methodology. Hence, akin to the text similarity scanners, actors might become familiar with registered reports through various communication channels, thus making the innovation more familiar.

Even though concerns about the 'replication crisis' in science currently seem to be spreading, they originated and still mainly seem to affect the medical science and (social) psychology (Wicherts 2017; Begley and Ioannidis 2015). Hence, the implementation of registered reports seems to be constrained to the area for which it provides a solution to an acknowledged and well-defined problem. In addition, the registered report format seems to be most applicable to certain areas of research (including the empirical, highly standardised fields, with low levels of researchers' degrees of freedom), while it is less applicable in fields with other methodological and epistemic traditions (such as the humanities).

Image manipulation scanners in biomedical journals

A third editorial innovation that we would like to single out comprises the use of image manipulation scanners. At present, they seem to be most commonly used in biomedical fields and, to a lesser extent, some journals in psychology (Scheman and Bennett 2017). Within these fields, they again provide a solution to an uncontested issue, being the manipulation of figures and images, such as western blots. While detecting image tweaking is still technically challenging, highly standardised representations such as western blots allow for some automated detection, or at least flagging of potential problems. Even though some prominent cases of fraud where detected through careful scanning of images and figures, including the

Schön case (Consoli 2006), such detection as yet relies on human skill. While techniques based on Artificial Intelligence promise to take this approach to a more automated level, such expectations remain to be fulfilled (BioMed Central 2017). Currently, the use of image manipulation scanners therefore seems to be constrained to (i) fields in which images commonly occur in manuscripts; and (ii) those fields that have highly standardised representations in images and figures, thereby allowing relatively simple technical tools to be of genuine assistance.

Open review at several publishers

The last peer review innovation implemented in specific niches is the open review model. Several publishers have now adopted this model, with some, such as BioMed Central and the British Medical Journal, launching a range of new journals adopting open review (Godlee 2002). This review procedure aligns with the more general call for opening up science and adhering to open science practices, including publishing open access, sharing data, and other forms of transparency in research (Nosek et al. 2015). Despite wide calls to follow these standards, our data show that implementation of the open review model is still rather modest and mainly confined to several individual publishers. Part of this may be due to the large variety of different forms of 'open review', a term that may encompass either the disclosure of reviewers' identities to the authors of a submitted manuscript, the disclosure of such identities to the wider public, or even the publication of entire review reports (Ross-Hellauer 2017). In fact, Ross-Hellauer (2017) found at least twenty-two different definitions of 'open peer review', showing that the phrase is currently highly ambiguous and has not yet settled into a single set of features or schema for implementation. This lack of uniformity may cause a serious obstacle for editors or publishers willing to implement some form of open review in their journals.

6.4. Conclusion

This study has been one of the first attempts to map the distribution and development of journal peer review and editorial procedures. Our work presents new perspectives on multiple aspects of review. First, it shows that editorial procedures are diverse. The 'common core' of editorial procedures that are shared by a wide variety of journals comprises a surprisingly small set of procedures. Journals commonly differ in their review procedures in small or subtle ways.

Second, we witness only minor variations in editorial procedures when aggregating over either scientific disciplines or academic publishers. Hence, while individual journals commonly differ slightly in their editorial procedures, on a larger scale, there are few systematic patterns in these differences.

Third, over the past decades, an abundance of new review procedures have been suggested and initiated. However, adoption of these innovative formats by other journals is slow. Since the beginning of the century, only a very limited number of journals have made substantial changes to their editorial process. Even today, the traditional forms of peer review, single- or double-blind pre-publication review, still prevail over more innovative formats, such as open, post-publication, or review assisted by the wider community and digital tools. Text similarity scanners are the exception to this pattern. In the past decade, their uptake has rapidly increased and using such scanners has now become more or less common practice.

Fourth, we obtained some data about journals' motivations to alter their editorial procedures. Despite the exploratory nature of this data, it suggests that innovations most commonly occur as a response to novel opportunities, rather than as a response to shifting expectations or general threats.

Last, we sketch out several niches in which particular innovations have found their way to implementation. Although these are, with the exception of text similarity scanners, not or hardly visible in the overall figures on the distribution of editorial models, the implementation of innovations in particular niches may provide an important step towards further distribution (e.g. Verbong et al. 2008; A. Smith and Raven 2012). A closer look at these niches portrays several contexts in which implementation becomes more likely: contexts in which an innovation offers a fairly simple solution to an uncontested problem, or contexts in which actors may acquire prior familiarity with the innovation. In contrast, ambiguity about the specific features of an innovation as well as technical limitations and epistemic or methodological diversity seem to hinder wide spread implementation.

Hence, we conclude that there are various reasons for, for example, text similarity scanners to be a special case in our sample of review procedures. Rather than spreading by the logic of a diffusion model, there are more subtle reasons for these tools to be specifically prone to implementation at a wide variety of journals. For the same reasons, a similar process might be expected to occur for other tools and review models in the future, including tools supporting statistics review and image manipulation scanners: They also promise a relatively simple fix for an uncontested issue as well as that they might be introduced and create familiarity in multiple contexts.

6.5. Discussion

Our study is the first to analyse the implementation of a wide variety of editorial formats over a wide variety of journals. Earlier work has commonly focussed on a single aspect of review (for instance mapping single-blind vs. double-blind procedures) or focussed on a specific research discipline or publisher (e.g. Taylor & Francis 2015).

Our study's findings may be somewhat limited by a number of factors. First, it is important to bear in mind the possible bias in the selection of and response by our study sample. When sampling email addresses of journal editors, we used the Web of Science database, searching for editorials written in 2017. This thus excludes journals not indexed by Web of Science, or those journals that do not publish editorials. In particular, this may have excluded several, young, non-English, or niche journals. Especially some of the young or niche journals might be particularly innovative in their editorial process, hence the selection bias may have caused us to underestimate the diversity of editorial procedures. In particular, it may have led us to overlook the implementation of editorial procedures within a specific community that is underrepresented within our sample. For example, the apparent increase in the usage of registered reports in psychology journals (Center for Open Science 2018) is a trend not visible in our data.

Conversely, from the sample of journals that were sent an invitation to participate, those journals paying specific attention to their editorial process, as well as those being particularly keen on innovating editorial procedures are arguably more likely to have responded to our survey. Hence, this potential response bias may in fact have resulted in an overestimation of diversity and innovation in editorial procedures. In fact, our data provides some hint to this phenomenon, with substantially more innovations in peer review reported by those journals that responded most quickly to our survey.

A final limitation of our study rests in the survey approach to collect data about editorial procedures. Even though we tested our survey before distributing it, this type of data collection is inherently prone to misunderstanding of the wording, as well as incomplete knowledge of the editors of (past) review procedures, or different interpretation of terms by researchers and respondents. This might have led editors to classify their journal's editorial procedures differently from how we intended it and hence have influenced our analyses. Specifically, editors' incomplete knowledge of past changes in review procedures might have influenced our analysis of the implementation of new editorial models, especially with implementations established further in the past. In addition, we acknowledge that we have studied formal peer review procedures, rather than actual review practices. How editors say their review is performed, might not for all manuscripts correspond to how things are done in practice.

Our findings raise several questions to be addressed in future research. Both the apparent lack in diversity of editorial procedures across disciplines and publishers as well as the slow implementation of novel procedures raise questions about the seeming inertia in review and factors hindering innovation. While the literature on the spread of innovations suggests that the academic publishing industry meets many of the conditions required for substantial adoption of innovations, we witness little of this in practice. In addition, the exploratory findings about reasons for journals to adopt new review procedures raise further questions to be explored, including about which actors are in the position to implement new review formats.

On what do they base their decisions regarding innovation? And more generally, do journals deliberately adopt a strategy of cautious innovation, or is the peer review system merely a sleepy giant?

In addition, future research could examine actual review practices, rather than formal editorial procedures as reported by editors or outlined on journal web pages. These questions are to be tackled by more qualitative research involving the various stakeholders in the editorial process, including editorial boards, managing editors, publishers, authors and reviewers.

Last, an open question to be addressed in future research concerns the effects of various procedures: do they deliver on the expectations? For instance, are specialist statistics reviewers or statistics scanners indeed effective in increasing the quality of statistical analyses? And do text similarity scanners indeed decrease the amount of (textual) duplication in journal manuscripts? While some work on this has been done (Horbach and Halffman 2018a), much remains to be elucidated. In the end, such information should be the leading motivation for journals to adopt specific review procedures.

6.6. References

Baldwin, Melinda. 2015. Credibility, peer review, and Nature, 1945–1990. *Notes and Records: the Royal Society journal of the history of science* 69 (3):337-352. doi:10.1098/rsnr.2015.0029.

Baldwin, Melinda. 2018. Scientific Autonomy, Public Accountability, and the Rise of "Peer Review" in the Cold War United States. *Isis* 109 (3):538-558. doi:10.1086/700070.

Begley, C Glenn, and John PA Ioannidis. 2015. Reproducibility in science. *Circulation research* 116 (1):116-126.

Biagioli, Mario. 2002. From book censorship to academic peer review. *Emergences: Journal for the Study of Media & Composite Cultures* 12 (1):11-45.

Bijker, Wiebe E, Thomas Parke Hughes, and Trevor J Pinch. 1987. *The social construction of technological systems: New directions in the sociology and history of technology*. Cambridge, Massachusetts: MIT press.

BioMed Central. 2017. What might peer review look like in 2030?, ed. BioMed Central. figshare.

Bohlin, Ingemar. 2004. Communication Regimes in Competition. *Social Studies of Science* 34 (3):365-391. doi:doi:10.1177/0306312704041522.

Brown, Nik, and Mike Michael. 2003. A Sociology of Expectations: Retrospecting Prospects and Prospecting Retrospects. *Technology Analysis & Strategic Management* 15 (1):3-18. doi:10.1080/0953732032000046024.

Burnham, J. C. 1990. The evolution of editorial peer review. *Jama* 263 (10):1323-1329. doi:10.1001/jama.1990.03440100023003.

Center for Open Science. 2018. Registered Reports: Peer review before results are known to align scientific values and practices - Participating Journals. https://cos.io/rr/. Accessed October 4 2018.

cOAlition S. 2018. Accelerating the transition to full and immediate Open Access to scientific publications. https://www.coalition-s.org/about/. Accessed.
Consoli, Luca. 2006. Scientific misconduct and science ethics: A case study based approach. *Science and Engineering Ethics* 12 (3):533-541.
Csiszar, Alex. 2016. Peer Review: Troubled from the start. *Nature*.
Csiszar, Alex. 2018. *The scientific journal: Authorship and the politics of knowledge in the nineteenth century*. University of Chicago Press.
CWTS. 2018. CWTS Leiden Ranking - Fields. http://www.leidenranking.com/information/fields. Accessed.
Directory of Open Access Journals. 2018. Search. https://doaj.org/search#.W74NsPkzaUk. Accessed October 10 2018.
Ercegovac, Zorana, and John V Richardson. 2004. Academic dishonesty, plagiarism included, in the digital age: A literature review. *College & Research Libraries* 65 (4):301-318.
Franke, Nikolaus, and Sonali Shah. 2003. How communities support innovative activities: an exploration of assistance and sharing among end-users. *Research Policy* 32 (1):157-178. doi:https://doi.org/10.1016/S0048-7333(02)00006-9.
Fyfe, Aileen, Kelly Coate, Stephen Curry, Stuart Lawson, Noah Moxham, and Camilla Mork Rostvik. 2017. Untangling academic publishing: a history of the relationship between commercial interests, academic prestige and the circulation of research.
Fyfe, Aileen, Julie McDougall-Waters, and Noah Moxham. 2015. 350 years of scientific periodicals. *Notes and Records: the Royal Society journal of the history of science* 69 (3):227-239. doi:10.1098/rsnr.2015.0036.
Fyfe, Aileen, Flaminio Squazzoni, Didier Torny, and Pierpaolo Dondio. In Press. Managing the Growth of Peer Review at the Royal Society Journals, 1865-1965. *Science, Technology, & Human Values* 0 (0):0162243919862868. doi:10.1177/0162243919862868.
Godlee, Fiona. 2002. Making reviewers visible: openness, accountability, and credit. *Jama* 287 (21):2762-2765.
Greenhalgh, T., G. Robert, F. Macfarlane, P. Bate, and O. Kyriakidou. 2004. Diffusion of innovations in service organizations: Systematic review and recommendations. *Milbank Quarterly* 82 (4):581-629. doi:10.1111/j.0887-378X.2004.00325.x.
Hames, Irene. 2014. The changing face of peer review. *Sci Ed* 1 (1):9-12. doi:10.6087/kcse.2014.1.9.
Hopp, Christian, and Gary A. Hoover. 2017. How prevalent is academic misconduct in management research? *Journal of Business Research* 80:73-81. doi:https://doi.org/10.1016/j.jbusres.2017.07.003.
Horbach, Serge P. J. M., and Willem Halffman. 2018a. The ability of different peer review procedures to flag problematic publications. *Scientometrics*. doi:10.1007/s11192-018-2969-2.
Horbach, Serge P. J. M., and Willem Halffman. 2018b. The changing forms and expectations of peer review. *Research Integrity and Peer Review* 3 (1):8. doi:10.1186/s41073-018-0051-5.
Knoepfler, Paul. 2015. Reviewing post-publication peer review. *Trends in Genetics* 31 (5):221-223.

LaFollette, Marcel Chotkowski. 1992. *Stealing into print: fraud, plagiarism, and misconduct in scientific publishing*. Univ of California Press.

Larivière, Vincent, Stefanie Haustein, and Philippe Mongeon. 2015. The Oligopoly of Academic Publishers in the Digital Era. *Plos One* 10 (6):e0127502. doi:10.1371/journal.pone.0127502.

Lee, C. J., C. R. Sugimoto, G. Zhang, and B. Cronin. 2013. Bias in peer review. *Journal of the American Society for Information Science and Technology* 64 (1):2-17. doi:10.1002/asi.22784.

MacKenzie, Donald, and Judy Wajcman. 1999. *The social shaping of technology*. vol. 2nd: Open university press.

Moxham, Noah, and Aileen Fyfe. 2017. The Royal Society and the prehistory of peer review, 1665–1965. *The Historical Journal*:1-27. doi:10.1017/S0018246X17000334.

Nguyen, V. M., N. R. Haddaway, L. F. G. Gutowsky, A. D. M. Wilson, A. J. Gallagher, M. R. Donaldson, N. Hammerschlag, and S. J. Cooke. 2015. How Long Is Too Long in Contemporary Peer Review? Perspectives from Authors Publishing in Conservation Biology Journals. *Plos One* 10 (8):20. doi:10.1371/journal.pone.0132557.

Nosek, Brian A., G. Alter, G. C. Banks, D. Borsboom, S. D. Bowman, S. J. Breckler, S. Buck et al. 2015. Promoting an open research culture. *Science* 348 (6242):1422-1425. doi:10.1126/science.aab2374.

Nosek, Brian A., and Daniël Lakens. 2014. Registered reports: A method to increase the credibility of published results. *Social Psychology* 45 (3):137-141. doi:10.1027/1864-9335/a000192.

Okike, K., K. T. Hug, M. S. Kocher, and S. S. Leopold. 2016. Single-blind vs double-blind peer review in the setting of author prestige. *Jama* 316 (12):1315-1316. doi:10.1001/jama.2016.11014.

Oudshoorn, N. E. J. , and T. Pinch. 2007. User-technology relationships: Some Recent Developments. In *Handbook for Social Studies of Science*, eds. E. J. Hackett, O. Amsterdamska, M. Lycnh, and J. Wajcman, 541 - 567. London: MIT Press.

Patel, J. 2014. Why training and specialization is needed for peer review: a case study of peer review for randomized controlled trials. *Bmc Medicine* 12:7. doi:10.1186/s12916-014-0128-z.

Peres, R., E. Muller, and V. Mahajan. 2010. Innovation diffusion and new product growth models: A critical review and research directions. *International Journal of Research in Marketing* 27 (2):91-106. doi:10.1016/j.ijresmar.2009.12.012.

Peters, Douglas P., and Stephen J. Ceci. 1982. Peer-review practices of psychological journals: The fate of published articles, submitted again. *Behavioral and Brain Sciences* 5 (2):187-195. doi:10.1017/S0140525X00011183.

Pontille, David, and Didier Torny. 2014. The Blind Shall See! The Question of Anonymity in Journal Peer Review. *Ada: A Journal of Gender, New Media, and Technology* 4. doi:doi:10.7264/N3542KVW.

Research Square. 2017. Rubriq. https://www.rubriq.com/. Accessed.

Rojas, Fabio. 2007. A modest proposal: triple blind review. Orgtheory.net. https://orgtheory.wordpress.com/2007/01/23/a-modest-proposal-triple-blind-review/. Accessed January 23, 2007.

Ross-Hellauer, T. 2017. *What is open peer review? A systematic review [version 2; referees: 1 approved, 3 approved with reservations]*. vol. 588.
Scheman, Rita, and Christina N. Bennett. 2017. Assessing the Outcomes of Introducing a Digital Image Quality Control Review Into the Publication Process for Research Articles in Physiology Journals. In *International Congress on Peer Review and Scientific Publication*. Chicago.
Smith, Adrian, and Rob Raven. 2012. What is protective space? Reconsidering niches in transitions to sustainability. *Research Policy* 41 (6):1025-1036. doi:https://doi.org/10.1016/j.respol.2011.12.012.
Smith, Richard. 2006. Peer review: a flawed process at the heart of science and journals. *Journal of the Royal Society of Medicine* 99 (4):178-182.
Smith, Richard. 2010. Classical peer review: an empty gun. *Breast Cancer Research* 12:4. doi:10.1186/bcr2742.
Stitzel, Brandli, Gary A. Hoover, and William Clark. 2018. More on Plagiarism in the Social Sciences*. *Social Science Quarterly*:n/a-n/a. doi:10.1111/ssqu.12481.
Taylor & Francis. 2015. "Peer Review in 2015: A global view", A white paper by Taylor & Francis. https://authorservices.taylorandfrancis.com/peer-review-global-view. Accessed.
Tennant, Jonathan P, Jonathan M Dugan, Daniel Graziotin, Damien C Jacques, François Waldner, Daniel Mietchen, Yehia Elkhatib, Lauren B Collister, Christina K Pikas, and Tom Crick. 2017. A multi-disciplinary perspective on emergent and future innovations in peer review. *F1000Research* 6 (1151).
Tennant, Jonathan P, T Ross-Hellauer, Lisa Matthias, Andy Byers, and Jonathan M Dugan. 2018. Thinking Outside the Black Box of Peer Review. Paper presented at the International Conference on Peer Review, Rome,
Thelwall, M., S. Haustein, V. Lariviere, and C. R. Sugimoto. 2013. Do Altmetrics Work? Twitter and Ten Other Social Web Services. *Plos One* 8 (5). doi:10.1371/journal.pone.0064841.
Van Lente, Harro. 1993. Promising technology. The dynamics of expectations in technological developments. Twente University, Delft.
Verbong, Geert, Frank W. Geels, and Rob Raven. 2008. Multi-niche analysis of dynamics and policies in Dutch renewable energy innovation journeys (1970–2006): hype-cycles, closed networks and technology-focused learning. *Technology Analysis & Strategic Management* 20 (5):555-573. doi:10.1080/09537320802292719.
von Hippel, Eric. 1976. The dominant role of users in the scientific instrument innovation process. *Research Policy* 5 (3):212-239. doi:https://doi.org/10.1016/0048-7333(76)90028-7.
Walker, Richard, and Pascal Rocha da Silva. 2015. Emerging trends in peer review—a survey. *Frontiers in Neuroscience* 9 (169). doi:10.3389/fnins.2015.00169.
Waltman, L., and N. J. van Eck. 2012. A new methodology for constructing a publication-level classification system of science. *Journal of the American Society for Information Science and Technology* 63 (12):2378-2392. doi:10.1002/asi.22748.
Wejnert, B. 2002. Integrating models of diffusion of innovations: A conceptual framework. *Annual review of sociology* 28:297-326. doi:10.1146/annurev.soc.28.110601.141051.
Wicherts, J. M. 2017. The Weak Spots in Contemporary Science (and How to Fix Them). *Animals (Basel)* 7 (12). doi:10.3390/ani7120090.

Wisdom, Jennifer P., Ka Ho Brian Chor, Kimberly E. Hoagwood, and Sarah M. Horwitz. 2014. Innovation Adoption: A Review of Theories and Constructs. *Administration and policy in mental health* 41 (4):480-502. doi:10.1007/s10488-013-0486-4.

Zuckerman, Harriet, and Robert K. Merton. 1971. Patterns of evaluation in science: Institutionalisation, structure and functions of the referee system. *Minerva* 9 (1):66-100. doi:10.1007/bf01553188.

Chapter 7 - Innovating the editorial process: An ethnography of publishing

This chapter is under review as:
Horbach, S. P. J. M., & Halffman, W. (Under review).
Innovating the editorial process: An ethnography of publishing.
Science, Technology & Human Values.

7

7.1. Introduction

Academic peer review plays a crucial role in many of research's core processes, including grant and career reviews, but especially in the editorial assessment of papers by research journals. The journal peer review system and the editorial process in which it is embedded are gatekeepers for the dissemination of research findings, act as a self-regulating mechanism, and, by acting as a selection mechanism, play a key role in the academic reward system (Guston, 2007; Fyfe et al., 2017; Zuckerman and Merton, 1971).

Following a series of scandals and controversies over the ability of peer review to guard research quality or integrity, several innovative peer review procedures and supporting technologies have been proposed by a host of enthusiastic innovators. These include the introduction of various software tools, such as text similarity or statistics scanners; procedures of blinding or disclosing actors' identities; shifting timing of peer review in the publication process; and new criteria for accepting or rejecting manuscripts. Each of these innovations has been motivated by specific concerns over flaws in the dominant approaches to peer review. Some of the suggested solutions even move in opposite directions, such as increased anonymity versus increased openness (Horbach and Halffman, 2018).

Despite the host of suggestions and innovations that promise to improve the peer review system, adoption of these new initiatives seems slow. On a global scale, peer review procedures are rather stable and traditional ways of doing peer review still prevail, despite evidence of flaws in old practices and proposed advantages of new ones (Bravo et al., 2019; Smith, 2006; Peters and Ceci, 1982). Implementation of novel review procedures seems to be restricted to specific niches (specialties, publishing platforms), with the exception of the implementation of text similarity software or 'plagiarism scanners' (Horbach and Halffman, submitted). If some of these innovations are indeed so much better than their predecessors, it may seem strange that new review practices do not convince a wider set of journals.

This raises questions about the wider conditions for peer review change and about the considerations that inform decisions to innovate. Many of the newly suggested procedures claim to improve the quality of published research or the fairness of the review process, but these are not the only considerations informing journal policies. Claims about how novel review formats improve quality, transparency or scrutiny of the published record – and ultimately might benefit research in its endeavour to produce and disseminate knowledge – are omnipresent in discussions on peer review (e.g. Godlee, 2002; Bravo et al., 2019). The advocates of peer review innovations assume that these are the features that will carry forward a transformation. While the discourse about these innovations is thus inspired by strong academic values, it does not account for the current practical conditions of running a research journal. From the perspective of fairness and quality alone, it remains unclear how other considerations, such as publishers' motives or other stakeholders' perspectives, may affect decisions to innovate the editorial process. In contrast, the perspectives and business considerations of publishers have been highlighted in discussions on publication models such as open access publishing, or the establishment of hybrid journals (Hansoti et al., 2016; Resnick, 2019). However, we show that such considerations extend much further and reach into the very heart of the editorial and peer review process.

From the perspective of a Science and Technology Studies (STS) understanding of innovations, it is not so strange that innovations fail to convince users by arguments of superior 'quality' alone. Apart from their meaning and specific performance, innovations require integration in existing practices and wider socio-technical configurations, with active involvement of users in light of practical concerns and relations to other actors. In the case of journal peer review, this comprises several user perspectives, including opinions about varying peer review procedures in research communities, the willingness of authors and reviewers to participate in innovative formats, but especially how peer review innovations relate to existing editorial practices and policies. Hence some of the currently proposed peer review innovations may be more than just marginal improvements, and are rather suggestions that require transformative change, affecting not only peer review, but also publishing strategies and economics. Understanding the appeal of review innovations therefore requires comprehension of the editorial practices in which they are to land.

In this study, we set out to research editorial practices to reveal how these might benefit from peer review innovation, and to describe the considerations that inform such decisions. We carried out ethnographic research at the editorial office of two large academic publishers to study the editorial process in close detail and better understand the day-to-day practices and concerns of their employees at all levels. In addition, we interviewed editors of journals that are not closely related to large publishers. Specifically, we were interested in understanding the considerations that inform editorial transformation, guided by questions such as: what does the process of transformation look like? Who makes decisions about such changes? And on what basis are changes made? We hence mainly focus on intended, deliberately planned instances of transformation.

We first describe our methodology and the editorial practices we have observed. Then we analyse the process of transformation and reasons to integrate or ignore peer review innovations. In a subsequent section, we situate the discussion about editorial changes in the currently fast changing academic publishing landscape, fuelled by discussions on Open Science, the added value of commercial publishers and the rise of alternative publication models such as pre-print servers. Ultimately, we conclude that the different meanings attached to editorial practices by those involved, are a prime explanation of why some practice transformations are more successful than others.

7.2. Theoretical framework

In this chapter, we study the practices involved in the editorial and peer review process of large, commercial, academic publishers. Following Schatzki (1996) and Reckwitz (2002), we understand practices simply as 'a routinized type of behaviour' and a temporally and spatially dispersed nexus of doings and sayings. We will understand individuals as the *carriers* of practices and hence 'know-how, meanings and purposes' are not taken to be personal attributes, but rather 'elements and qualities of a practices in which the single individual participates' (Reckwitz, 2002).

Following Shove et al. (2012), we will distinguish three elements of a practice: materials, competence and meaning. There is now broad agreement that things (material objects) should be treated as elements of practice (Shove et al., 2012). Competence refers to the know-how, background knowledge and understanding required to either perform or evaluate a performance of a practice (though there might be a difference between the skills required to do both tasks). Meaning refers to the social and symbolic significance of participation in a practice at any one moment (Shove et al., 2012).

In the setting of the peer review and editorial process, the material elements of a practice include manuscripts, the email and the electronic submission system, and digital tools such as plagiarism scanners or statistics scanners. The competences involved comprise an academic knowledge level expressed as an academic degree, expertise in the relevant subject area, know-how of the review and editorial procedures, familiarity of the editor with the reviewer, English language skills, legal expertise or ethical background knowledge. Last, the meaning constituting the practice may involve a sense of academic duty, a willingness to improve research, an appreciation for keeping up with the literature, a desire to stay anonymous, a desire to make money or create business value, a lack of time, or commitment to a journal, a research field, or a company.

Unpacking the notion of practice a little further, we may consider peer review and editorial practices as an example of what Schatzki calls complex 'integrative' practices (Schatzki, 1996), since they embrace 'a set of hierarchically organized doings, sayings, tasks and projects'. Shove et al. (2012) use a somewhat different notion and speak of 'complexes of practices' of which peer reviewing may constitute an example (embracing the more mundane practices of reading, writing, judging, emailing, etc.). Complexes of practices are cases in which practices come to depend upon each other, either in terms of sequence, synchronization, proximity or necessary co-existence. In such cases, emergent characteristics of the complex of practice cannot be reduced to the individual practices of which it is composed.

7.2.1. Transforming practices

An important aspect of our study is the analysis of how editorial practices may be transformed or, on the contrary, what keeps them stable. In previous work on transformation of practices, authors have distinguished between gradual and radical transformation of a practice. In the former, learning, carrying and sharing may lead to the capture, commitment and change of some practices and practitioners: the processes are transformative both of the practitioners involved and of the practices they reproduce. In the latter, practices 'die' and new practices are born as 'changes in organization are vast or wholesale, or a practice's projects and tasks are simply no longer carried out' (Schatzki, 2002). One of the explanations for the extinction of old (bundles or complexes of) practices is that they had too little internal rewards and were hence not valued for their own sake, but rather as an instrument to obtain something else (MacIntyre, 1985). Other explanations refer to a lack of symbolic or normative anchoring as well as a lack of connection with and dependence on other practices. In the current publishing landscape, with large scale shifts in publication models triggered by open science and open access initiatives (cOALition S, 2018), both gradual and more radical transformation in the editorial process may be expected, we will hence focus on both in our analysis.

In short, practices die out when links between their constituting elements are no longer reproduced. Similarly, bundles or complexes of practices discontinue when one of the practices constituting them disappears. This can either happen through materials not being available anymore (or changes in materials, such as modifications in the electronic submission system), competences disappearing, or shifting meanings. Our study will examine which factors have most impact on shifting, disappearing and evolving editorial practices.

Besides focussing on what initiates transformation of practices, a fruitful lens may be to look at what keeps practices constant and facilitates reproduction over time. One of the factors important in keeping practices stable comprises the infrastructures in which practices are embedded (Rinkinen et al., 2017). These infrastructures allow for routinized actions and maintain the links between different (elements of) practices, thereby keeping them stable over space and time (de Wit et al., 2002). Monitoring and feedback play another essential role in the maintenance of practices. It helps them develop, adapt or stay constant over time, as well as help them travel through space and time. In this, it is useful to distinguish between forms of monitoring and feedback that link one instance of performance to the next, and those implicated in the unfolding careers of practices-as-entities. However, both forms critically interact and often connect. This may happen in at least three ways (Shove et al., 2012):

- When the careers of individuals and practices intersect, monitoring may reveal important signs of progress and hence encourage further effort and investment of time and energy in future performances of a practice.
- Methods of measurement may end up changing the performances and the practices they are designed to monitor. Ample examples of this phenomenon have been described in the literature (Campbell, 1979).
- Systems of classification and standards constitute 'invisible mediators of action' (Bowker and Star, 2000). By setting these standards, templates are established by which performances are compared and which define what one enactment is a performance of.

Thus, technologies and instruments of feedback are of concrete relevance to establish and maintain circuits of reproduction, which in turn are of direct consequence for the survival and transformation of relations between practices and of practices (and their constituting elements) themselves.

7.3. Methods

Our findings are based on ethnographic field visits (by SH) to two large, commercial academic publishers. In total, the ethnographic material consists of notes gathered during 41 interviews or individual meetings, and 10 group meetings. With our work we follow the tradition of ethnographic fieldwork at publishers or publishing related organisations, for example by Hirschauer (2010) and Jacob (2019). In addition, we conducted three interviews with (managing) editors of journals not closely related to such large publishers, mainly for contrast

The first fieldwork visit took place at the editorial office of publisher A. At this publisher, we mainly focussed on a set of open access journals and the team managing and working for this set. The second fieldwork period involved the editorial offices of publisher B. Both publishers have a portfolio of hundreds of journals and book series. Extensive field notes were taken during the interviews, meetings and the remainder of the fieldwork. We refrained from making audio recordings for several reasons: In the open-plan style offices the quality of the recordings can often be poor, due to background noise. This means formal interviews can only take place in separate, quiet areas, requiring the interviewees to move away from their desk, where they would not be able to show what they are doing (but rather have to rely on explaining it verbally), and they are in a less familiar or comfortable place. Some interviews did take place away from the desk, but only at the interviewee's suggestion. In general, we noticed that asking for recordings and having the recorder visible throughout the interview had an impact on the interviewees. They clearly tended to feel more reserved. Combining these considerations, we felt that we were able to collect richer and more accurate data when not recording the interviews. Alternatively, extensive notes were taken during the field visit, which were consequently processed on the same day. Admittedly, this creates an extra layer of interpretation by the researcher.

In this chapter, data is anonymised to the extent that all names of publishers, journals and individuals are omitted. When presenting empirical data, we provide generic job titles or descriptions to contextualise the quote or data while protecting the anonymity of the individual. Representatives of the publishers involved, as well as all editors from the more or less standalone journals, read the manuscript prior to submission and were given the opportunity to comment on issues of anonymity and factual mistakes.

7.4. Innovating the editorial process

In this section, we will first describe the practice of handling a manuscript by large publishers, following the editorial process from submission to the final decision of acceptance or rejection. Subsequently, we indicate how changes to this process may occur, commenting on the different roles of actors in the decision-making process leading to changes, the factors that influence whether a certain new initiative will be followed up, will be more widely adopted, or terminated. In this description, we will point to instances where different practices, roles and infrastructures meet and connect, commenting on how this affects the innovation process.

7.4.1. Following a manuscript: the editorial process

Even though some differences exist in the way different publishers handle their manuscripts, the process is fairly similar in broad terms, at least among the large publishers (both the ones we visited as well as other large publishers). When a manuscript is submitted through the online editorial management system, it will be handled by various actors, each with their own highly differentiated task:

- The first stage of manuscript processing is typically handled by people usually referred to as *manuscript editors.* Commonly, those editors are based in low-wage countries. They might either be employed by the publisher or work for a vendor company. Manuscript editors perform basic checks on, for example, the manuscript's structure, the plagiarism scan, declarations of ethical consent or issues, or compliance with reporting guidelines. They flag potential issues to the assistant editor and may send a bundled query to the author (the latter can typically be done only once in order not to slow down the process).
- The assistant editors or 'editorial assistants' (usually based in first-world head offices) consult the input from the manuscript editors and can perform some additional checks, for instance assessing ethical issues, third party rights, and duplication or overlap with other manuscripts. They might also send out queries to the author. In general, they try to balance several concerns: trying not to slow down the submission process, but also

aiming for completeness and clarity before peer review. This was a common theme that was raised several times i.e. the importance of making necessary changes or requiring additional information so as not to inconvenience the peer reviewers, safeguarding them from instances of unclarity, poor language, or with manuscript structure issues. At publisher A, the assistant editors felt checking for ethical issues is the most important part of their job and maintains the company's reputation: "We are very careful. We have to be very careful in order to protect the reputation of the publisher." However, this also has a more formal side to it: "We make sure we're not getting sued" (assistant editor).

- The assistant editors have clear targets with a number of manuscripts they are expected to handle per day. Handling a single manuscript is a highly standardised practice, which takes them between five and 25 minutes (depending on the level of experience of the individual and the complexity of the manuscript. Most editors reported to need on average about 10 minutes per manuscript).
  The assistant editor then asks an associate editor whether s/he wants to handle the manuscript through the actual review process. In some cases, the assistant editor might be in charge of sourcing and inviting reviewers, for which they will commonly consult the publisher's database, or external databases such as Scopus, Web of Science, or PubMed. Commonly, they will resort to keyword searches to find reviewers, while keeping track of the number of papers an academic has recently reviewed and how reviewers were rated by associate editors in the past.
- The associate editors or editorial board members (who usually are external academics, i.e. not employed by the publisher) receive the comments from the assistant editor. They source reviewers, when this is not already done, usually through their personal networks. In addition, they coordinate the review process and make a recommendation about acceptance/rejection to the executive editor. This work may be done on a voluntary, unpaid basis or editors may receive an honorarium. While some in the publishers' offices argue these external editors are enrolled because of their expertise and connection with the research field and the community, others also point to a cost argument. They argue that publishers aim to have as many manuscripts handled by associate editors as possible: "ideally, all manuscripts are handled by external editors" (executive editor), for obvious reasons of lowering workloads for internal editors and hence reducing costs. This makes recruitment of external editors "an important part of journal development" done by executive editors (executive editor).
- The recommendations for acceptance or rejection are then passed on to the executive editor, or editor-in-chief, who makes the decision. Executive editors may also have several additional roles: (i) if no external editor can be found, they take up editorial tasks of sourcing reviewers and managing the review process; (ii) they recruit new

associate editors for the editorial board; (iii) they keep in contact with / manage the editorial board; (iv) they undertake journal development (redefining the journal's scope, introducing new sections, managing editorial processes, marketing, etc.), and (v) they commission specific content for the journal (reviews, commentaries, thematic issues, etc.).

- The executive editor then reports the decision back to the assistant editor, who communicates the decision to the author. In the case that the manuscript is accepted, some final checks (similar to the initial checks after submission, but now stricter) are carried out by the assistant editor. After some potential final revisions, the manuscript is sent to the production units, which may be located in other countries. The various steps of the editorial process explained above are schematically depicted in figure 7.1.

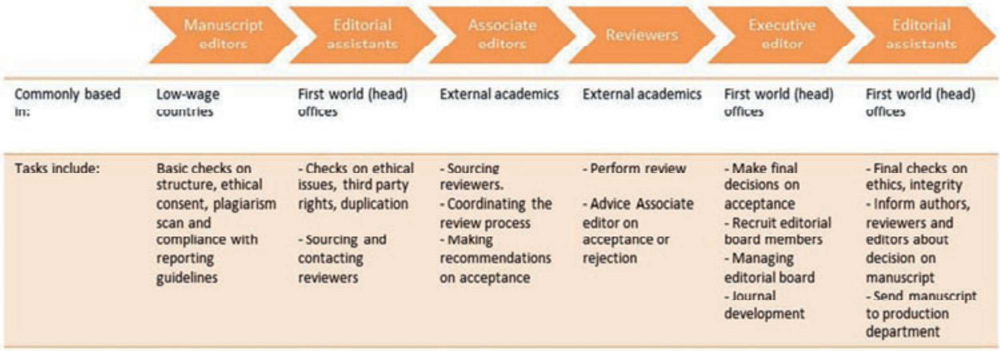


Figure 7.1: Schematic overview of the editorial process

Besides many journals owned and operated by the publishers themselves, both publishers have a number of journals in their portfolio that are owned by professional or learned societies. For these journals the publisher might take on slightly different roles. Most commonly, the publisher and society negotiate a division of the tasks mentioned above, in which the publisher performs a subset of the tasks, while others are maintained by the society. Within such societies, the task differentiation is commonly far less pronounced than in the large publishers: usually many of the different tasks mentioned above are carried out by only one or very few people.

In addition to individuals in roles directly related to the handling of manuscripts, several other people are involved in the core business of daily journal management (i.e. apart from technical, maintenance, catering or safety processes).

- Managing editors or team leaders typically manage a group of assistant editors. They keep track of their performances (in terms of targets, which usually centre on the

number of published/checked manuscripts), are an 'escalation point' for (difficult) issues with manuscripts, they provide training to new members, and they distribute new projects over the team members / journals. They might also have a say in decisions on the journals policy or peer review model.
- Journal managers manage the entire editorial process for a series or portfolio of journals. This person could be accountable for the process management of the journal, managing the editorial and peer review process, as well as for strategic management of the journal. This could include ensuring that journals are maintaining certain standards, for instance with respect to growth, author service, publishing integrity, etc. They can have a prominent role to play in journal policy and might be in contact with, and undertake negotiations with members of societies affiliated with the publisher.
- There are often various support teams, such as those that may help source reviewers in case a manuscript needs to be handled in-house or an associate editor cannot find reviewers. They do this based on bibliometric techniques, such as by using keyword searches, and use databases of academics doing reviews for similar journals. Another support team include for example those that manage editorial inboxes, by taking care of all author queries related to manuscripts 'under review'.

The review process in the publishers' journals is conducted pre-publication, mainly through either single-blind review (in which the reviewer knows the author's identity, but not vice versa), or open review (in which both authors and reviewers know each other's identity – in this case review reports may also be published online along with the published article). Some journals also uses the double-blind format (in which neither reviewers nor authors know each other's identity). Text similarity scanners are used, but no other (semi-)automated digital tools (such as statistics or image manipulation scanners). Reviewers, who are selected by the editors, do not get the opportunity to communicate with each other in the review process. Some journals at both publishers now also offer the Registered Reports model for doing review (Nosek and Lakens, 2014). However, editors report that the uptake of this model is still low and hardly any manuscripts are reviewed in this way.

The publication process at the big publishers is clearly very layered and hierarchical, organising the editorial process in a long procedural chain, with highly specialised division of labour. The process depends heavily on connections between, and coordination of the individual actors' practices. These consist of procedural solutions of providing and delivering information, but also of signalling potential issues that may need intervention of higher layers in the hierarchy. The material components of the practices, mainly embedded in the infrastructure of the electronic manuscript management system, play a key role in aligning and connecting the different tasks. This facilitates fast interaction between the different layers of the process and coordinates it by keeping track of a manuscript's status and of internal and external actors involved. This is also reflected in the editorial process of some journals not directly related to large publishers. One of the managing editors of such a journal mentioned: "We use a very old-fashioned online system. It is very basic, it cannot do much, but it works." In such smaller editorial teams, with less specialised division of labour, the online management system is less required to align the different actors in the editorial process.

Now that we understand the structure of and roles within the editorial process, we will turn our attention to how the process may change or transform. The hierarchical nature of the editorial process and the commercial logics that crucially inform its decisions will turn out to be essential factors to explain why and whether specific changes take place.

7.4.2. Innovating the process: getting ideas

Editorial teams learn about the editorial innovations currently suggested or tried out in academic publishing in various ways, including some channels of particular interest for academic publishers.

Several of the ideas travel with people moving between publishers or units within the same publisher: "The idea for the project comes from [the new team leader], who was on [another journal] and now joined our team. He took the idea with him. He made people aware of the project." (team leader of group of editors). With people switching jobs between publishers, not only knowledge concerning (new) editorial practices is exchanged, but also competences required to perform such practices, hence allowing practices to spread. Other ideas might reach the team via professional contacts, such as software developers or academics, e.g.: "I got to know the people who are involved in establishing the platform" (executive editor). Similarly, editorial board members might suggest new approaches, or they might indicate issues with the current approach, thereby triggering discussion about how to solve such issues. This is one of the main information routes for suggestions to publishers from their community and the editorial board keeps the journal in touch with the community. Another editor mentioned the value of social media: "The main source would be Twitter, actually", referring to online discussions and fast circulation of ideas on these platforms. Another occasion for new procedures or practices occurs when "the electronic management systems that we are using, enables something new. That is a major breakthrough for us" (Publisher). Such software innovations give publishers the opportunity to quite easily implement something new, but they also show that other publishers are working on similar initiatives, creating pressure to keep up with competitors. What we thus see is that personal networks of people working for the publisher are important, especially for those working close the community of authors and reviewers. However, also publishing technology can be a conduit for innovations and can be crucial to their introduction to work processes.

In this discussion it is important to distinguish between larger and smaller projects, the larger projects being those either rolled out over many journals or those requiring large investments. Smaller projects may find their way to the project team via the above mentioned formal and informal networks. In contrast, larger initiatives will come “from very high up” (team leader of editors), referring to the top management of the company. One of the team members initiating many of the large new projects explained that projects arrive at their desk mainly in two ways: either the head of an innovation department suggests new initiatives after consulting with technology companies, which demonstrate novel opportunities, or “management makes contractual deals with other companies.” This might include external partners, usually commercial providers of automated tools or review services, which can assist in review and collaborate through large deals. A journal portfolio’s publisher added that, for the larger innovations: “We will take the same approach as any other company might take, by looking at the market and seeing where gaps appear. […] We will look at spots where growth may be realised […] or where we see an opportunity” (Publisher, referring to market analyses carried out by a dedicated team within the company). Therefore, information channels do not just pertain to knowledge about academic benefits of innovations, but also about business economics, in particular for larger innovations.

7.4.3. Innovating the process: decision makers

The same factors mentioned earlier, the companies’ internal hierarchy and the distinction between smaller and larger projects, become evident when considering who can decide to start, continue or terminate new initiatives. Commonly, when ideas reach a publisher’s management team, they will first be implemented on a small number of journals, during pilot phases, before deciding to roll them out over a wider set of journals. Publishers hence opt for rather gradual transformation of the editorial practices, containing risks and creating learning opportunities. Currently, several trials involving smaller or larger modifications to the publication process are executed in journals managed by the publishers. These include the introduction of new checks carried out by the assistant editors, or more intrusive modifications, such as offering pre-print possibilities to authors; implementing some form of Registered Reports, in which research protocols or plans are reviewed before results or conclusions are known; and the introduction of variants of open peer review, in which peer review reports are published alongside the articles.

In the discussions about who can decide to implement innovations, the publisher’s internal hierarchy becomes very visible. Speaking to several members at the office, the phrase ‘people higher up’ was very common to describe where decision-making power lies: “These decisions come from many levels above me” (senior member of product development team). “But I am not the one making the decisions about this. That will happen higher up” (team leader of editors). This is mainly the case when decisions about larger projects have to be made: "The decisions about such large projects will be made on the very top level of the chief publishing director. In general, decisions about such large projects, enrolled [publisher]-wide, will be on this level. Smaller projects might be handled at a lower level. It depends on the potential impact that the project has, in terms of investment, required structural changes and potential of improvement" (publisher).

However, even in decisions about smaller projects, the hierarchical structure of large publishers' organisation becomes evident: "[…] in this case it was the chief editor who decided it, but she had to get permission from her manager" (executive editor) and "Management would then decide whether this is a good thing to do from a business perspective. If they decide to go for it, we would just do the implementation right away" (associate in policy team). In practice, decisions are made by managers (who manage portfolios of up to several dozens of journals) and people above them in hierarchy.

The situation is more nuanced for society journals, i.e. journals owned by professional or learned societies in which the publisher has a mainly supporting role. These journals are largely autonomous and the publishers have only limited involvement with their editorial policies and processes. Hence, decisions about how to organise the editorial and peer review process are largely made by these societies and their editorial managers, thereby creating additional layers next to the publisher's internal hierarchy. Societies can frequently lead the way in particular community developments and be a source of inspiration for the publisher. Conversely, if decisions involve discussions over a longer time period it may be more quickly implemented in journals entirely owned by the publisher.

The decision-making structure both shows clear patterns of hierarchy and layering of the different editorial practices. Larger projects travel top-down through the company, driven by the decision power at the higher levels in light of what innovations mean for their concerns. What constitutes a convincing argument for change varies between different layers in the hierarchy. It is to these meanings and arguments for change that we will now turn our attention.

7.4.4. Innovating the process: convincing arguments

Observing editorial transformation processes or asking about past transformations, we analysed reasons that convince decision makers to enrol or implement new initiatives. Not surprisingly, the commercial aspect of the publisher plays a crucial role in these decisions, with typical expressions such as: "At the end of the day we're still a business." However, on closer analysis, this commercial argument manifests itself in quite specific ways, related to the publishers' business model and corporate strategy. It is not just *that* business interests play a role in innovation decisions, but *how* these business interests are understood.

First, the publishers seem to be particularly keen on protecting or strengthening their reputation, even at the expense of additional resources. This becomes clear in the following excerpt from my field notes, from a conversation with a staff member about new initiatives to detect misconduct cases:

'I bring up that, because there are only relatively few integrity cases, it might not be worth a huge investment. She responds that: "One case can bring a lot of reputation damage" explaining that even though they have a lot of papers published and only relatively few of those contain issues, these still can cause major harm to the reputation of the publisher. I reply that hence, even though they might need to set up systems that take a lot of resources and may only catch a few cases, this may still be worth it, because it may prevent reputational damage: "Yes, absolutely", she confirms. "There needs to be a lot of trust in the system."'

Another consideration for the publisher, that is related to commercial considerations, is how the publisher can add value for researchers. This usually comes in a drive to speed up the editorial process, improve handling of submissions, decreasing turnaround times etc. However, it was also apparent that a balance is needed with on the one hand improving efficiency without compromising on delivering quality. From conversations with a number of publishing colleagues - at all levels - it was clear that projects are introduced where the publisher believes they are doing the right thing by the community in upholding standards that the community supports and are not introducing changes or innovations that would not receive support from the community.

A second way in which the commercial argument manifests itself, is through a continuous push to balance a need to speed up the editorial process, decreasing turnaround times and thereby decreasing costs, with the quality of the editorial process and consideration of the research community's needs. An executive editor responds to my question of what would be convincing arguments for her manager: "A lot of it will be time: we obviously don't do things that increase our turnaround times. Other factors include concerns about whether it is not too labour intensive, and we balance time and benefits" (executive editor). A member of the product development team explains: "The biggest cost for the publisher is the editorial process. […] And we do not want to undo any of the efforts that we made via other means in speeding up the process." She claims that "this [the increased cost due to more time-consuming editorial processes] will form the major barrier to enrol the project on a wider scale" (senior member of product development team).

A closely related third manifestation of the commercial argument concerns the number of submissions a journal expects. Short turnaround times are not only beneficial to reduced costs, they also indirectly lead to increased submissions and hence revenue potential: "Speeding up is required for authors, because they want quick turnaround times, and for the reputation of the journal" (executive editor). Another executive editor explains: "In general, when you want to convince people, you need to show that you have the backing of the field. […] In the end it all comes down to the number of submissions we get" (executive editor). Or a more direct claim: "We could never suggest anything to journals if it either makes the authors or the reviewers less likely to work with the journal" (Senior Manager Publishing Team). Another senior staff member, working alongside society owned journals, told me that: "We persuade people to get more and better content." Referring directly to the business value and corporate strategy of the publisher by strengthening its reputation, an assistant editor team leader stated: "… we always think about how we can improve the turnaround time, since we may use that as a selling point." However, this commercial logic has to be balanced with academic standards and this may involve measures that may conflict with individual researchers' interests. As another staff member later commented: "We reject papers, we retract papers, we publish papers, and each of these might be an unhappy circumstance for an individual author or individual reviewer. But what we try to do is the right thing by the community, upholding the standards that the community supports (..)"

A fourth argument also refers to adding business value, through responding to the needs of researchers in their various roles as editors, authors and reviewers. Factors here involve costs of introducing a new project while maintaining appropriate legal considerations and the perceived benefits to all parties of making a change. This becomes clear in a conversation with a publishing associate in the publishing policy and strategy team:

Me: "So if I understand you correctly, there are two main pillars on which you base your decisions, being first of all the legal aspect: Are we allowed to do this? And secondly the commercial or business aspects: "Can we actually make money out of this?"

She: "Yes, yes, that is true. But usually we are not directly looking at how much money we can make with it, because obviously many of these projects, such as the […] project, are not directly making money. Instead, we usually ask how much money we can afford to put into this. Because it will just cost us money to build the system, but if reviewers like it, it will make the reviewers happier and they might be willing to review for us another time."

Me: "And that might then make the review process go faster and hence lead to a more cost-effective process?"

She: “Yes, exactly. That is how it works” (publishing associate in the publishing policy, development and strategy team).

Another indication of what arguments are convincing can be observed in the monitoring and evaluation indicators for new project pilot phases. Several members of the editorial board explained how they would measure: “Turnaround times, the number of reviewers engaging, and the rate of reviewers accepting to review” (Publisher), and: “The usual thing we would measure is the uptake by authors” [i.e. the number of authors opting for a newly offered service] (member of product team). Such performance indicators closely articulate the business concerns in operational terms and translate them into terms directly relevant to editorial innovations. As we noted earlier, these feedback and monitoring mechanisms are crucial mechanisms to maintain or transform practices. Specifying monitoring criteria that operationalise speed and uptake will increase the endurance likelihood of practices that align with business strategy.

This attempt to increase efficiency might have further consequences for specific scientific disciplines. As we noted earlier, there is an increasing desire to standardise editorial processes across journals within the same publisher, mainly driven by efficiency and economy-of-scale considerations. Is this effort towards more standardisation, approaches are “commonly aligned with those of the largest set of journals that already use them or in which they naturally fit” (editor-in-chief). Hence approaches are typically modelled on disciplines in which the publisher holds most journals. This may decrease diversity in review procedures and potentially go at the expensive of approaches more suitable in other, smaller disciplines or research areas in which the publisher holds only few journals.

However, commercial arguments are not the sole factor facilitating or hindering innovation. In the complex hierarchy of tasks within the publisher, a single group of people is particularly involved with the academic aspects of the editorial process. This group, consisting of executive editors or editors-in-chief, is fairly distanced from direct financial and business considerations. It is this group that seems to be particularly keen on improving science and for whom a publisher is clearly distinct from ‘any other company’. One of them claimed: “We are a company, but we are not a manuscript accepting machine. […] I really want to do it well.” He was positive about his colleague editors thinking about it the same way. “It needs to go well” he claimed, “else I do not want to work for the journal any longer” (executive editor). Other members of the company acknowledged this role, describing the executive editors as “the guardians of quality” (in-house editor).

Once again, we can hence observe how different meanings are attached to editorial practices at different levels of the organisation. Whereas to some, the practices are instrumental in facilitating company growth, fostering reputation or creating unique selling points, others attach more academically informed meanings to the process, aiming to improve science and dissemination of scientific results. It is in the interplay of those layers and meanings that decisions about transformations of the editorial process are made.

7.4.5. Innovating the process: potential hurdles

Apart from analysing convincing arguments to implement or introduce new initiatives, the analysis of what constitute major hurdles to editorial transformations provides another interesting lens. Besides the legal restrictions mentioned above (such as privacy concerns or issues related to the General Data Protection Regulation), two main categories of impediments were foregrounded. These were explicitly articulated and summarised by one of our interviewees: “There are two main hurdles to innovation: People are very reluctant to change. In the end, we all want to keep things as they are. And there are the technical issues involved” (executive editor). Both impediments were expressed multiple times by members of the editorial team, which may be indicative that either habits and conventions among certain actors, or the technical configuration of the electronic editorial system are the main hurdles to innovation.

“Sometimes the reviewers have issues with that [a new format of the review process], even though they accept to review in this way at the very start” (executive editor), implying that the instructions did not come across.

“I hope to see it being implemented soon. This will require some changes to the [name of the electronic editorial system]. That is always complicated. It will take a while” (team leader of assistant editors).

Especially the latter hurdle is closely related to the material aspects of editorial practices, whereas the former commonly ties to the competence aspects of such practices. The preference for either of the two solutions can depend on the level of configuration that is possible within a manuscript management system and the costs involved. It may also be constrained by whether technical change and “automated solutions” could be preferable, because they require less behavioural change than individuals involved in the processes.

Hence some of the material and competence elements of (complexes of) practices, may provide barriers to transformation or innovation. As we discussed earlier, the infrastructural online manuscript management system plays a major role in connecting and coordinating the various editorial practices. However, while usually taken for granted and to some extent invisible, it gets foregrounded when transformations to the system have to be made. This aspect, common to, or even defining of, infrastructures (Star, 1999), makes the system one of the major hurdles in innovating editorial practices. Interestingly, these infrastructural aspects are not fully determined by the publisher itself. Both the structure of the electronic manuscript management system, the developers of such systems, and the competences and habits of reviewers represent and originate in connections to the editorial practices at other journals or publishers, as well as the practices of grant review at funding institutes.

7.5. Discussion: publishing in a changing landscape

These connections between publishers and the wider publishing and research community are obviously not restricted to habits of reviewers and the electronic systems used in the editorial process. Currently, several potentially fundamental changes are taking place in scientific publishing. These changes include a move towards open access publishing, increasingly demanded by major funding agencies (cOALition S, 2018), which has major impact on the publishers' business models. They also include a growing discomfort of multiple stakeholders in research about the role of (commercial) academic publishers. Fuelled by discussions about large profit margins of such publishers (Larivière et al., 2015), rising prices of subscriptions and open access fees, combined with simultaneous budget cuts in academic research and have done considerable harm to publishers' reputation.

In addition, the rise of other publishing formats, such as pre-print servers that no longer require the direct involvement of publishers (Walker and Rocha da Silva, 2015), raises further issues about the role of publishers. People have started to ask questions about how publishers add value to the publishing process or the published literature and about the justification for spending large amounts of, mainly public, funds on publishing through large, commercial publishers (Stern and O'Shea, 2019; Resnick, 2019).

Unsurprisingly, members of the offices we visited were well aware of this changing landscape and the potential implications for their work and products. The awareness of the publishers' need to demonstrate added value manifests itself in several ways. At one of the publishers, research integrity plays an important role in this. In a conversation with a senior member, she explained that there is currently a lot of outrage against publishers:

"People say that we don't need the publishers anymore, because they can just post research on their personal webpage or submit it to pre-print archives and have other people review it on these platforms. We therefore have to show the added value of publishing and our work [at the research integrity team] is a way of doing this."

She expected that, by demonstrating an effort to uphold integrity and publishing ethics, publishers can increase trust in the work they publish. "It conveys a message that we are taking this seriously. It shows people that they can trust the work we publish and thereby we show the added value of publishers" (staff member).

In fact, such reputational considerations might even warrant deliberate financial losses: "Yes, I don't think we are making any money on this journal. But it is our main open data sharing journal so we keep it because we want to seem like a publisher that supports open and transparent data" (member of product team). This is confirmed by a senior managing editor at the other large publisher: "there is no direct financial incentive for publishers to rock the boat to make a change. However, there is a long-term incentive to get involved. Publishers have to show that publishing is different from Wikipedia."

This is in line with the dominant view on the changing status of publishers, which requires the publishers to take considerable action. Part of this action, driven by the shift towards more open access publishing, consists of a changing focus on 'what the community wants'. While previously librarians would be the main source and spokesmen of 'what the community wants', publishers are quickly shifting their attention towards needs and desires of researchers, either in their role as authors or as reviewers. This aligns with the publishers' business needs: in the subscription-model librarians were involved in deciding which subscriptions to buy, but in the open access model researchers themselves are more directly involved in deciding where to publish and hence where to spend money on publishing.

A stronger focus on transparency constitutes another trend among publishers that is fuelled by external changes in the publishing landscape. By being more transparent about publishing work, for instance by showing how many reviewers had to be invited, could allow the publishers to demonstrate the effort that goes into the review process, thereby showing their added value: "We need to do a better job in showing how we have added value. Being open about review and the system is a way of doing this" (Senior Manager).

We thus see the specific meanings attached to parts of the editorial process. The practices of upholding research integrity and publishing ethics or increasing transparency might genuinely contribute to better research, but they simultaneously serve a direct business need. By several members of the publishers, they are understood in a fairly instrumentalist way, as mechanisms to safeguard or strengthen the publisher's reputation.

In this respect, it is necessary to make a distinction between the meanings and values attached to specific practices by the publisher as a whole and by people in their individual capacities. In the conversations we had with employees we noticed a remarkable usage of the first person singular or plural. When referring to the company's values and reputation, as well as their efforts to protect those, interviewees almost exclusively spoke in first person plural, e.g.: "We have to be very careful in order to protect the reputation of the publisher", "We make sure we're not getting sued" (assistant editor), and "we show the added value of publishers" (senior team leader). In contrast, when discussing more academic interest not as closely tied to the publisher's commercial interest, interviewees were more common to speak in the first person singular: “I really want to do it well” (in-house editor) and “I dream about a world not so much focussed on indicators and turnaround times” (executive editor). This suggests subtle distinctions between personal or professional concerns and company concerns that attach a different meaning to innovations’ potential.

7.6. Conclusion

This study identified several factors that are important in transforming editorial practices of commercial publishers. To understand innovation considerations, we first note that the editorial process at large publishers is very hierarchically structured, with distinct tasks for distinct layers of the process and thereby a clear division of labour among these layers. Extensive training for in-house editors and elaborate guidelines and manuals maintain a highly standardised and routinized process. The many layers of the process clearly express the complexity and inter-relatedness of editorial practices, as a combination of many mundane, simple practices distributed over various people and places (Shove et al., 2012).

Analysing how editorial practices may be transformed, we conclude that, while information about new initiatives circulates widely, projects tend to be typically implemented only on a relatively small scale. For larger projects, managerial approval has to be obtained. Analysing the convincing arguments for management to make changes in editorial practices, we observe several recurring themes. The major factor encompasses a commercial interest, which is understood as the importance to uphold reputation, shorten the editorial process and turnaround times, and increase willingness of researchers to be involved (either as editors, authors and reviewers).

Different meanings are at stake here, attached to editorial practices by different actors. Some assess and support editorial practices or innovations for how they might improve research. For others, the main question is what editorial processes mean for the publisher’s business model. Because the latter meaning is more common among people in higher layers of the companies’ hierarchy, this meaning tends to prevail in decisions on large-scale innovation projects. Hitherto, these commercial considerations have mainly been discussed concerning publishers’ attitude towards open access publishing or concerning predatory journals (e.g. Shen and Björk, 2015). However, we show that they reach much deeper, influencing the core of the editorial process. The potential of editorial or peer review innovations is not just assessed in terms of whether they will improve research, but also in light of whether they strengthen the company.

Last, factors commonly impeding rapid or large-scale changes in the editorial process are often related to infrastructural aspects such as the electronic editorial system or habits and conventions of individuals involved. These impeding factors, usually comprising the material and competence parts of editorial practices, are all examples of instances where the publisher’s editorial practices connect to those of other publishers, organisations or the wider publishing community.

The debate about new initiatives to develop or improve the editorial system or peer review system is usually centred on academic arguments: how the review system might improve the research enterprise. Arguments about the advantages of open peer review or similar innovations highlight advantages for research, obtaining knowledge and distributing it. As new ideas about how to organise the editorial process emerge, academic or societal considerations are therefore predominant (Horbach and Halffman, 2018). On a small scale, such considerations may well drive innovative projects in academic publishing. However, when it comes to large-scale implementation, other considerations and motives come to the fore. We show that the editorial process is closely connected to commercial practices of creating business value, and the very specific terms in which this is understood, such as reputation considerations and the urge to increase efficiency. This might help explain why some innovations have currently been successfully implemented on a wide scale, such as plagiarism detection software, whereas others remain peripheral, in spite of strong arguments from their supporters. Arguably, those innovations aligning with the specific understanding of the publisher's business model are most likely to witness successful implementation. This might provide valuable insights for future endeavours to innovate the academic peer review system and improve its functionalities.

7.7. References

Bowker GC and Star SL. (2000) Invisible Mediators of Action: Classification and the Ubiquity of Standards. *Mind, Culture, and Activity* 7: 147-163.

Bravo G, Grimaldo F, López-Iñesta E, et al. (2019) The effect of publishing peer review reports on referee behavior in five scholarly journals. *Nature Communications* 10: 322.

Campbell DT. (1979) Assessing the impact of planned social change. *Evaluation and Program Planning* 2: 67-90.

cOALition S. (2018) *Accelerating the transition to full and immediate Open Access to scientific publications*. Available at: https://www.coalition-s.org/about/.

de Wit O, van den Ende J, Schot J, et al. (2002) Innovation Junctions: Office Technologies in the Netherlands, 1880-1980. *Technology and Culture* 43: 50-72.

Fyfe A, Coate K, Curry S, et al. (2017) Untangling academic publishing: a history of the relationship between commercial interests, academic prestige and the circulation of research.

Godlee F. (2002) Making reviewers visible: openness, accountability, and credit. *Jama* 287: 2762-2765.

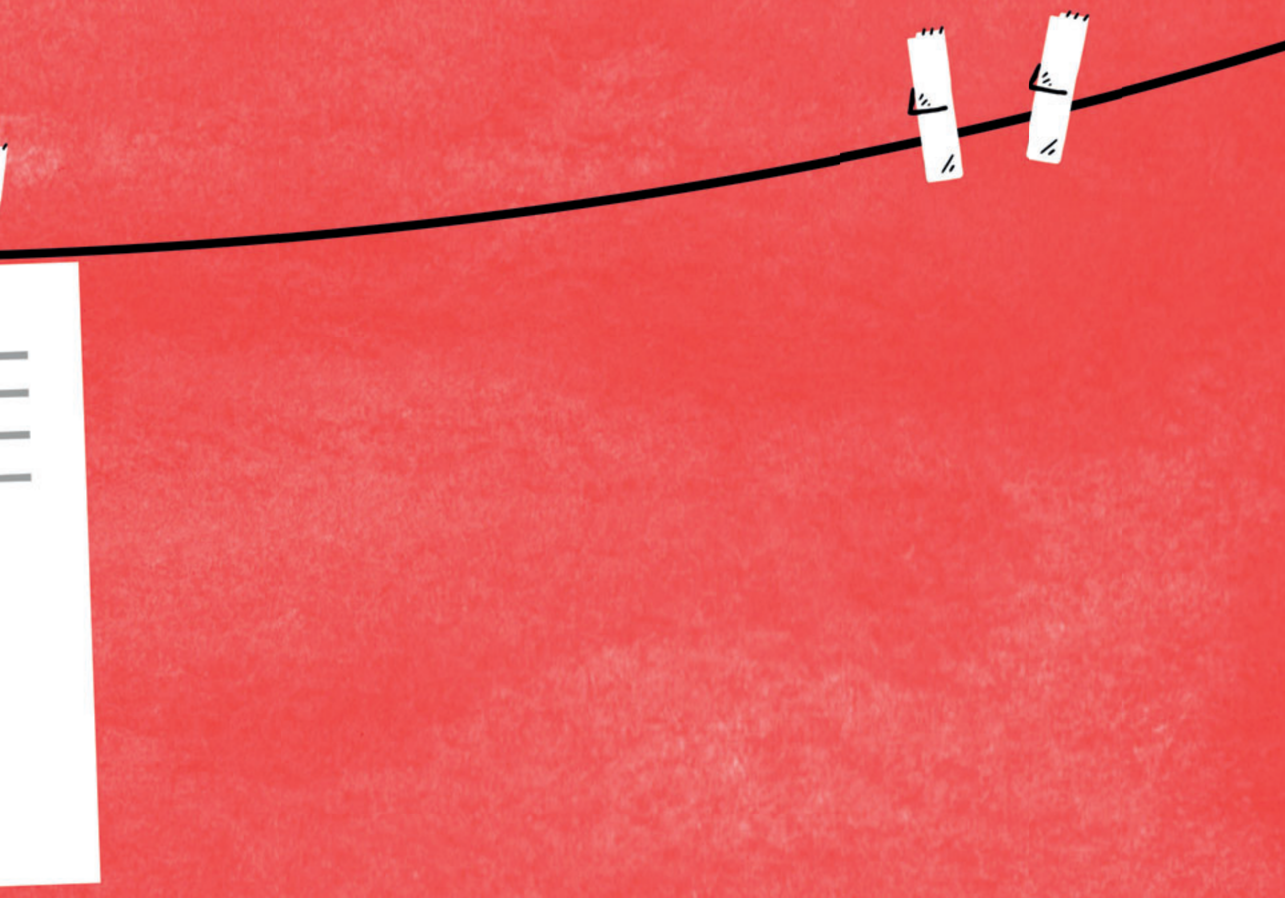
Guston DH. (2007) *Between politics and science: Assuring the integrity and productivity of reseach*: Cambridge University Press.

Hansoti B, Langdorf MI and Murphy LS. (2016) Discriminating Between Legitimate and Predatory Open Access Journals: Report from the International Federation for

Emergency Medicine Research Committee. *Western Journal of Emergency Medicine* 17: 497-507.
Hirschauer S. (2010) Editorial Judgments: A Praxeology of 'Voting' in Peer Review. 40: 71-103.
Horbach SPJM and Halffman W. (2018) The changing forms and expectations of peer review. *Research Integrity and Peer Review* 3: 8.
Horbach SPJM and Halffman W. (submitted) Journal Peer Review: Cautious Innovator or Sleepy Giant? *Minerva*.
Jacob M-A. (2019) Under repair: A publication ethics and research record in the making. *Social Studies of Science* 49: 77-101.
Larivière V, Haustein S and Mongeon P. (2015) The Oligopoly of Academic Publishers in the Digital Era. *Plos One* 10: e0127502.
MacIntyre A. (1985) *After virtue : a study in moral theory,* London: Duckworth.
Nosek BA and Lakens D. (2014) Registered reports: A method to increase the credibility of published results. *Social Psychology* 45: 137-141.
Peters DP and Ceci SJ. (1982) Peer-review practices of psychological journals: The fate of published articles, submitted again. *Behavioral and Brain Sciences* 5: 187-195.
Reckwitz A. (2002) Toward a Theory of Social Practices:A Development in Culturalist Theorizing. *European Journal of Social Theory* 5: 243-263.
Resnick B. (2019) *The costs of academic publishing are absurd. The University of California is fighting back.* Available at: https://www.vox.com/science-and-health/2019/3/1/18245235/university-of-california-elsevier-subscription-open-access.
Rinkinen J, Shove E and Smits M. (2017) Cold chains in Hanoi and Bangkok: Changing systems of provision and practice. *Journal of Consumer Culture* 0: 1469540517717783.
Schatzki TR. (1996) *Social practices: A Wittgensteinian approach to human activity and the social*: Cambridge University Press.
Schatzki TR. (2002) *The Site of the Social: A Philosophical Account of the Constitution of Social Life and Change*: Penn State University Press.
Shen C and Björk B-C. (2015) 'Predatory'open access: a longitudinal study of article volumes and market characteristics. *Bmc Medicine* 13: 230.
Shove E, Pantzar M and Watson M. (2012) The dynamics of social practice : everyday life and how it changes. Los Angeles: SAGE.
Smith R. (2006) Peer review: a flawed process at the heart of science and journals. *Journal of the Royal Society of Medicine* 99: 178-182.
Star SL. (1999) The Ethnography of Infrastructure. *American Behavioral Scientist* 43: 377-391.
Stern BM and O'Shea EK. (2019) A proposal for the future of scientific publishing in the life sciences. *PLOS Biology* 17: e3000116.
Walker R and Rocha da Silva P. (2015) Emerging trends in peer review—a survey. *Frontiers in Neuroscience* 9.
Zuckerman H and Merton RK. (1971) Patterns of evaluation in science: Institutionalisation, structure and functions of the referee system. *Minerva* 9: 66-100.

Part III

TO CLEAN



In the previous parts of this dissertation we have seen how problematic research may spread through the academic literature, and we got an impression of how the editorial process may be designed to filter such research. However, one can hardly expect this system to be completely impermeable, if only because of the controversy about what constitutes problematic research. Hence, cases of contentious research will continue to slip through and end up in the research literature. This requires further systems of regulation to be developed and studied. In addition, it necessitates proper ways to deal with cases of problematic research practices once they have come to light. It is to these issues that the last empirical part of this dissertation will turn its attention.

In *To Clean* we will divert our attention from academic journals to study the context of research performing organisations, most notably that of universities. These organisations obviously play a crucial role in the research process. Therefore they are prominent places where issues with research integrity may appear and where they should be dealt with. While the role of dealing with cases of problematic research practices may be distributed among actors, (leading) academic journals tend to shift the responsibility for investigating allegations of misconduct to the organisation at which the alleged fraudster works (Bauchner, Fontanarosa, Flanagin, & Thornton, 2018). Only after such organisation has investigated a case, these journals feel the need to respond according to the investigation's conclusions.

In this part's chapters we revisit some of the phenomena that we came across in previous parts of this dissertation. Most notably, we continue on the role of power structures and hierarchy, which became evident in our ethnographic study of academic publishers. We extend the study of these phenomena beyond the editorial processes to encompass relations between active researchers in organisational constellations. We do this both to study the reporting of cases allegedly involving research misconduct as well as the handling of such cases by universities.

In chapter 8 we first develop a theoretically and empirically informed understanding of the causes and consequences of reporting research misconduct in terms of power relations. To this end, we use a multinational survey in a set of European universities to study the whistleblowing process. Based on qualitative data provided by witnesses of problematic research practices, we examine actors' rationales for reporting and not reporting misconduct. We also study how they report cases, e.g. to whom, and the consequences of reporting.

In the final empirical chapter of this dissertation, chapter 9, we continue the study of reported misconduct cases. Instead of focussing on the reporting, we will now study what happens once an allegation has been filed. While obtaining very little attention in the context of research misconduct, such processes have received ample attention in management and organisational studies. Specifically, the effect of an organisation's responses in the aftermath of misconduct cases has been subject of several studies. These studies mainly build upon the theory of organisational sensemaking (Gangloff, 2014; Gioia & Chittipeddi, 1991; Weick, 1995; Weick, Sutcliffe, & Obstfeld, 2005). Key elements in this literature are the processes invoked when organisations experience extraordinary events, such as accidents, crises, or, in our context, potential cases of misconduct.

We extend these studies to the context of academia by employing a comparative case study approach to describe and assess the handling of four cases of alleged misconduct by their university, respectively in the Netherlands and Norway. Chapter 9 proposes a theoretical model that explains how organisational responses to misconduct emerge and evolve as iterations of the processes of sensemaking, sensegiving, and sensehiding. In addition, it links these iterations to a set of background premises, such as organisation's characteristics, nurturing the organisational responses. Last, it analyses the perceived consequences of organisations' way of dealing with cases of alleged misconduct, both for the organisations themselves and for the individuals involved.

With these analyses, *To Clean* aims to make our studies of research integrity issues full circle. Supplementing the previous parts, it now studies both different actors and contexts, i.e. those of research performing organisations, and a different aspect of problematic research, i.e. the handling of detected cases of alleged misconduct. Relying on survey and interview data gathered at universities across Europe, it aims to understand what challenges arise in this set of actors and how their practices and procedures may be improved to foster research integrity.

Chapter 8 – On the willingness to and consequences of reporting research misconduct: the role of power relations

This chapter is under review as:
Horbach, S. P. J. M., Breit, E., Halffman, W., & Mamelund, S.-E. (Revisions under review). On the willingness to and consequences of reporting research misconduct: the role of power relations Science and Engineering Ethics.

8

8.1. Introduction

In recent years, the attention paid to research integrity and misconduct has increased. Besides attempting to measure or estimate the extent of misconduct, several scholars have also investigated its causes. This literature shows that important drivers of misconduct range from individual personality traits to systemic factors, which include productivity pressure and corporate influences (Fanelli et al. 2015; Tijdink et al. 2016; Horbach and Halffman 2019). Much less attention has been paid to how scientific misconduct is denounced and detected, for instance, through the peer review system (Guston 2007; LaFollette 1992) or social control mechanisms such as whistleblowing (Stroebe et al. 2012). These processes are crucial for signalling misconduct and articulating what the research community deems un/acceptable behaviour. In addition, the detection and sanctioning of research misconduct depend almost entirely on discovery and reporting by peers, with the potential exception of plagiarism and some forms of statistical manipulation, which require automated detection by means of ‘scanners’.

In this paper, we aim to provide a more elaborate theoretical and empirical understanding of the causes and consequences of reporting research misconduct. We do so by approaching the issue from a whistleblowing perspective (Near and Miceli 2016; Santoro and Kumar 2018; Vandekerckhove 2016) and by applying theories of power and power differences. The literature provides different definitions of ‘whistleblowing’ and to avoid confusion, we use the term ‘reporting’. We understand this term as formally giving account of alleged research misbehaviour to a party tasked with handling such cases. In our use of the term, we therefore do not involve ‘softer’ forms of reporting, such as talking to colleagues, and we do not distinguish between internal and external whistleblowing (e.g. Gao et al. 2015). The concept of reporting may thus involve reporting by following a university’s procedures, or informing one’s superior or other actors with formal tasks and responsibilities (e.g., an ombudsperson, members of an ethics board), or even the press.

A better understanding of research misconduct reporting could contribute to improved early warnings and more effective preventive policies to promote research integrity. Rather than appealing to individuals to take responsibility and relying on sanctions, such policies should pay more attention to social processes, such as power imbalances, group pressure and performance pressure. More specifically, appropriate reporting policies should target a culture of complacency and cynicism that normalises questionable research practices, or even outright misconduct (Clair 2015; Martinson et al. 2010). The literature on organisational integrity in general has focussed on power imbalances, retribution concerns and career consequences (Bowie 2010; Palazzo 2007), which probably also play a role in research integrity.

In this study, we aim to better understand the processes that facilitate or inhibit the reporting of alleged research misconduct by analysing a large sample of direct and indirect witnesses of research misbehaviour. We draw on qualitative responses from a 2017 quantitative, multinational survey in eight European countries. In this survey, respondents who indicated they had directly or indirectly witnessed an instance of misconduct were asked to respond to open-ended questions about this instance and how they handled it.

We focus on the following research question: How do varying power positions influence the reporting, or not reporting, of alleged research misconduct? Given the data available to us, we focus on three specific power positions: academic seniority (i.e. the formal work position), work contracts (i.e. permanent vs. temporary appointments) and gender. These power elements have been identified as key factors in most commercial organisations' studies on organisational integrity (Dozier and Miceli 1985; Cassematis and Wortley 2013; Culiberg and Mihelic 2017). We also study the influence of the specific type of misconduct, i.e. whether it involves a clear-cut type, such as plagiarism, or a more contested 'questionable research practice' (QRP), such as the disputed attribution of authorship.

To our knowledge, this is the first systematic study of researchers' reasons for and accounts of reporting or not reporting witnessed misconduct. Our aim is to extend studies of whistleblowing and misconduct reporting from a commercial, mostly private, setting to that of academia. Our article is structured as follows: section two presents an overview of the literature on research integrity and misconduct. In section three, we introduce studies of whistleblowing and power, deriving factors potentially affecting researchers' willingness to and the consequences of reporting alleged misconduct. Section four describes our study's survey methodology, while section five presents this survey's main qualitative empirical results. In section six, these findings are formulated as propositions relating power relations to the reporting of alleged misconduct. We suggest these propositions can be used to explore further hypothesis-testing research on the relationship between power differences and misconduct reporting. Finally, section seven offers concluding remarks and policy recommendations.

8.2. Literature

Over the past decades, research misconduct has drawn the attention of scholars from various fields. Currently, an extensive literature focusses by and large on the prevalence and causes of misconduct, or on questionable research practices (QRP). The literature focuses to a lesser extent on the consequences of misconduct, for example, how retracted journal articles' career effects (Azoulay et al. 2017), how institutions deal with alleged misconduct cases (Horbach et al. 2018b, 2018a), the consequences for research reliability (Horbach and Halffman 2017; Al-Marzouki et al. 2005), and on the role of scientific misconduct in general (Schulz et al. 2016).

This literature has also developed an inventory of 'novel' forms of academic misbehaviour (Callaway 2015; Sacco et al. 2018; Biagioli et al. 2019; Bouter et al. 2016), including estimations of how often certain of these forms occur (e.g. Hopp and Hoover 2017; Fanelli 2009) and several 'risk' categories of authors, including scientific fields and geographical areas where research misconduct or QRPs occur more frequently (e.g. Fanelli et al. 2015; Yang 2013; Stitzel et al. 2018). Similarly, the literature has outlined several potential causes of misbehaviour in science, including individual researchers' personality traits (Tijdink et al. 2016); the organisational context in which these researchers operate (Anderson et al. 2007; Forsberg et al. 2018); and more systemic causes, such as competitive research funding and 'publish or perish' pressures (Fanelli et al. 2017; Sarewitz 2016).

Much less attention has been paid to the mechanisms that might detect and identify scientific misconduct. Some have argued that the peer review system is a prime example of such a mechanism (Guston 2007; LaFollette 1992), others' hopes rest on social control mechanisms, most notably whistle-blowers and the close colleagues of misbehaving scientists (Stroebe et al. 2012). Despite the little evidence gathered on this topic, case studies of misconduct cases suggest that alleged culprits' close colleagues and peers are the most likely way of bringing misconduct to light (Horbach et al. 2018a).

Specifically, the processes involved in signalling and reporting alleged misconduct are not well researched. For example, we are not aware of any study examining researchers' motivations for reporting alleged academic misconduct, or of any research on such actions' effectiveness. Some studies have, however, made a case for establishing safe whistleblowing procedures in academic organisations (Forsberg et al. 2018). There is also a significant literature on whistleblowing procedures in the business ethics and management fields (e.g. Culiberg and Mihelic 2017; Palazzo 2007; Vandekerckhove 2016). Nevertheless, little is known about how these processes are applied in academic research institutions such as universities. Questions about who are most likely to report, their motivations for doing so, and the effectiveness and potential negative consequences of reporting remain unanswered. A deeper understanding of how and why researchers raise concerns or keep quiet can contribute to developing organisational conditions that will support a stronger culture of research integrity.

8.3. Theoretical framework

In addition to the literature on research integrity, there is a vast body of research on integrity in organisations in general, based on studies of wrongdoing in (or by) organisations (e.g. Palmer 2012), and research on organisational integrity management (e.g. Paine 1994). The literature on organisational integrity has paid more explicit attention to whistleblowing and the reporting of misbehaviour (Vandekerckhove 2016; Near and Miceli 2016). Among the more central theoretical questions in the literature on whistleblowing are the factors influencing (a) a witness of wrongdoing's decision whether or not to report such instances; (b) the extent to which a reporter faces, or fears, retaliation; and (c) reporting's effectiveness in terms of addressing wrongdoing. Besides the likelihood and effectiveness of reporting, the literature has also highlighted whistleblowing's potentially negative consequences, such as psychosocial and reputational consequences (Bjørkelo and Matthiesen 2012; Culiberg and Mihelic 2017; Park and Lewis 2018).

The role of power relations is widely acknowledged in respect of whistleblowing or the reporting of wrongdoing in organisations. We therefore highlight two central conceptualisations of power in this literature. As theorised in the resource dependence theory, the first notion understands power as a central resource or asset that an individual may possess (Lukes 2005). From this perspective, an organisation's less powerful members – such as younger employees, people with temporary work contracts, women, or people lower in the organisation's hierarchy – are less likely to report alleged misconduct.

Several factors constitute this decreased reporting likelihood. For example, these actors may have less access to powerful social networks in the organisation and therefore have less social capital. In addition, younger, and thus less experienced employees, may have less knowledge of the procedures and of how these are applied in practice. Low-resource members may also face more retaliation and are generally less able to achieve genuine and desirable change as a result of reporting a case, such as adequate intervention in wrongdoing cases, or even improved integrity policies (Gao et al. 2015). This is especially true in cases involving more powerful wrongdoers.

A second theoretical dimension is French and Raven's theory of social power, i.e. having the ability or being in a position to influence others. The theory involves five power bases: legitimate power (based on the legitimate right to prescribe behaviour), referent power (based on identification with one another), expert power (based on special knowledge or expertise), reward power (based on the ability to award resources), or coercive power (based on threats of punishment) (French et al. 1959). This suggests that reporters of alleged misconduct lacking such power bases are less likely to be effective, especially when reporting the misbehaviour of more powerful organisation members. This lack of power may affect the likelihood that they will report misconduct and the outcome of their reporting.

Based on the above, we expect researchers with fewer resources to be less likely to report alleged misconduct and their reporting to be less likely to result in effective interventions (Mesmer-Magnus and Viswesvaran 2005; Gao et al. 2015). Such researchers may include those in junior positions, such as doctoral students and post-docs, as well as those with temporary contracts. In addition, social power theory indicates that researchers with higher seniority (i.e.

who have worked in academia longer) are more likely to (effectively) report misconduct cases (Cassematis and Wortley 2013; Gao et al. 2015).

Consequently, even though the role of power in whistleblowing has not been studied in an academic context, the variables, (i) academic seniority and (ii) temporal vs permanent work appointments, are expected to affect the willingness to and the consequences of reporting research misconduct. In addition, some of the obstacles to and the potential consequences of reporting are believed to affect women more, although the evidence for this from the general integrity literature is inconclusive (Mesmer-Magnus and Viswesvaran 2005). To shed more light on the topic, gender will be explored as a third variable potentially influencing reporting.

Lastly, several studies have outlined the influence of the type of wrongdoing on the likelihood of reporting. In particular, clear-cut instances of misbehaviour are more likely to be reported than nuanced cases, which may be prone to different interpretations and normative assessments (Near and Miceli 1985; Mesmer-Magnus and Viswesvaran 2005). Accordingly, more indisputable forms of research misconduct, such as fabrication, falsification and plagiarism (FFP), are more likely to be reported than more questionable research practices, such as disputes over authorship or text recycling. Furthermore, if witnesses of wrongdoing perceive that there is high probability of their complaint being taken seriously and it is less likely to backfire, they are more likely to report it. We also study the type of witnessed misconduct's influence as an additional variable in our model.

8.4. Methods

8.4.1 Data collection

Data on research misconduct as a workplace issue were collected by means of a web-based, cross-sectional survey (Questback). Central themes in this survey were organisational policies regarding misconduct and integrity, whistleblowing mechanisms and attitudes, tensions arising from and the risks of research misconduct, perceptions of integrity measures, and the prevalence of research misconduct (Mamelund et al. 2018).

In this paper, we draw on data collected as part of the survey, which consisted of open-ended questions on the respondents' possible first-hand knowledge of a research misconduct incident. This approach was adopted from the validated and revised Scientific Misconduct Questionnaire (SMQ-R) (Habermann et al. 2010; Broome et al. 2005). The open-ended questions were:

- How did you first learn about the instance of research misconduct?
- Please describe the specific instance of research misconduct.
- What did you do when you became aware of it?

- Whom (titles only) did you talk to?
- Were you able to talk to the individuals who were involved?
- Was the instance reported? To whom and by whom?
- What was the outcome? How did you feel about how it was handled?
- Did you think anything changed as a result?
- Is there anything you would have done differently?

8. 4.1.1 Participant selection

The survey was conducted with the employees of the European PRINTEGER project's eight partner universities (PRINTEGER 2016). A PRINTEGER member sent a link to the survey questionnaire to the principal investigators in each of the partner universities. These investigators subsequently forwarded information about the survey and the link to a senior manager at their institution, who distributed these to the target population at their universities, i.e. the academic staff, excluding the technical and administrative staff. The reason for this distributed approach was to ensure a high response rate. The senior managers were encouraged to inform their academic staff of the survey and to highlight its importance.

Data collection took place from 7 March to 1 August 2017. The total population across the eight partner institutes comprises 20,815 academic staff members. Overall, 1126 respondents participated in the survey, with 194 responding to the open-ended questions. The demographical characteristics of the open-ended questions' and the survey's respondents are provided in table 8.1. Table 8.1 provides information on the potential reporting bias by indicating the differences between the qualitative sample's respondents and the participating universities' general population. As shown in the table, the qualitative sample consists of older researchers and those with more senior positions (i.e. professors and associate professors) than found in the general population. The qualitative sample also consists of a higher proportion of social and behavioural sciences' researchers.

Each of the eight universities provided population data. Two challenges arose from the population analyses. Firstly, the coding system across the universities revealed inconsistencies regarding academic positions, especially regarding the meaning of 'teacher' and 'academic field'. Where available, we used individual-level data and discarded the faculty information. Secondly, not all universities had access to staff members to cover all the demographic variables. For example, one of the universities could not reach its PhD students.

Variables	Subgroups	Qualitative sample (%)	Survey sample (%)	Population (%)	Difference between qualitative sample and population
Gender	Male	52.7	47.8	53.6	-0.9
	Female	47.3	52.2	46.4	0.9
Age	20-29	12.6	20.8	24.6	-12
	30-39	27.8	30.5	32.7	-4.9
	40-49	20.4	21.2	18.8	1.6
	50-59	24.0	16.2	15.4	8.6
	60+	15.2	11.3	8.6	6.6
Position	Professor	31.9	18.1	12.2	19.7
	Associate professor	33.5	36.4	28.7	4.8
	Teacher	0.0	1.3	5.7	-5.7
	Post-doc	12.4	10.6	22.5	-10.1
	PhD student/TA	22.2	33.5	30.9	-8.7
Appointment	Temporary	45.0	51.0	NA	-
	Permanent	55.0	49.0	NA	-
Academic field	Engineering	5.3	5.0	5.3	0.0
	Language, info, comm.	2.6	3.8	5.7	-3.1
	Law, arts, humanities	15.3	17.1	20.1	-4.8
	Medical/life science	27.4	23.6	27.7	-0.3
	Natural sciences	21.0	17.3	24.7	-3.7
	Social and behavioural	28.4	33.2	16.6	11.8

Table 8.1 Demographic distribution of responses

8. 4.1.2 Privacy and ethics approval

The relevant ethics committees at each participating institution granted their ethics approval of the survey. The privacy policy was explained before the participants started the survey and they were asked to agree to this, thus providing informed consent. The responses were collected anonymously and are not traceable to the respondents' institutes. We present quotes from the responses with the demographic information that the relevant respondent provided

8.4.2 Data analysis

The analysis used for this paper is primarily qualitative. Owing to the relatively small sample size and the diversity of the universities studied, as well as the richness of the qualitative responses to the open-ended questions, we refrain from an in-depth statistical analysis, focussing instead on the content of the open-ended questions' responses. Our analysis is therefore primarily inductive. We explore the issues raised in the responses and develop propositions that future research can test. The relatively high number of qualitative responses allowed us to examine the different responses' frequency to demonstrate more and less common responses, but not to estimate precise rates in terms of the population. Consequently, we do not provide significance levels and do not claim that these proportions can be generalised.

The responses were coded using NVivo. We collapsed the nine initial questions into the following five categories due to the overlap between some of the questions: (1) type of research misconduct witnessed, (2) source that led to awareness of the misconduct, (3) initial reaction to the awareness, (4) type of reporting and (5) outcome of the reporting. Broome et al. (2005) took a similar approach, although the categories are not entirely similar.

We adopted an open coding approach within these five categories, which meant that we assigned labels to the responses as articulated, rather than using a pre-existing template (Strauss and Corbin 1990). For the 'type of misconduct witnessed' category, we also drew on existing categorizations of misconduct, such as falsification, fabrication and plagiarism. In the overall analysis, we categorised some responses into multiple codes. Consequently, the frequency of the codes may be larger than the overall number of respondents. Appendix A provides an overview of the coding and the responses, including illustrative examples.

After coding, we analysed the relationships between the responses and the demographical variables, using simple cross-tables. Given our interest in power relations, we focused especially on gender, age, academic position and work contract. We also counted the frequency of the codes related to the different demographic variables, whose results are provided in the following sections, but mainly serve as a contextualisation of the qualitative material forming the core of our results.

In our analysis, we focus on several aspects of reporting alleged misconduct. In respect of each of the variables, we analyse the way power relations are related to: (i) researchers' willingness to report alleged misconduct, (ii) the way in which cases are reported (e.g. to whom) and (iii) the reporting's perceived effect. In the analysis of the latter, we distinguish between reporting having no consequences, negative consequences and constructive consequences. In respect of negative consequences, we refer to cases leading to retaliation against reporters, whereas in respect of constructive consequences, we refer to respondents' descriptions of adequate intervention in misconduct cases or improved integrity codes and procedures.

8.5. Results

In this section, we will first (5.1.) look more closely at the distribution of the responses with respect to the three variables outlined in section 3: seniority, work contracts and gender. We analyse the responses in terms of each variable in three ways: (i) whether or not a case was reported, (ii) how and to whom a case was reported, and (iii) the respondent's perception of the consequences of reporting. We then (5.2) explore in detail how the respondents' reporting varies across the different types of misconduct witnessed. These findings suggest how elements of power are involved in misconduct and its potential reporting's social dimension, which, in section 6, lead to several propositions on power relations' role in the reporting of alleged misconduct.

8.5.1 Effect of power relations on reporting

8.5.1.1 Academic seniority

The existing literature on whistleblowing suggests that researchers in junior or lower academic positions are less likely to report alleged misconduct compared to those in more senior positions. In our data, there is indeed a division between professors and researchers in lower positions. In our limited sample, professors reported a witnessed case of alleged misconduct more often (67% reported vs 29% not reported) than other members of academia, such as associate professors (37% vs 53%), post docs (35% vs 61%) and PhD students or TAs (39% vs 51%). Since we defined reporting as giving account of a case to any party tasked with handling such cases, these observations hold even when taking into consideration that some junior researchers would report misconduct to senior researchers, who would then report it to official channels, such as a research integrity committee.

When we examine the responses across age groups, more senior researchers are also more likely to report misconduct. While respondents in the age group 20-29 claim to have reported misconduct in 33% of the witnessed instances, the other age groups' percentages are 32% (30-39), 51% (40-49), 65% (50-59), and 51% (60+). In this small set, the effect of age and academic rank could not be isolated, but it does suggest that age should currently not be ignored as a significant factor.

The qualitative responses provide more insight into the reasons for reporting or not. Notions of power, in this case in the sense of resource availability, seem to play a crucial role. A prominent reason for junior researchers not reporting was their fear of negative consequences for them, such as losing future opportunities or the hampering of their social relations at work. This is exemplified in the following quotes:

No, I did not push it further for fear of career consequences (30-39, female, Law/Arts/Humanities, PhD student, temporary, 0-5 years).

I reported to the direct supervisor, but was sure I could not go beyond that, as that would directly have impeded my own relation with my own supervisor (I am a PhD student) (20-29, male, Natural Sciences, PhD student, temporary, 0-5 years).

No, I decided to not report it, because I'm a junior researcher and afraid that it would affect my career possibilities (20-29, female, Medical/Life Sciences, PhD student, temporary, 0-5 years).

Even though less common, some more senior researchers also reported this fear of negative consequences:

Nothing. If I said sth., I'd be disadvantaged in all [aspects] of my work. Office politics (No age, no gender, Natural Sciences, permanent, assistant/associate prof, 16+ years).

Another frequently mentioned reason was distrust of the management's willingness to take any corrective action, as exemplified below:

No. My department manager never takes any action on any problem (40-49, female, Social/Behavioural Sciences, PhD student, temporary, 0-5 years).

A third reason mentioned is a belief that reporting would not lead to any changes, notably due to certain individuals being protected. These responses show how issues of seniority, hierarchy and power affected respondents' decision not to report an alleged case:

No. My old boss is too powerful in the community (30-39, male, Natural Sciences, leadership role, assistant/associate prof, temporary, 6-10 years).

Nothing, there is no going against my boss. There have even been lawsuits in the past, but the University has always covered for her (30-39, female, Law, Arts and Humanities, PhD student, temporary, 0-5 years).

In all these examples, the common hierarchical structure in academia plays a prominent role in an actor's decision to report or not. Both conceptions of power outlined in section three (resource dependence and French and Raven's theory of social power) become visible in the respondents' comments. Specifically, the control of resources, such as senior colleagues' research and promotion opportunities, seems to be a prime concern. The latter even extends beyond the research organisation's immediate environment to the wider research field and peer community. The organisational and wider research context may thus be a source of normalising behaviour – among others due to restricted reporting.

Another response mentioned that power relations, in particular seniority, may not only affect the reporting of alleged misbehaviour, but may actually be one of its causes. A female professor in the life sciences explains how she was ‘pressured’ into behaving in dubious ways:

> I was working with a more senior professor to promote findings of some research through a prestigious impact/ knowledge mobilisation event which was presenting the 'best evidence to inform practice'. The prof wanted to promote a tool we had developed as having a positive impact. Myself and the wider research team had concerns that we had no evidence of the tool's efficacy and in fact the small feasibility study had raised some concerns about its effect. We wanted to wait until the full trial was complete before promoting it as a tool. Although we had agreed as a team the limits of what could be said about the tool, I was the only team member working with the professor on this impact event and a few days before, he called me to [participate in] a teleconference with the sponsor of the event, and they both put a lot of pressure on me to allow the tool to be presented as effective. When I started to explain to the sponsor what the concerns of the team were, the professor muted the call and told me not to tell her that! It was a very intimidating situation and I felt I had to withdraw (40-49, female, Medical/Life Sciences, leadership role, assistant/associate prof, permanent, 16+ years).

The mentioned fear does not only concern fear of superiors, but also fear – especially related to such superiors – of the misconduct incident becoming public. Thus, while power plays a key role in discouraging reporting, senior researchers may not perceive their wielded power as self-serving, but as an effort to protect a collective interest, often the research institution’s image (although this could be considered misguided). The following quotation is from a female PhD student explaining why she did not report an incident after consultation with her professor:

> Only to the professor, who insisted it [should] not [be] reported to the ethics officer. They didn't want it to become public and wanted to fix it themselves (20-29, female, Medical and Life Sciences, PhD student, temporary, 0-5 years).

The following is a similar response from a more senior respondent:

> The instance was not reported to maintain the reputation of the faculty. The decision was made solely by the professor involved (60-69, female, Natural Sciences, researcher, permanent).

We also examined to whom the incidents were reported. Our data suggest that the different levels of seniority have different reactions. Of our respondents, professors (26%) and associate professors (27%) informed a supervisor more often about an incident than PhD students/TAs

(12%) and postdocs (4%). Conversely, PhD students/TAs (29%) and postdocs (35%) responded more often by talking to their colleagues about the misconduct than professors (7%) and assistant/associate professors (11%) did. Furthermore, professors and associate professors confronted the culprits more often (26% and 24%) than postdocs (17%) and PhD students/TAs (12%). Although the tendency is not very strong, this may be cautiously interpreted as 'soft' responses dominating with regard to junior researchers, while more senior academics make use of 'harder' means.

Our material also showed that respondents in more senior positions more often perceived the outcome of reporting alleged misconduct as constructive, than those in junior ones. Professors experienced a constructive change (34%) most often, followed by associate professors (23%), postdocs (22%) and PhD students/TAs (17%). The tendency is similar across ages. Overall, this level of reporting indicates that a perceived constructive change is uncommon across all positions, but even more so regarding the most junior academic positions; however, the limited observation size indicates that the findings should be regarded with caution.

The following responses by a female PhD student are an example of 'no change' after reporting an incident. She reported an issue of undeserved authorship, primarily due to work-related frustrations:

To [an] ombudsperson by me (PhD student). Decision was made because I was suffering greatly as a PhD student under her (sic.) supervisor; several other instances of misconduct also applied in our work relationship, in addition to emotional abuse (being personally criticized, being called an[d] yelled at after hours, being pressured into misconduct, ...) (20-29, female, Social and Behavioural Sciences, PhD student, temporary, 0-5 years).

She continued to explain that reporting her supervisor was difficult due to the latter's social position and, crucially, that she had refrained from reporting any misconduct ever since:

The attitude seemed to be that as a senior she [had to] know what she was doing, and as a junior researcher, I felt met with disbelief. Little action was undertaken, and I have refrained from reporting any misconduct ever since.

The expectation that action will be taken and the perceived guarantee that reporting a case will have an effective outcome can affect academics' willingness to report cases. Organisational procedures, such as reporting to an ombudsman, could offer powerful resources to redress imbalances, but these have to provide convincing intervention opportunities.

Finally, the issue of power was not always mentioned as a barrier to reporting others' misconduct. Our material also had an example of a PhD student who had benefited from her supervisor's power use. This student used the survey to reflect on this incident:

I talked with the professor, who put my name on the paper; he felt he had done me a favour and I did not object. Not reported. I decided it was good for my career and I should just leave it (regarding a case where “my name was put on a paper that I had had nothing to do with”; 20-29, female, Medical/Life Science, PhD student, temporary, 0-5 years).

8.5.1.2. Work appointment

The second variable we explored constitutes the relation between employment precarity and misconduct reporting. This was based on the assumption that researchers with temporary contracts are less likely to report alleged misconduct compared to researchers with permanent contracts. As we will show, our data suggest that the employment conditions and academic seniority patterns are similar. A confounding factor between the two variables could be that, in general, professors and associate/assistant professors are likely to have permanent positions, while PhD, teaching assistant and Post-doc positions are usually, if not always, temporary.

Examining the answers to the respondents’ willingness to report, we found that researchers in permanent positions report incidences of suspected misconduct twice as often as those in temporary positions. Whereas 59% of researchers in permanent appointments reported such incidence, only 31% of those in temporary appointments did so.

In the qualitative responses, the respondents never explicitly mentioned temporal employment, which was only indirectly mentioned in their responses through references to power and hierarchy. For example, a common reason for researchers in temporary positions not reporting was their fear of this having negative effects on their career possibilities. The following quotation exemplifies this fear:

No, I can't because of hierarchy. It is a superior (sic) and denouncing could have a negative impact on my job (40-49, gender: other, Law, Arts and Humanities, assist/assoc prof, temporary, 11-15 years).

No, because this study was important [for] the dissertation of the PhD student who had limited time. I didn't feel like I had enough support to get out of this unharmed (male, 30-39, left academia, temporary, Social Sciences).

Others attributed their lack of reporting to the management not taking them seriously:

I did not report this explicit situation. I have however gone to the ombudsperson for similar situations (p-hacking, unauthorized authorship, ...) and was again discarded (sic) as […] the junior one with no experience (20-29, female, Social Sciences, PhD student, temporary, 0-5 years)

Such responses and ways of reasoning were far more common in respect of respondents with temporary work contracts compared to their permanently appointed colleagues. The same holds for another common response, namely a lack of knowledge regarding what to do, i.e. where and how to report:

No, I wasn't directly working with him anymore when I found out and I didn't know what to do (30-39, female, Social and Behavioural Sciences, assistant/associate prof, temporary, 11-15 years).

No, no idea where I can report this (30-39, female, Law, Arts and Humanities, assistant/associate prof, temporary, 6-10 years).

We also examined the variation in researchers with permanent and temporary work contracts' types of responses. The results indicate that researchers in permanent positions confront the culprits more often (26%) than those in temporary positions (14%). Such confrontations often involved students or other co-authors, as in the following example:

[I] talked to the persons involved, co-authors, and reported [this misconduct] to a faculty representative specialized in misconduct (60-69, male, Medical and Life Sciences, professor, permanent, 16+ years).

Researchers in permanent positions also informed their superiors more often (26%) than those in temporary positions (14%). Similarly, researchers in temporary positions 'did nothing' more often (23%) than their colleagues in permanent positions (15%). These researchers also talked to colleagues more often (24%) than those in permanent positions (11%). In the qualitative responses, these statements were usually not explained or narrated, but emerged in the form of "I did nothing" or "I discussed it with my colleagues". However, there were noteworthy exceptions:

Nothing, because the present head of department (new last author [of] this paper) tries to eliminate me and I am completely dependent [on] him (female, 60-69, Medical/Life Sciences, assistant/associate prof, temporary, 16+ years).

Nothing, due to the principal researcher's wish (30-39, male, Medical/Life Sciences, temporary, assistant/associate prof, 0-5 years).

Here too, the dependency on research and career resources that staff members in more permanent positions control seems crucial in terms of the use of power relations.

Finally, we examined the variation in researchers with temporary and those with permanent positions' perception of the outcomes of reporting. Our data suggest that a larger percentage of researchers with permanent positions report a constructive change, as shown in the following:

> [I] asked the editor of the journal to withdraw the paper; the thesis itself did not suffer, as the citation and reference were included in another chapter in the same thesis (60-69, male, Medical and Life Sciences, professor, permanent, 16+ years).

> This person got fired (30-39, female, Social and Behavioural Sciences, associate/assistant professor, temporary, 6-10 years).

Only researchers in permanent positions reported negative outcomes (7%), such as “the relationship with the author involved is somewhat troubled” (60-69, female, Medical and Life Sciences, professor, permanent, 11-15 years). Consequently, while a fear of reporting’s potential negative consequences is often presented as a reason for not reporting a case, there are hardly any consequences at all in practice. This refers specifically to researchers with temporary contracts.

Finally, researchers in temporary positions reported ‘no change’ to a greater extent (55%) than those in permanent positions (38%). These responses were usually not narrated, but expressed in the form of “no change” or “Not much – but at least more of my earlier work is cited” (30-39, male, Natural Sciences, assistant/associate professor, temporary, 6-10 years).

Overall, therefore, researchers in permanent positions reported misconduct more often than those in temporary positions, and their reporting is more likely to have a constructive result. This indicates that, in academia, the type of work appointment may have an effect on the reporting practices and their outcomes. Possible reasons for the difference between the two groups are that researchers in temporary positions feel they have more to lose by reporting and are less interested in it, because they identify less with the work organisation.

8.5.1.3 Gender

The third variable we explore concerns gender differences in the reporting of alleged misconduct and the consequences of this. This exploration was based on the literature’s assertion that men are more likely to report misconduct and to perceive the consequences as more constructive than women. However, contrary to our expectation and as we will show, few of the differences are related to gender.

There was little difference between men and women regarding the reporting of alleged research misconduct. 51% (N=50) of the men and 45% (N=40) of the women claimed to have reported a witnessed case of misbehaviour.

Neither does there seem to be substantial gender differences regarding reactions to misconduct. With regard to the most commonly reported forms of acting upon cases of alleged misconduct, 24% of women and 20% of men reported having confronted the culprits. The same is true of 'talking to colleagues', which 19% of women and 14% of men did, and of 'informing supervisor', which 20% of women and 22% of men did. 'No response' was selected by 16% of the women and 18% of the men.

Finally, our data show that there are no substantial differences between men (27%) and women's (26%) perceptions of reporting having constructive consequences. Women report a slightly higher number of negative consequences (6%) than men (3%), but the relative difference is small and the absolute number is so low that we cannot draw any clear conclusions from this result.

In the qualitative responses, there were also no noteworthy differences between men and women regarding how the instances of the reporting of misconduct and reactions to this were articulated and made sense of. None of the responses used wording related to gender. In terms of our data, gender does not seem to be distinctively related to researchers' reporting of misconduct, or to the outcomes of their reporting.

8.5.2. Type of misconduct reported

The last variable we analysed concerns the type of misconduct witnessed. Although not directly related to power structures, this variable likely influences reporting behaviour, as more clear-cut types of misconduct are more readily reported than more nuanced forms of misconduct (Mesmer-Magnus & Viswesvaran, 2005; Janet P. Near & Miceli, 1985). We distinguish between fabrication, falsification and plagiarism (FFP) as the clear-cut forms of misconduct, and the rest (QRP) as more nuanced forms of misconduct (See Appendix A).

Table 8.2 shows how plagiarism was the most commonly reported type of misconduct, followed by authorship issues, fabrication, cherry picking, falsification, text recycling, and data manipulation. Considering relative, rather than absolute numbers, we conclude that the more contentious forms of misconduct such as authorship and cherry picking had a lower reporting ratio than the more clear-cur forms such as plagiarism, falsification and fabrication.

Was it reported?	**Yes**	**No**	**Don't know**	**NA**
Plagiarism (N=60)	38 (63%)	14 (23%)	5 (8%)	3 (5%)
Authorship (N=49)	16 (33%)	31 (63%)	0 (0%)	2 (4%)
Cherrypicking (N=28)	9 (32%)	18 (64%)	1 (4%)	0 (0%)
Falsification (N=14)	6 (43%)	7 (50%)	0 (0%)	1 (7%)
Fabrication (N=13)	11 (85%)	2 (15%)	0 (0%)	0 (0%)
Text recycling (N=7)	5 (71%)	2 (29%)	0 (0%)	0 (0%)

Data manipulation (N=5)	3 (40%)	2 (40%)	0 (0%)	0 (0%)
Don't understand (N=5)	2 (40%)	3 (60%)	0 (0%)	0 (0%)
Do not wish to answer (N=1)	1 (100%)	0 (0%)	0 (0%)	0 (0%)

Table 8.2: Types of misconduct reported[2]

We also examined the perceived outcomes of reporting for the different forms of misconduct. Table 8.3 shows that plagiarism had the highest number of constructive consequences resulting from the reporting, followed by cherry picking and falsification. In terms of no change, authorship had the highest number, followed by plagiarism and cherry picking.

Did anything change?	Constructive consequences	Negative consequences	No change	Don't know	NA
Plagiarism (N=60)	20 (33%)	3 (5%)	19 (32%)	10 (17%)	8 (13%)
Cherry picking (N=28)	8 (29%)	0 (0%)	14 (50%)	0 (0%)	6 (21%)
Falsification (N=14)	7 (50%)	0 (0%)	6 (43%)	0 (0%)	1 (7%)
Fabrication (N=13)	6 (46%)	3 (23%)	4 (31%)	0 (0%)	0 (0%)
Authorship (N=49)	4 (8%)	1 (2%)	35 (71%)	2 (4%)	7 (14%)
Text recycling (N=7)	3 (43%)	1 (14%)	2 (29%)	0 (0%)	1 (14%)
Data Manipulation (N=5)	2 (40%)	0 (0%)	3 (60%)	0 (0%)	0 (0%)
Don't understand (N=5)	2 %	0%	40%	0%	40%
Do not wish to answer (N=1)	100%	0%	0%	0%	0%

Table 8.3: Perceived consequences of reporting

These numbers also reflect the contours of the difference between the clear-cut forms of misconduct and the more contentious forms. Likewise, as we noted earlier, there is a generally remarkably low chance of experiencing a positive, constructive consequence of reporting, with a much higher rate of respondents mentioning perceiving no changes at all.

Overall, therefore, we found that clear-cut cases of misbehaviour are reported more often than nuanced cases. Since nuanced cases are difficult to assess normatively, which could make a (formal) misconduct case a precarious endeavour, this could explain the difference. Likewise, given the difficulty of justifying such accusations, researchers may feel it is a too uncertain undertaking, risking long, intensive procedures and potential repercussions. Several respondents did indeed indicate as much in their answers:

No... since it's not actually forbidden, but still [an] unethical research practice (cherry picking, referring to "[w]ild and unjustified analyses until there is a significant result"; 30-39, female, Social Sciences, PhD student, temporary, 0-5 years).

[2] In the tables we only include the subset of the total sample, only including forms of misconduct where N>5.

I reported to the direct supervisor, but was sure I could not go beyond that, as that would directly have impeded my own relation with my own supervisor (I am a PhD student). Since this is an instance of [an] incomplete assessment of [an] error on results, this does not seem to classify as direct misconduct (20-29, male, natural sciences, PhD student, temporary, 0-5 years).

8.6. Discussion

Our survey results suggest that demographic differences affect the likelihood of cases of alleged misconduct being acted upon, i.e. reported and dealt with constructively. Firstly, the analysis suggests that younger researchers, researchers with temporary appointments and those in lower academic positions are less likely to report misconduct, compared to their more senior and permanently appointed colleagues. Secondly, contested forms of misconduct (e.g. authorship, cherry picking of data) seem to be reported less than more clear-cut instances of misconduct (e.g. plagiarism, text recycling and the falsification of data).

These trends, which emerge in the quantitative survey data, can be meaningfully interpreted in combination with the qualitative data. In the respondents' answers to the open-ended survey questions, they frequently attribute their decisions regarding whether and how to report to hierarchy or power issues. Some respondents do this very explicitly by referring to their previous superior, or to their current hierarchical relationship with their supervisor. Others hint more implicitly at hierarchy and power issues, implying power imbalances in the sense of resource dependence.

Based on the numerical data and our respondents' interpretation of these and the context, we can refine and translate the general claims made in the whistleblowing literature into specific propositions on reporting alleged research misbehaviour at academic institutes. Owing to the sample size limitations and the participating research institutes' high diversity, our results might be insufficient to draw definitive, statistically relevant conclusions, but we do believe that they provide strong indications that future studies could verify. We will formulate the propositions relating to the variables studied separately: academic seniority, the temporality of work contracts, gender, and the type of misconduct witnessed.

The first set of statements involves academic seniority and age. Overall, the results suggest that seniority in research constitutes a valuable resource in respect of reporting alleged research misconduct. This arguably manifests itself through access to important resources during the reporting process. Our respondents maintained that these resources include, for example, the knowledge to identify such misconduct and the appropriate reporting channel, the social capital and power position to act on misconduct (e.g. due to the low perceived career risk), as well as the ability to follow up on cases in order to secure a constructive outcome. Junior researchers may not possess the same knowledge and social capital, and fear harming their academic careers. These exploratory findings suggest the following propositions for further hypothesis-testing research:

Proposition 1: Researchers in junior academic positions and younger researchers are less likely to report instances of alleged research misconduct compared to more senior and older researchers.

Proposition 2: Reporting by researchers in junior positions or younger researchers is less likely to lead to constructive consequences compared to that of more senior and older researchers.

The second set of statements involves the temporality of work appointments. It is clear that for researchers and all other forms of employees, temporary contracts constitute an element of power imbalance. Temporary employment essentially entails that these employees are excluded from tenured work agreements and therefore face the risk of their work contracts not being renewed. It is reasonable to believe this, and our results also provide initial evidence of such a lack of power and social capital manifesting itself in these researchers' reporting behaviour. For example, their fear of negative personal consequence in the aftermath of reporting alleged misconduct, which respondents with temporary contracts expressed more often than others, is a manifestation of their lack of power. This leads us to the second set of propositions:

Proposition 3: Researchers with temporary contracts are less likely to report instances of alleged research misconduct than those with permanent contracts.

Proposition 4: Reporting by researchers with temporary contracts is less likely to lead to constructive consequences than reporting by those with permanent contracts.

The third set of statements involves gender. Somewhat surprisingly, we did not find any strong indications that gender differences play a role in reporting alleged research misconduct. However, we believe the theoretical underpinnings of gender as a central dimension of power imbalances in organisations and in academia more specifically (Aagaard 2016; Grilli and Allesina 2017; Treviño et al. 2017) are strong enough to warrant further studies. Contrary to our findings, we thus propose the following propositions:

Proposition 5: Female witnesses of alleged research misconduct are less likely to report such instances than their male colleagues.

Proposition 6: Reporting by female researchers is less likely to lead to constructive consequences than reporting by male researchers.

The fourth and final statement category involves the characteristics of the alleged misconduct. Although not a distinct power dimension, it does shed light on the likelihood of researchers' perception that their complaint will be regarded seriously, acted upon and lead to constructive consequences. Given that clear cases of misconduct are easier to identify and their reporting easier to justify (Near and Miceli 2005), we propose the following final statement:

Proposition 7: Researchers are more likely to report clear-cut instances of alleged research misconduct than more nuanced or 'grey areas' of misbehaviour.

The lack of reporting of 'grey' forms of misconduct is due to the crucial negative effect of such forms of misconduct being potentially continued. In other words, not only are such forms of misconduct per definition difficult to assess normatively, they are also likely to be more unspoken and implied in research. This involves the risk of such practices becoming embedded and institutionalised rather than openly discussed and reflected upon. Indeed, institutional or national integrity committees' processing of allegations of misconduct has often led to the codification of research practices (Horbach et al. 2018a). Consequently, if cases of specific types of alleged misconduct are not reported and integrity committees cannot assess them subsequently, they may not be classified as either proper or improper research practices.

8.7. Conclusion and recommendations

In this study, we have analysed reporting as one of the social control mechanisms flagging misbehaviour in science. In particular, we have studied the actors who are most likely to report alleged misconduct, how they report, and the consequences of reporting. We found differences in the rate of reporting and the consequences thereof, depending on the demographic characteristics of the person witnessing the case.

These insights contribute to the literature on research misconduct in two ways. Firstly, to our knowledge, we provide the first systematic insights into researchers' reasons and explanations for reporting, or not reporting, witnessed misconduct. We find indications that younger researchers, researchers with temporary appointments and those in lower academic positions are less likely to act and report than their senior and permanently appointed colleagues. The crucial hurdles for not reporting are these researchers' concerns that this may harm their career and their expectation of not being taken seriously, both of which are rooted in power relations and hierarchical differences leading to resource dependence.

We also find that contested forms of misconduct (e.g. authorship, cherry picking of data and fabrication of data) are less likely to be reported than more clear-cut instances of misconduct (e.g. plagiarism, text recycling and falsification of data). The respondents mention that minor misbehaviour is not considered worth reporting, or express doubts about the effectiveness of reporting a case when the witnessed behaviour does not explicitly transgress norms, such as with many of the QRPs. Concern about reporting's negative consequences, such as career opportunities or organisational reputations being harmed, is always taken into considerations.

Secondly, we have theorised the relationship between power differences and researchers' willingness to report – in particular the role of seniority, work appointments and gender. We have derived a list of seven propositions that we believe warrant testing and refinement in future studies using a larger sample to help with further theory building about power differences and research misconduct. More specifically, by focusing on such structural power dimensions, we provide a different perspective than most prior studies of scientific misconduct, which have mainly focused on the negative consequences for the individual wrongdoer and his/her colleagues. We thus open up a broader organisational understanding of the mechanisms that impact researchers' ability and willingness to successfully report misconduct.

Based on our study, we argue that establishing fair whistleblowing procedures is a prime requirement to specifically empower less powerful members of the research community to report scientific misbehaviour. This may also specifically strengthen one of science's most important social control mechanisms. Following Lukes's and Bachrach and Baratz's conception of power in the form of agenda setting (Lukes 2005; Bachrach and Baratz 1962), whistleblower procedures are a prime way of making latent and covert interests visible, thereby demanding decisions from those in power.

Our findings may have several implications for policy. We argue that policy interventions, such as research integrity courses for junior researchers, the articulation of research integrity codes, or integrity boards have to consider the power imbalances in research organisations. Our results suggest a need for improved early-warning systems and particularly for safe whistleblowing procedures. Specifically, such procedures should take the position of an organisation's less powerful members, such as junior researchers and people with temporary work appointments, into account and facilitate their whistleblowing. This requires procedures that effectively address issues of power imbalance and the fear of not being taken seriously. The implementation of such procedures could help target a culture of complacency and cynicism that normalises questionable research practices. In addition, it may contribute to a sense of organisational responsibility that should ultimately foster a climate of research integrity.

8.8. Acknowledgement

The current version of this chapter has benefitted from copy editing by Ilse Evertse (Ilse Evertse & Associates).

8.9. References

Aagaard, K. (2016). New and persistent gender equality challenges in academia. *Scandinavian Journal of Public Administration, 20*(1), 87-90.

Al-Marzouki, S., Evans, S., Marshall, T., & Roberts, I. (2005). Are these data real? Statistical methods for the detection of data fabrication in clinical trials. *BMJ, 331*(7511), 267-270. doi: 10.1136/bmj.331.7511.267

Anderson, M. S., Ronning, E. A., De Vries, R., & Martinson, B. C. (2007). The perverse effects of competition on scientists' work and relationships. *Science and Engineering Ethics, 13*(4), 437-461.

Azoulay, P., Bonatti, A., & Krieger, J. L. (2017). The career effects of scandal: Evidence from scientific retractions. *Research Policy, 46*(9), 1552-1569. doi: https://doi.org/10.1016/j.respol.2017.07.003

Bachrach, P., & Baratz, M. S. (1962). The Two Faces of Power. *American Political Science Review, 56*, 941-952.

Biagioli, M., Kenney, M., Martin, B. R., & Walsh, J. P. (2019). Academic misconduct, misrepresentation and gaming: A reassessment. *Research Policy, 48*(2), 401-413. doi: 10.1016/j.respol.2018.10.025

Bjørkelo, B., & Matthiesen, S. B. (2012). Preventing and Dealing with Retaliation Against Whistleblowers. In W. Vandekerckhove & D. Lewis (Eds.), *Whistleblowing and Democratic Values*: The international whistleblowing research network.

Bouter, L. M., Tijdink, J., Axelsen, N., Martinson, B. C., & ter Riet, G. (2016). Ranking major and minor research misbehaviors: results from a survey among participants of four World Conferences on Research Integrity. *Research Integrity and Peer Review, 1*(1), 17.

Bowie, N. E. (2010). Organizational integrity and moral climates *The oxford handbook of business ethics*: Oxford University Press.

Broome, M. E., Pryor, E., Habermann, B., Pulley, L., & Kincaid, H. (2005). The Scientific Misconduct Questionnaire—Revised (SMQ-R): Validation and Psychometric Testing. *Accountability in Research, 12*(4), 263-280. doi: 10.1080/08989620500440253

Callaway, E. (2015). Faked peer reviews prompt 64 retractions. *Nature News*.

Cassematis, P. G., & Wortley, R. (2013). Prediction of Whistleblowing or Non-reporting Observation: The Role of Personal and Situational Factors. *Journal of business ethics, 117*(3), 615-634. doi: 10.1007/s10551-012-1548-3

Clair, J. A. (2015). Procedural Injustice in the System of Peer Review and Scientific Misconduct. *Academy of Management Learning & Education, 14*(2), 159-172. doi: 10.5465/amle.2013.0243

Culiberg, B., & Mihelic, K. K. (2017). The Evolution of Whistleblowing Studies: A Critical Review and Research Agenda. *Journal of business ethics, 146*(4), 787-803. doi: 10.1007/s10551-016-3237-0

Dozier, J. B., & Miceli, M. P. (1985). Potential predictors of whistleblowing - a pro-social behavior perspective. *Academy of Management Review, 10*(4), 823-836. doi: 10.2307/258050

Fanelli, D. (2009). How Many Scientists Fabricate and Falsify Research? A Systematic Review and Meta-Analysis of Survey Data. *Plos One, 4*(5), 11. doi: 10.1371/journal.pone.0005738

Fanelli, D., Costas, R., Fang, F. C., Casadevall, A., & Bik, E. M. (2017). Why Do Scientists Fabricate And Falsify Data? A Matched-Control Analysis Of Papers Containing Problematic Image Duplications. *bioRxiv*. doi: 10.1101/126805

Fanelli, D., Costas, R., & Lariviere, V. (2015). Misconduct Policies, Academic Culture and Career Stage, Not Gender or Pressures to Publish, Affect Scientific Integrity. *Plos One, 10*(6), 18. doi: 10.1371/journal.pone.0127556

Forsberg, E.-M., Anthun, F. O., Bailey, S., Birchley, G., Bout, H., Casonato, C., . . . Zöller, M. (2018). Working with Research Integrity—Guidance for Research Performing Organisations: The Bonn PRINTEGER Statement. *Science and Engineering Ethics, 24*(4), 1023-1034. doi: 10.1007/s11948-018-0034-4

French, J., Raven, B., & Cartwright, D. (1959). The bases of social power. *Classics of organization theory, 7*, 311-320.

Gao, J. Y., Greenberg, R., & Wong-On-Wing, B. (2015). Whistleblowing Intentions of Lower-Level Employees: The Effect of Reporting Channel, Bystanders, and Wrongdoer Power Status. *Journal of business ethics, 126*(1), 85-99. doi: 10.1007/s10551-013-2008-4

Grilli, J., & Allesina, S. (2017). Last name analysis of mobility, gender imbalance, and nepotism across academic systems. *Proceedings of the National Academy of Sciences, 114*(29), 7600-7605.

Guston, D. H. (2007). *Between politics and science: Assuring the integrity and productivity of reseach*: Cambridge University Press.

Habermann, B., Broome, M., Pryor, E. R., & Ziner, K. W. (2010). Research Coordinators Experiences with Scientific Misconduct and Research Integrity. *Nursing research, 59*(1), 51-57. doi: 10.1097/NNR.0b013e3181c3b9f2

Hopp, C., & Hoover, G. A. (2017). How prevalent is academic misconduct in management research? *Journal of Business Research, 80*, 73-81. doi: https://doi.org/10.1016/j.jbusres.2017.07.003

Horbach, S. P. J. M., Breit, E., & Mamelund, S.-E. (2018a). Organisational responses to alleged scientific misconduct: Sensemaking, sensegiving, and sensehiding. *Science and Public Policy*, scy068-scy068. doi: 10.1093/scipol/scy068

Horbach, S. P. J. M., Breit, E., & Mamelund, S.-E. (2018b). Organisational responses to alleged scientific misconduct: Sensemaking, sensegiving, and sensehiding. *Science and Public Policy, 46*(3), 415-429. doi: 10.1093/scipol/scy068

Horbach, S. P. J. M., & Halffman, W. (2017). The ghosts of HeLa: How cell line misidentification contaminates the scientific literature. *Plos One, 12*(10), e0186281. doi: 10.1371/journal.pone.0186281

Horbach, S. P. J. M., & Halffman, W. (2019). The extent and causes of academic text recycling or 'self-plagiarism'. *Research Policy, 48*(2), 492-502. doi: https://doi.org/10.1016/j.respol.2017.09.004

LaFollette, M. C. (1992). *Stealing into print: fraud, plagiarism, and misconduct in scientific publishing*: Univ of California Press.

Lukes, S. (2005). *Power: A Radical View* (2nd edition ed.). Basingstoke and New York: Palgrave MacMillan.

Mamelund, S.-E., Breit, E., & Forsberg, E.-M. (2018). A multinational survey on research misconduct and integrity: a workfloor perspective (DIV.2). In PRINTEGER (Ed.). Oslo.

Martinson, B. C., Crain, A. L., De Vries, R., & Anderson, M. S. (2010). The importance of organisational justice in ensuring research integrity. *Journal of Empirical Research on Human Research Ethics, 5*(3), 67-83. doi: 10.1525/jer.2010.5.3.67

Mesmer-Magnus, J. R., & Viswesvaran, C. (2005). Whistleblowing in organizations: An examination of correlates of whistleblowing intentions, actions, and retaliation. *Journal of business ethics, 62*(3), 277-297. doi: 10.1007/s10551-005-0849-1

Near, J. P., & Miceli, M. P. (1985). Organizational dissidence - The case of whistle-blowing. *Journal of business ethics, 4*(1), 1-16. doi: 10.1007/bf00382668

Near, J. P., & Miceli, M. P. (2005). Standing Up or Standing By: What Predicts Blowing the Whistle on Organizational Wrongdoing? *Research in Personnel and Human Resources Management* (pp. 95-136).

Near, J. P., & Miceli, M. P. (2016). After the wrongdoing: What managers should know about whistleblowing. *Business Horizons, 59*(1), 105-114.

Paine, L. S. (1994). Managing for organizational integrity. *Harvard business review, 72*(2), 106-117.

Palazzo, G. (2007). Organizational integrity—understanding the dimensions of ethical and unethical behavior in corporations. *Corporate ethics and corporate governance*, 113-128.

Palmer, D. (2012). *Normal organizational wrongdoing: A critical analysis of theories of misconduct in and by organizations*. Oxford: Oxford University Press.

Park, H., & Lewis, D. (2018). The negative health effects of external whistleblowing: A study of some key factors. *The Social Science Journal*. doi: https://doi.org/10.1016/j.soscij.2018.04.002

PRINTEGER. (2016). Documents and Results. Retrieved 20-04-2016, 2016, from https://printeger.eu/documents-results/

Sacco, D. F., Bruton, S. V., & Brown, M. (2018). In Defense of the Questionable: Defining the Basis of Research Scientists' Engagement in Questionable Research Practices. *Journal of Empirical Research on Human Research Ethics, 13*(1), 101-110. doi: 10.1177/1556264617743834

Santoro, D., & Kumar, M. (2018). *Speaking Truth to Power: A Theory of Whistleblowing* (Vol. 6): Springer.

Sarewitz, D. (2016). The pressure to publish pushes down quality. *Nature, 533*(7602), 147-147.

Schulz, J. B., Cookson, M. R., & Hausmann, L. (2016). The impact of fraudulent and irreproducible data to the translational research crisis - solutions and implementation. *Journal of Neurochemistry, 139*, 253-270. doi: 10.1111/jnc.13844

Stitzel, B., Hoover, G. A., & Clark, W. (2018). More on Plagiarism in the Social Sciences*. *Social Science Quarterly*, n/a-n/a. doi: 10.1111/ssqu.12481

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research*: Sage publications.

Stroebe, W., Postmes, T., & Spears, R. (2012). Scientific Misconduct and the Myth of Self-Correction in Science. *Perspectives on Psychological Science, 7*(6), 670-688. doi: 10.1177/1745691612460687

Tijdink, J. K., Bouter, L. M., Veldkamp, C. L. S., van de Ven, P. M., Wicherts, J. M., & Smulders, Y. M. (2016). Personality Traits Are Associated with Research Misbehavior in Dutch Scientists: A Cross-Sectional Study. *Plos One, 11*(9), e0163251. doi: 10.1371/journal.pone.0163251

Treviño, L. J., Balkin, D. B., & Gomez-Mejia, L. R. (2017). How “doing gender” leads to gender imbalances in the higher ranks in colleges of business [and how to “undo gender”]. *Academy of Management Learning & Education, 16*(3), 439-453.

Vandekerckhove, W. (2016). *Whistleblowing and organizational social responsibility: A global assessment*: Routledge.

Yang, W. (2013). Research Integrity in China. *Science, 342*(6162), 1019-1019. doi: 10.1126/science.1247700

Chapter 9 - Organisational responses to cases of alleged scientific misconduct: Sensemaking, Sensegiving and Sensehiding

This chapter was published as:
Horbach, S. P. J. M., Breit, E., & Mamelund, S.-E. (2018). Organisational responses to alleged scientific misconduct: Sensemaking, sensegiving, and sensehiding. Science and Public Policy, 46(3): 415-429.. doi: 10.1093/scipol/scy068

9.1. Introduction

Several high-profile cases of misconduct have since the 1980s attracted substantial attention to the conduct and functioning of science (Guston, 1999). Ever since, science has witnessed an apparent increase in the number of, sometimes spectacular, cases of scientific misconduct. Scientists and science policy makers increasingly express concern about the effective functioning of the scientific enterprise (Fanelli, Costas, Fang, Casadevall, & Bik, 2017; Fanelli, Costas, & Lariviere, 2015; Hesselmann, Graf, Schmidt, & Reinhart, 2017; Wagner & Bates, 2016).

Whereas it was once thought that misconduct in science could hardly exists due to science's self-regulating mechanisms, scientists and policy makers alike have currently expressed their concern about the apparent frequency of reported cases of misconduct, which by some are considered only the 'tip of the iceberg' (Bozeman & Youtie, 2016; Fanelli, 2009; Hiney, 2015). This concern has led to a now substantial amount of literature on the determinants, consequences, costs and incidence of misconduct in science. In addition, the European concept of misconduct in science has been expressed more refined, now frequently encompassing not only the core misdoings in science - Fabrication, Falsification, Plagiarism (FFP) - but also including a variety of more subtle forms of potentially detrimental behaviour under the heading of *Questionable Research Practices (QRP)* (Horbach & Halffman, 2017b; B. Penders, Vos, & Horstman, 2009; N. H. Steneck, 2006). Commonly, these latter practices are considered the more injurious to the scientific enterprise, due to their apparent wide spread nature and systemic causes, such as pressure to publish and lack of social control (Bouter, Tijdink, Axelsen, Martinson, & ter Riet, 2016; B. C. Martinson, Anderson, & de Vries, 2005).

The growing concern over scientific misconduct has led to a substantial amount of research on the incidence, causes and patterns of misconduct (Faria, 2014, 2015; Hackett, 1994; Vaughan, 2002). However, while scholars increasingly acknowledge the importance for organisations to provide a healthy working climate that fosters research integrity (Brian C. Martinson, Thrush, & Lauren Crain, 2013), little is known about how universities and research organisations respond to cases of alleged misconduct. Researchers have a great deal of autonomy in their work, through the notion of "academic freedom"; yet, researchers are also employees, and they work in formal organisational surroundings, where universities, as employers, have responsibilities to act in situations of alleged misconduct. Even though we know at least implicitly that universities have such obligations, we lack knowledge on how they respond to alleged scientific misconduct and why they respond as they do. Even though some studies have addressed the issue of responses to scientific misconduct empirically (e.g. Bonito, Titus, & Wright, 2012; Mazur, 1989; Nicholas H Steneck, 1994) much uncertainty still exists about the specific mechanisms at play as well as their consequences (Ben-Yehuda & Oliver-Lumerman, 2017). In addition, previous studies on responses to alleged misconduct have been predominantly empirical in nature and have so far failed to add sufficient theoretical knowledge.

To address part of this knowledge gap, we use a comparative case-study approach to describe and assess the way in which four cases of alleged misconduct have been handled by their respective universities – two cases in the Netherlands and two cases in Norway. We have two main objectives: First to empirically explore and describe the four influential cases of alleged misconduct, including the response processes and the outcomes of the cases. Second, we aim to explain the reasons and rationale of mechanisms at the base of these processes and outcomes. We do so by adopting the concepts of organisational sensemaking, sensegiving and sensehiding (Degn, 2018; Gangloff, 2014; Weick, 1995) which address how organisations experience surprising events – such as alleged misconduct – and seek to restore status quo in the aftermath of the events. On the basis of our analysis, we propose a theoretical model that depicts organisational responses to misconduct as iterations between these three forms of social processes. We thus aim to contribute to knowledge about research misconduct by highlighting the impact of organisational behaviour on understandings and assessments of alleged misconduct.

9.2. Theoretical framework

9.2.1. Scientific misconduct

Debates over scientific misconduct are not new to science. For decades, scholars have been interested in the phenomenon, which, over the years, has gradually shifted from being a discussion on ‘fraud’ to one of integrity, misconduct and ethics (Ben-Yehuda & Oliver-Lumerman, 2017; Horbach & Halffman, 2017b). Within this debate, scientists and policy makers alike have distinguished between behaviours which are related to scientific integrity and those which are not: plagiarism usually is related to integrity, ‘self-plagiarism’ sometimes is, and behaviours such as harassment, or financial fraud usually are not. Even among those behaviours classified as misconduct, scholars distinguish between more or less severe practices. Manipulating data or tweaking statistics is considered particularly harmful in biomedical sciences and human services fields, with potentially life-threatening consequences (Montgomery & Oliver, 2017). Alternatively, plagiarism, duplication and text recycling is considered more severe in the humanities, because “the wording is the essence of the novelty” (Chrousos, Kalantaridou, Margioris, & Gravanis, 2012). In addition, classifying certain behaviour as misconduct or QRP allowed scientists to analyse the prevalence and causes of such behaviour, both in general terms as well as in specific research areas. Although much uncertainty still exists, some QRP’s are expected to occur in substantial extent (e.g. Horbach & Halffman, 2017a; John, Loewenstein, & Prelec, 2012), whereas the core sins of science, FFP, are thought to be less common (Fanelli, 2009).

The definition of misconduct has been anything but static. Norms and guidelines have shifted over the years as a result of institutional and organisational pressures (Montgomery & Oliver, 2009), shifting power relations (Martin, 1992) and novel detection abilities (Callahan, 2017). In the light of this shifting organisational landscape, with novel actors articulating and standardising norms and guidelines, it is of particular interest how universities, at the centre of these discussions, respond to cases of alleged misconduct (Montgomery & Oliver, 2009).

9.2.2. Organisational causes for misconduct

In order to study responses to misconduct in science, we distinguish three layers of the academic system that influence the emergence of misconduct: the environment of institutions (macro-level), organisational characteristics (meso-level), and the understandings and practices of individuals (micro-level). Although these three levels of analysis are often held separate, the literature shows that the social origin of nonconformitive behaviour is in the connections between them (Faria, 2015; Hackett, 1994; Vaughan, 2002). Two of the layers, the research environment and the individual behaviour of its practitioners, have received ample attention in the literature on scientific integrity and misconduct, leading to discussion about the 'publish-or-perish' dictum and analyses of psychological and demographical characteristics of suspected fraudsters (Anderson, Ronning, De Vries, & Martinson, 2007; Fanelli et al., 2017; Fanelli et al., 2015). However, the role of research organisations, including their characteristics and responsibilities, has only marginally been addressed.

Being notably absent in the discussion and analysis of misconduct in science, organisational characteristics have been widely studied in the context of other forms of misconduct, such as white-colour crime, corruption, financial fraud and medical errors (Ashforth, Gioia, Robinson, & Trevino, 2008; Greve, Palmer, & Pozner, 2010; Murphy & Dacin, 2011; Thurman, 2001; Trinkle, Phillips, Hall, & Moffatt, 2017; Vaughan, 1999). These studies demonstrate the influence of institutional characteristics on the occurrence of dubious behaviour and outline the ways in which the micro, meso and macro-levels interact to provide incentives, opportunities and motives for such behaviour. Among others, the literature highlights the importance of (the lack of) social control. Traditionally, science is characterised by a high extent of social control, which manifests itself through mechanisms such as peer review, reproducibility studies, collaborative work and numerous external oversight and regulating platforms and institutions (Collins, 1968; Leahey & Montgomery, 2011; Montgomery & Oliver, 2009). It has long been claimed that these mechanisms contribute to the self-regulating nature of science, i.e. that science naturally 'has' integrity (Merton, 1973; Zuckerman, 1977). However, due to the growth of and increased competition in science, social control is thought to be under threat (Zuckerman, 1984). This is considered one of the main explanatory factors for the occurrence of scientific misconduct (Faria, 2015; Vaughan, 2002). In order to understand the occurrence of organisational misconduct, other factors have also been studied. They include the presence of regulatory frameworks, the extent to which organisations are hierarchically structured, and the perceived level of managerial pressure on individual scholars (Greve et al., 2010; Murphy & Dacin, 2011; Vaughan, 1999, 2002; Zuckerman, 1977).

9.2.3. Organisational responses to research misconduct: A sensemaking perspective

On top of the above, there is also an extensive literature that has focused on organisational responses to misconduct, albeit that this literature has only marginally focussed on the academic or scientific context. In this literature, a central theory is that of organisational sensemaking (Gangloff, 2014; Gioia & Chittipeddi, 1991; Weick, 1995; Weick, Sutcliffe, & Obstfeld, 2005), which highlights how organisations experience and respond to extraordinary events such as accidents, crises, or, in our context, alleged misconduct. It comprises three interrelated concepts of sensemaking, sensegiving, and sensehiding.

The concept of sensemaking has been described as ‘the ongoing retrospective development of plausible images that rationalise what people are doing’ (Weick, 1995; Weick et al., 2005). Violations of commonly held expectations of acceptable or ‘normal’ behaviour create a discrepancy that demands organisations – such as universities – to attach meaning to these events (Gangloff, 2014; Weick et al., 2005), thereby answering the basic questions of ‘what does this [incident] mean?’ and ‘what should I [or we] do next?’ (Gangloff, 2014). Rather than focusing on isolated responses, sensemaking involves a “continued redrafting of an emerging story so that it becomes more comprehensive” (van Vuuren, 2012; Weick et al., 2005). In the light of misconduct, sensemaking serves as a primary mechanism to establish a common understanding of appropriate behaviour (Greve et al., 2010; Palazzo, 2007).

Sensemaking is not solely an individual endeavour occurring in a vacuum of personal interpretations. On the contrary, actors, both individuals as well as organisations, may use strategic actions aimed at influencing, nudging, or forcing others to make sense in a particular way. This is called ‘sensegiving’ and refers to attempts to guide the ‘meaning construction of others toward a preferred redefinition of organisational reality’ (Gioia & Chittipeddi, 1991; van Vuuren, 2012). As such, sensegiving operates in a reciprocal relationship with sensemaking (Gangloff, 2014; Rouleau, 2005). For example, in cases of alleged misconduct, organisations might implicitly or explicitly engage in practices of sensegiving in order to exploit the information asymmetry between the organisation and outside spectators, thus shaping external evaluations of the discussed research practices into desirable directions.

Two main forms of sensegiving have been outlined – explanatory framing (i.e. what the organisation says) and corrective action (i.e. what the organisation does) (Gangloff, 2014; Williams & Benford, 2000). Giving sense through explanatory framing occurs by concentrating information through highlighting some aspects or punctuating certain clues (Williams & Benford, 2000). Examples include the actors stating to have ‘no other choices’, or stressing the fact that no uncommon deviation from common practices has occurred. Second, sensegiving through corrective actions involves intentions to prevent repetition of the non-conforming behaviour (Maitlis & Lawrence, 2007; Monin, Noorderhaven, Vaara, & Kroon, 2013). Examples may include the execution of certain sanctions, the establishment of new procedures or the installation of new equipment. Overall, a consistency between the explanatory framing and corrective action – i.e. “walking the talk” – has been documented as most effective in order to restore relationships with external stakeholders (Gangloff & Connelly, 2015).

Constitutive of the practice of sensegiving, is the notion of ‘sensehiding’ (Monin et al., 2013; van Vuuren, 2012). By performing activities of sensegiving, one inevitably engages in processes of distorting and manipulating images through holding back particular aspects or cues. By leaving out these specific aspects, an actor aims to create a favourable image or meaning. In such cases, it is of primary interest to whom such image or meaning is favourable. Employing the organisation’s status over those of individual participants involved in the case of alleged misconduct, the practice of sensehiding might be a particularly effective way of influencing (public) perceptions in the aftermath of the case (Gangloff, 2014).

In our analysis, we use the sensemaking, sensegiving and sensehiding concepts to answer the question of how organisations respond to alleged cases of scientific misconduct. We see the three concepts as interrelated and mutually constitutive. We thus examine the extent to which the concepts, which have predominantly been applied in management studies of organisational misconduct, apply in an academic setting. Examples of central themes that we expect to be relevant also in organisational sensemaking of scientific misconduct includes the types of behaviour that are ‘made sense of’ in the first place; the discrepancy between formal statements and actual practice in organisations’ responses to the alleged misconduct (Degn, 2018; MacLean & Behnam, 2010); and the tendency to reduce reputational damage rather than facilitating effective learning practices (Coombs, 2007; Davies & Olmedo-Cifuentes, 2016). Thereby we address the knowledge gap concerning how cases of alleged scientific misconduct are responded to and we extend the theory of sensemaking as a response to misconduct to the setting of universities.

9.3. Methodology

9.3.1. Selection of cases

In our analysis we adopt a comparative case study design (Eisenhardt & Graebner, 2007; Yin, 2013) to assess the four cases of alleged misconduct. The cases were theoretically sampled to extend extant theory based on different characteristics. First, the cases were chosen from two different countries, the Netherlands and Norway. Both countries have relatively well-established procedures regarding research integrity and scientific misconduct; they may in fact be regarded as forerunners in the field of scientific integrity policy within Europe. This is witnessed among others by the existence of a legal framework regarding scientific misconduct in Norway, defining misconduct in legal terms and outlining sanctions for transgressing those. At the time of selecting our cases, (early 2016) Norway was the only European country to have such legislation defining misconduct and its sanctionability, even though this was not yet enforced when one of the described cases took place. Later, other countries, including Denmark, also introduced national legal frameworks regarding research misconduct. Another example demonstrating Norway's and the Netherland's proactive stance regarding research integrity is the existence of a national committee on scientific integrity (*LOWI*) and explicit guidelines regarding transparency concerning cases of alleged misconduct in the Netherlands. However, similar initiatives were taken elsewhere as well, including in Denmark, the UK and Croatia. Furthermore, both countries provide publicly available information on these and other cases of alleged misconduct. For example in the Netherlands, the national association of universities collects anonymised case reports on alleged cases of scientific misconduct that were handled by institutional integrity committees. All Dutch universities are obliged to participate in this and annually send in their reports. The same holds true for Norwegian research institutes since the installation of the new legislative framework. This relative openness about organisational handlings of alleged cases of misconduct in both countries allowed us to perform detailed case studies.

Second, from each country, we selected one 'black-and-white' case of relatively clear instances of misconduct, and one case within the 'grey' area of scientific integrity's spectrum. This selection thus involved a theoretical distinction between FFP in the black-and-white cases and forms of QRP in the grey area cases. Furthermore, the black-and-white case involved high-profile academics (and universities) as well as extensive media coverage, whereas the grey area case have involved more 'average' scientists with a lower academic status and have generally stayed under the radar of public and media attention. In so doing, the theoretical sampling enables us to extend our findings across national contexts and types of misconduct. Due to privacy issues and constraints, we are cannot reveal the identities of the directly involved or any of our interviewees in the 'grey' cases.

9.3.2. Data collection

The data used in the analysis are publicly available documents, supplemented with interviews (see appendix A). The documents include the official allegations; investigation reports both from local and national authorities (universities and national bodies on research integrity); newspaper articles; and public letters and statements by the involved parties.

The document analysis was complemented by nine face-to-face and five telephone interviews with actors involved. In all cases, attempts were made to interview the accused, but all declined or simply did not respond to our repeated invitations. The interviews were conducted in Dutch and Norwegian respectively, lasted for approximately one hour and were tape-recorded upon interviewee approval to allow for detailed analysis. The interviewees in both Norway and the Netherlands were given the opportunity to comment on drafts of the case-analyses and all agreed to the manuscript's final draft. The interviews have been anonymised and the interviewees will not be identified in this chapter.

We used content analysis (Krippendorff & Bock, 2009) to establish the narrative of the cases. We did so by reading the publicly available documents and identifying involved actors; form and origin of the allegation; official processing of the allegation; conclusions of official procedure; and responses to the allegation and conclusions of official procedure. Information from the interviewees were used to complement and triangulate information from the available documents as well as to illustrate personal consequences and viewpoints from involved actors. Interview reports were coded along the above topics, i.e. actors, allegation, processing, conclusions, and responses. In the interview reports, we also searched for commonalities and differences in rationales and motives of involved actors to act in the ways they did. The comparative nature of the case study approach manifests itself in this latter aspect.

By sampling cases from a diverse background we aimed to find commonalities and differences in their structure and the motives of the involved actors, thus engaging in 'asymmetrical comparison' (Krause, 2016). This has helped us provide a detailed understanding of how cases of alleged misconduct are dealt with and the common patterns entailed in such cases.

9.4. Case descriptions

9.4.1. Dutch case 1

This case involves the Dutch emeritus professor in regional economics and economic geography of the Vrije Universiteit Amsterdam (VU), Peter Nijkamp, and his PhD-student Karima Kourtit. Several anonymous allegations were filled against them between May 2013 and June 2014. Additionally, allegations of self-plagiarism were put forward by a national newspaper in January 2014. These allegations were processed by four local integrity committees at VU, with each committee looking into one allegation. In two instances, the accused appealed their cases at the national committee on scientific integrity (LOWI) which handled both cases. Nijkamp is a particularly prominent scientist within the Dutch research system, being a former president of the governing board of the Netherlands Research Council, chairman of the Dutch Social Science Council (SWR) and vice-president of the Royal Netherlands Academy of Sciences (KNAW). On top, Nijkamp has become widely known as one of the most productive scholars in his field. From 1975 onwards he published over 2300 scientific articles and more than 100 edited volumes.

9.4.1.1. The organisational context

Prof. Nijkamp was a full professor at the department of Spatial Economics at VU. The department holds 60 staff members who are involved in both fundamental as well as national and international commissioned research. RePEc ranks the Department in the top 5% of the world in multiple domains of spatial economics (VU, 2017).

The Department of Spatial Economics is part of the faculty of Economics and Business Administration. The faculty's webpage explicitly mentions the importance of scientific integrity. In addition, it explicitly refers to the codes of conduct by VSNU (VSNU, 2012) and ALLEA (ESF/ALLEA, 2011) that faculty members are to adhere to. Furthermore, it has a reference to a university wide 'Academic Integrity Complaints Procedure' (although the document cannot be found on the universities webpage), to faculty specific trustees and it mentions the university-wide committee on scientific integrity.

Within the department there is, officially, a clear focus on the quality of research, rather than the quantity. This shows for example in the fact that research resources are distributed over the department members on the basis of their top-5 publications from the past 5 years, rather than their total amount of publications (Zwemmer, Gunning, & Grobbee, 2015) (Interviewee 2,3). This system was implemented prior to the start of Nijkamp-case and is still in place. Contrary to this, Prof. Nijkamp's colleagues indicate that some members of the department, including Prof. Nijkamp himself, had a clear focus on quantity anyway: "prof. Nijkamp was a good professor and a good scientist. He completely dedicated himself to his research, but he particularly focussed on its quantity, rather than its quality. This mainly showed when he participated in large European projects. […] In such cases he was very eager in forming many (or several) journal articles from one project report." (Interviewee 2)

Concerning the position of Nijkamp within the department, several of the interviewees note that it is best described as an 'island-like structure' in which prof. Nijkamp gathered a group of researchers who interacted only very limited with the rest of the department. Both this fragmented structure and the prevailing hierarchy, in which Nijkamp is described as an incontestable leader within his sub-group, gave rise to minimal levels of social control (Interviewee 1, 2, 3).

9.4.1.2. The allegations

The allegations by the anonymous whistleblower (who used the pseudonym N.N.) concern (self-) plagiarism in the PhD-thesis by Karima Kourtit, largely co-authored by Nijkamp; plagiarism in journal articles (co-)authored by Nijkamp; and data fabrication and manipulation within Nijkamp's work (LOWI, 2015, 2016; Struiksma, van Mil, & Widdershoven, 2015; Zwemmer et al., 2015). On top, the national newspaper NRC put forward an allegation of self-plagiarism within large parts of Nijkamp's oeuvre. Subsequently, the case invoked several public allegations against Nijkamp in the form of newspaper articles, blog posts and commentaries (e.g. Remmie, 2014; Verbon, 2016).

9.4.1.3. The responses

The VU installed several committees to officially handle the filled allegations. Three ad hoc committees were installed as a response to the official allegations put forward by N.N. to the VU university board. The fourth ad hoc committee was charged with the investigation into citation practices in Nijkamp's work, thereby investigating the newspaper allegations of self-plagiarism. In addition, two of the committees' conclusions were appealed at LOWI, yielding yet another two investigations into the allegations.

The committees come to diverse and sometimes contradictory conclusions. Most notably, LOWI repeatedly comes to different, milder conclusions that the ad hoc committees installed at VU. Because of the high number of committees and resulting investigation reports, we will not go into great depth into all conclusions, but rather limit the discussion to the core findings and the organisational responses to them. For more details we refer to the (public) investigation reports presented at the VSNU and LOWI webpages (LOWI, 2015, 2016; Struiksma et al., 2015; VSNU, 2015; Zwemmer et al., 2015).

The conclusions of the first committee stating that plagiarism was found in the doctoral thesis of Karima Kourtit, led the VU to postpone her public defence and give her the opportunity to resubmit her dissertation after rewriting the alleged passages.

The second VU committee finds plagiarism in several articles co-authored by Nijkamp. After appeal, the national integrity committee, LOWI contradicts this statement by stating that no clear intention to deceive can be found in the re-use of some text fragments and hence concludes that no plagiarism has occurred. These findings did not revoke any official response by the VU.

The third committee installed by the VU, investigating allegations of data fraud, concludes that no clear signs of data fraud can be found within the alleged publications, amongst others because no 'intentional misleading to obtain an advantage' could be identified.

The findings of the previous committees, most notably the first and second VU committee, as well as the newspaper allegations of self-plagiarism in Nijkamp's work, led the VU to install a fourth ad hoc committee to study the citation practices in Nijkamp's oeuvre. The committee concludes that 'systematic copy-pasting' within Nijkamp's work and qualifies this as 'Questionable Research Practices', intended to 'lead to a high number of publications, rather than an original oeuvre' (Zwemmer et al., 2015). After requesting a second opinion at LOWI, the national integrity committee concluded that the fourth VU committee did not use appropriate methods to assess Nijkamp's citation practices and hence that the conclusions by the VU committee are poorly grounded.

The findings of the committees in turn sparked various debates, most notably in the media, resulting in dozens of newspaper articles, blog posts and comments, as well as (inter)national debates on the acceptability of 'self-plagiarism' or text recycling (Horbach & Halffman, 2017a). This in turn resulted in novel regulations and guidelines on citation practices and text recycling (KNAW, 2014).

The directory board of the VU consents with the findings of the Committee but does not approve of the judgement that Nijkamp committed 'Questionable Research Practices', because this, according to the directory board, can only apply to individual publications (instead of an entire oeuvre) and because no clear regulations and norms regarding self-plagiarism were articulated at the time of publication of the alleged articles. The directory board specifically focussed on prof. Nijkamp, rather than to include his co-authors in the investigation, because the allegations in the newspaper did so. The investigation was aimed as clarification of the adopted practices by Nijkamp and, in case the results of the investigation allowed for this, to his name (interviewee 1). Because no other co-authors were accused of committing any undue practices, the directory board did not feel the need to clear their names and did not intend to raise any discussion on them (Interviewee 1).

Succeeding the publication of Committee 4's final report, the co-authors affiliated to the VU were individually informed about the Committee's findings by the Rector (Interviewee 1 and 2). One of the involved co-authors describes this as: "We were invited to visit the rector magnificus. He gave a sermon of ten minutes with almost no room for interaction or discussion […]. All in all, we as co-authors felt that the VU was not there to protect us. It did not feel as a safe environment and my confidence in the institute has dropped dramatically" (Interviewee 2). While co-authors complain about the lack of a protective environment, others pointed to co-authors' responsibilities within the publication process and the fact that none of them signalled prof. Nijkamp's potential dubious behaviour (Interviewee 1).

The VU decided not to take any measures against Prof. Nijkamp and did not further investigate the individual publications on plagiarism. During the case, Prof. Nijkamp reached the age of retirement and was made an emeritus professor in January 2015. In this role, he is still able to work at the VU and continue his scientific career.

9.4.2. Dutch case 2

This second Dutch case report concerns the allegations of plagiarism in, and poor scientific quality of, an external doctoral student's dissertation at Erasmus University Rotterdam (EUR) filled in 2013. The alleged culprit was a female doctoral student working in the consultancy sector and writing her dissertation at the Rotterdam School of Management on the topic of leadership models. Her Promotor, who was targeted in the allegations for not properly executing his role as supervisor, held a part-time professorship at EUR, where he acted as supervisor for a total of ten PhD-students, the majority of which were external PhD-students.

The case was handled by various committees at EUR and went to appeal at LOWI. Interestingly, the case stirred a debate about the possibility to retreat a doctoral degree after it has been rewarded by a university, a process of which no antecedents are known in the Netherlands. Despite the discussion of these rigorous measures, the case received only very limited (media) attention and mostly went by unnoticed.

9.4.2.1. The organisational context

Doctoral student X. was affiliated as an external doctoral student with the Erasmus University in Rotterdam. Here she was supervised by Dr. Z. and Promotor Y. from the Erasmus Research Institute of Management (ERIM). In this, the daily supervision was in hands of Dr. Z., while Promotor Y. acted as the official promotor (Interviewee 8). The process of writing and performing the research was mainly supervised by Dr. Z, while all results were subsequently shared with Promotor Y. to allow him to give feedback on all produced material (Interviewee 8). Together, Dr. Z. and Promotor Y. supervised between 25 and 30 external doctoral students during their careers (interviewee 6). In addition, prof. Y. was president of the exam committee at the Rotterdam School of Management.

As external doctoral student, doctoral student X. was only limited embedded within the department. She did not physically spend time at the department, nor did she cooperate with its members other than her supervisors (Interviewee 5, 6 and 8). Because she, as external doctoral student, was not enrolled in any of the training activities at Erasmus University, rules and regulations regarding scientific integrity were expected to reach doctoral student X. via her supervisor.

9.4.2.2. The allegations

On November 22nd 2013, Dr. B. filled an official allegation of plagiarism in Doctoral student X.'s dissertation to prof. dr. Pols, Rector of Erasmus University Rotterdam (Basten, 2013). According to Dr. B. large parts of the thesis have been copied from other sources without proper reference to them. The whistleblower finds at least thirteen sources ranging from Wikipedia articles to other PhD-theses. In several cases, Doctoral student X. does mention the use of a source in the beginning of a paragraph, but later copies large parts of the source without showing that the text was not her own. The allegation comprises a detailed list of parts of A's dissertation and the sources from which they have allegedly been copied.

In addition, the whistleblower mentions that two members of the doctoral committee and Doctoral student X.'s promotor, Promotor Y., were involved in the supervision of two theses from which material was plagiarised. This leaves the whistleblower to wonder whether the supervisors were aware of plagiarism in A's dissertation and whether intend to deceive is in play (Basten, 2013).

On top of pointing to plagiarism in Doctoral student X.'s thesis, the whistleblower considers the thesis to be of extremely low scientific quality. However, she decides not to include statements on this subject in her allegations because she did not feel like having discussions on the content of the thesis and considered the amount of plagiarism so overwhelming that discussion on quality were no longer needed in the light of rules on scientific integrity so flagrantly being broken (interviewee 7). In later stages of the allegation's handling by integrity committees, the scientific quality of the dissertation nevertheless came to play a central role.

9.4.2.3. The responses

EUR responded to the allegation by installing an ad-hoc integrity committee to investigate the allegation. Basing itself on the code of conduct by EUR (EUR, 2013), KNAW (KNAW, 2014), VSNU (VSNU, 2012) and ALLEA (ESF/ALLEA, 2011), the committee concludes that the allegation is grounded and that doctoral student X's dissertation indeed contains plagiarised material. The amount of copied text without proper citation is of such extent that one cannot speak of honest error or negligence. Therefore, according to the committee, the EUR has unjustly provided the doctor-degree to Doctoral student X. Concerning the supervision, the committee concludes that the promotor, Promotor Y., has provided insufficient support to his student and has been imputably inadequate (Doelder, Bensing, & Bosch, 2014).

These conclusions led the committee to advice the EUR directory board to repeal their decision of rewarding Doctoral student X. a PhD-degree. In addition, it advices the directory board to investigate the other theses written under supervision by promotor Y. Apart from that, the committee makes several recommendations on policy adjustments to prevent similar cases of fraud in future projects (Doelder et al., 2014).

Commenting on the report by the integrity committee, the directory board of the EUR states that it does not fully consent with the findings of the committee (College van Bestuur EUR, 2014). Although the board acknowledges that substantial amounts of text have been copied, it asserts that clear rules about what constitutes plagiarism are lacking. In addition, the directory board feels that the supervision of Doctoral student X. was insufficient. Finally, the majority of the plagiarised text concerns parts of the introduction, theoretical framework and methodology sections, rather than the core of the results or conclusions. Given these contextual circumstances, the directory board feels that repealing the decision to provide Doctoral student X. a PhD-degree would be too grave of a sanction. Instead, the directory board decided to reprimand Doctoral student X. and to demand her to rewrite the plagiarised parts of her thesis and provide proper citations. The directory board consents with the committee's advice to investigate the other PhD-theses written under promotor Y.'s supervision.

Following the decision by the EUR's directory board to provide Doctoral student X. a second chance to rewrite her thesis, the whistleblower appealed the case at national integrity board, LOWI. LOWI concludes that Doctoral student X. committed a severe form of plagiarism that cannot be attributed to negligence or mistake. The LOWI Committee advices the directory board to reconsider the imposed sanctions on Doctoral student X. (LOWI, 2014).

Nevertheless Doctoral student X gets the chance to rewrite her thesis and denude it from plagiarism. She takes on the job. Subsequently, EUR installs a second committee, not referred to as an 'integrity committee', tasked with judging the rewritten version of A's dissertation and testing the other theses written under supervision of Promotor Y. on plagiarism. The committee concludes that the resubmitted dissertation is free of plagiarism but is of too poor scientific quality to warrant a doctoral degree. In addition, it concludes that the other theses written under supervision of promotor Y. do not contain significantly more overlap with other work, than the theses in a control group (Leeflang, Dijk, & Bunt, 2015).

Basing itself on the report by Committee 2 and the LOWI Committee, the EUR directory board made a final decision in the Doctoral student X.-case on June 25th 2015 (College van Bestuur EUR, 2015). The final decision targets Doctoral student X., her promotor and general aspects of EUR policy for doctoral students. The directory board decides to:

- Make the (anonymised version of the) final report by Committee 2 publicly available;
- Repay the government subsidy obtained after the PhD-defence of Doctoral student X.;
- Not store any hardcopy or digital version of Doctoral student X.'s thesis in the university library;

- Advice other universities to remove (the original version of) Doctoral student X.'s thesis from their libraries.
- Promotor Y. cannot act as promotor of a doctoral candidate, nor can he become member of a doctoral or manuscript committee judging the quality of a PhD-thesis at the EUR.
- Every internal and external doctoral candidate should write an educational and supervision plan for his/her doctoral study;
- Every doctoral student should be supervised by at least two staff members;
- More severe demands will be set on the composition of doctoral committees judging the quality of a thesis;
- Every doctoral thesis should be subject to a plagiarism detection scan and the results of the scan should be analysed in context (NB such a scan was performed in this case prior to submitting the first version of the thesis, without flagging unacceptable duplication);
- Every doctoral student should acknowledge to be aware of the code of conduct for scientific practices at the EUR (College van Bestuur EUR, 2015).

Because there are no set means to repeal Doctoral student X.'s degree, the directory board decided to request Doctoral student X. to voluntarily renounce her degree, but to not take any legal steps in demanding her to do so. Doctoral student X has not yet followed up on this request.

This case remained relatively 'below the radar', yielding only minor media attention (Kolfschooten, 2015) and public discussion. Nevertheless, the measures taken by EUR did have some implications for national policy on research integrity and supervision of (external) doctoral students. This will be discussed more elaborately in section 5.

9.4.3. Norwegian case 1

The first Norwegian case concerns Jon Sudbø, who was an academic super-star until a whistleblower suggested that the patient data used in his 2005 Lancet article did not yet exist. An investigation into Sudbø's record of accomplishment, showed that he had fabricated data in at least half of his scientific articles including his PhD. Sudbø lost his job and was revoked his PhD and authorizations as a doctor and a dentist.

9.4.3.1. The organisational context

Sudbø worked within the disciplines of medicine and odontology. He was an oral cancer researcher at the Rikshospitalet-Radiumhospitalet. Sudbø was relatively young (45 years) medical doctor and a dentist when he was accused of scientific fraud in 2006. He had top grades (best in his odontology class), with an impressive CV and several large research grants. One of his last grants was from the National Cancer Institute in the USA, which he got in March 2004 together with distinguished American cancer researchers. In an interview with the

Norwegian newspaper Aftenposten, Sudbø tells “about periods of intense engagement where work is all-embracing and that the research gives the same kick as paragliding and diving. I'm probably a risk-taker and curious person” (Kluge, Hafstad, & Torp, 2006). Sudbø was an independent lone wolf ever since he started his PhD in 1993.Few had insight into his work. Lack of (social) control over his work and a competitive culture therefore also explain why he was able to carry on with his fraudulent behaviour without questions for more than a decade.

9.4.3.2. The allegations

The allegations in Sudbø’s case were filled by Camilla Stoltenberg, M.D., a researcher and Director at the Division of Epidemiology at the Norwegian Institute of Public Health, who had studied a new article by Sudbø in the October 2005 issue of Lancet (Sudbø, Lee, et al., 2005). She noticed that the data used was from an epidemiological database (CONOR) that did not exist at the time of the supposed data collection.

On the 10th of January 2006, Sudbø and his PhD advisor met with the Cancer registry, where also Professor Vatten from the University of Trondheim was present. Sudbø verbally admitted to his employer on 12th of January 2006 that he had fabricated the patient data (Ekbom, 2006). The Radium hospital informed the Lancet and the press that Sudbø had manipulated the data in his article the day after. The Lancet itself first expressed a concern about the article on January 21st (Horton, 2006a) and retracted the article on February 4th (Horton, 2006b) after Anders Ekbom, the leader of the investigation committee, confirmed in a letter to the Lancet that it contained fabricated data. The New England Journal of Medicine expressed a similar concern on two other papers by Sudbø on February 9th (Curfman, Morrissey, & Drazen, 2006a), but awaited the results from the investigation committee before they retracted both papers on November 2nd 2006 (Curfman, Morrissey, & Drazen, 2006b; Sudbø et al., 2001; Sudbø et al., 2004).

9.4.3.3. The responses

On the 18th of January 2006, an independent investigation committee was appointed by Rikshospitalet – Radiumhospitalet Medical Center and the University of Oslo to investigate the admitted fraud and determine the role of the co-authors (among others Sudbø’s wife and twin brother, and prominent cancer researchers from the USA and Finland) and whether Sudbø’s prior work was fraudulent (Ekbom, 2006). The commission was led by the Swedish epidemiology professor Anders Ekbom, at Karolinska hospital in Stockholm, Sweden. The committee also included statisticians, researchers and staff of the Norwegian Institute of Public Health, the Norwegian Research Council, Cancer Registry of Norway, and the Cancer Clinic at the Radium Hospital.

All of Sudbø's scientific work from 1993-2006 was investigated in a report of almost 150 pages (Ekbom, 2006). In total 60 scientists from 6 different countries and 38 scientific papers were investigated. 15 of Sudbø's 38 articles, including parts of his doctoral work as well as articles in Lancet, New England Journal of Medicine and Journal of Clinical Oncology were considered fraudulent due to manipulation and fabrication of patient data.

The investigation committee documented that large part of the patient data were manipulated, for example: 1) several were fictitious persons with inserted birth dates; 2) data from a single patient were reused; and 3) half of the patients had already been diagnosed with oral cancer before or at the same time as the leukoplakia was diagnosed. These latter patients could not be studied for later development of cancer, since they already had cancer. In addition, in the October 2005 Lancet article, Sudbø seemed to have used a database including 908 subjects (Sudbø, Lee, et al., 2005). It turned out that these data came from a database which did not yet exist and that the data once again were fabricated with 250 subjects with identical birthdays (Ekbom, 2006). Sudbø also admitted partly to the investigating commission (Ekbom, 2006) that he had fabricated data in an article in Clinical Oncology (Sudbø, Samuelsson, et al., 2005).

Responding to the committee's findings, Sudbø was the first to have his authorizations as a physician and dentist revoked because of scientific fraud in Norway. Moreover, he was the first to have his doctoral thesis revoked at the University of Oslo (Haug, 2007). In addition, he was fired from Rikshospitalet-Radiumhospitalet and and his 20% position at the Medical Faculty, University of Oslo.

The committee reprimanded Rikshospitalet-Radiumhospitalet for a lack of control on Sudbø's projects, lack of educational measures in the area of research ethics, and lack of routines for dealing with misconduct. As a response, Rikshospitalet-Radiumhospitalet introduced regulatory systems for securing better institutional control over their research activities, among other things, the introduction of larger research groups (Harboe, 2006).

The Sudbø-case undoubtedly increased the awareness of research misconduct not only in the health care sector in Norway, but also at other research institutions, colleges and universities. Several institutions have introduced more elaborate supervisory and regulatory systems to monitor research programmes and routines (Nylenna, 2007). Mandatory courses in research ethics have also been introduced and several institutions have made their own local research integrity guidelines.

The Sudbø-case was an eye-opener for the Norwegian authorities. Although a discussion on research integrity had been going on for several years, a new law on research ethics was finalized quicker than anticipated, in 2007. The law requires institutions to bear the responsibility of regulating research misconduct. In the wake of the new law, the National

Commission for the Investigation of Research Misconduct ("Granskningsutvalget") and the regional ethics committees for medical research also saw its birth and became statutory (Nylenna, 2016; Tavare, 2011). Moreover, scientific dishonesty was given a legal definition in the form of "falsification, fabrication, plagiarism and other serious breaches of good scientific practice committed intentionally or grossly negligently in planning, conducting or reporting research".

We found 315 matches 2006-2017 searching for "Jon Sudbø" in Norwegian newspapers. As most of these articles were mainly commenting on the allegations and the conclusions from the Investigation report from 2006 (Sudbø, Lee, et al., 2005), we have for the most part not cited them here. However, the number of matches shows that this was a high media profile case in Norway, and this was also the case internationally. The Sudbø-case also sparked an international debate on peer-review, the internal quality-control process and co-authorship.

9.4.4. Norwegian case 2

The second Norwegian case concerns a low media profile 'grey case' of alleged plagiarism, in which all involved actors were working at the same Christian University College in Bergen, Norway. The case was 'grey' for two reasons. First, the case shows that there were tensions between using ethics guidelines and ethics legislations to judge whether or not a case of plagiarism had occurred. Second, the case shows that there might be diffuse distinctions between processing and handling a case of workplace conflict and a case of research misconduct (Forskerforum, 2015; Komite for vurdering av klagesak ved NLA Høgskolen, 2014; Nasjonal utvalg for granskning av redelighet i forskning, 2015; Sævik, 2015a, 2015b).

9.4.4.1. The organisational context

During this case of alleged scientific misconduct, the whistleblower and the accused were both affiliated with the private Christian NLA University College with 2000 students and 200 employees. The College was founded in 1968 and has, in 2017, campuses in Bergen (1600 students), Kristiansand (200 students) and in Oslo (300 students).

NLA University College is the only private college in Norway offering primary school teacher education. This education is offered in Bergen, where also the kindergarten teacher education is offered together with bachelor and master's degrees in pedagogics, intercultural understanding and theology / practical theology and management.

The actors involved in the alleged case of misconduct were stationed at the Bergen campus in the field of pedagogics. The whistleblower worked at the department for pedagogics (Sandviken) and the accused at the department for pre-school teacher education (Breistein). The accused is a lecturer in pedagogics for bachelor and master students. The accused is a mid-career researcher with a master's degree in pedagogics and an unfinished PhD who has published three book-chapters in Norwegian anthologies on the topic of pedagogic creed for teachers. The whistleblowers are two lecturers in pedagogics, one for pre-school teacher students, the other for both pre-school teacher and primary school teacher students. Both were mid-career researchers. They are co-editors of one Norwegian anthology on pedagogic creed for kindergarten teachers, published August 2015: two of the chapters in this anthology were included in the later search for plagiarism.

The departments at Sandviken and Breistein are located 15 kilometers apart. The two units merged in 2010. Sandviken has a more academic profile, hosting more staff with PhDs and professors, and publishes more in academic journals compared to Breistein, which is more involved in teaching and applied science. The NLA University College is private, they are therefore not obliged to report openly to the public.

9.4.4.2. The allegations

In February 2014, the whistleblower at Sandviken unit in Bergen accused two colleagues at the Breistein unit for plagiarism, violation of good reference practice and for "uncollegial" behaviour. The whistleblower claimed that the two colleagues had plagiarized the concept "pedagogic creed" in 5 unpublished drafts, including two chapters meant for a Norwegian anthology (later published in 2015), and an abstract, paper and a power-point presentation used at an international conference (Forskerforum, 2015; Nasjonal utvalg for granskning av redelighet i forskning, 2015).

9.4.4.3. The responses

The leadership of the NLA University College first sought to resolve the matter between the parties internally. When this effort did not lead to a solution, it was decided to appoint a local expert committee to assess the alleged scientific misconduct. Two of the appointed members were employed at the University of Bergen and a third member was a retired professor still associated with the NLA University College.

The internal committee concluded that the accused had plagiarized the whistleblower, that they had violated good reference practices and behaved "uncollegially". The internal committee used *the Copyright act* (1961) and *the National Committee for Research Ethics guidelines in the Social Sciences and the Humanities* from 2006 (NESH) for supporting their argument that this was a case of plagiarism. Finally, the local commission stated that the institution had a responsibility for preventing scientific research misconduct in the future; this should be achieved in courses in research ethics, other follow-up and in establishing a sufficient organizational culture (Forskerforum, 2015; Nasjonal utvalg for granskning av redelighet i forskning, 2015).

The management at the NLA University College decided to take note of the committee's statement and lay the foundation for further follow-up of the case. This led to the delay of a book publishing and disciplinary warnings for the two accused employees. The NLA management concluded that there were grounds for termination of the accused's positions, but allowed them to hold their position if they would accept certain conditions. These included the accused to relate to general standards for research ethics in the field of pedagogics and in the institution, and follow a ban on sending e-mails or publish any documents on the case to internal or external colleagues. The accused disagreed with the conclusions, quit their jobs in 2015 and pointed to what they considered were law suiting- and procedural errors (Komite for vurdering av klagesak ved NLA Høgskolen, 2014; Nasjonal utvalg for granskning av redelighet i forskning, 2015). The accused henceforth appealed to the National Committees for Research Ethics (Forskerforum, 2015).Using the definition of the Law of Research Ethics, the national review committee concluded that there was no serious breach of good scientific practice and that the accused had not acted fraudulently. In the cases of similarity in text (in PPT-presentations), the investigators had recited text from the study guide at NLA Breistein, rather than copied the text from the whistleblower. Setting the question of plagiarism aside, the national committee criticized the local report for procedural errors and the NLA University College for a lack of routines for dealing with alleged cases of scientific misconduct and education on research ethics (Forskerforum, 2015; Nasjonal utvalg for granskning av redelighet i forskning, 2015).

As a direct consequence of the case, NLA published new ethics guidelines and new ideals and guidelines for handling and processing of alleged scientific misconduct. Their webpage now also includes a Scientific Misconduct Notification Form as well as links to web-sources on research integrity. Additionally, the institution published new guidelines for the prevention and management of conflicts between employees. Lastly, in the aftermath of the case, NLA initiate several courses and workshops related to scientific integrity and research ethics.

There were no national or long-reaching consequences of this case. The case had a low media profile. It was not covered by national media. We found a few newspaper reports made by the Bergen-based newspapers "Dagen" (19 and 21 November 2015) and "Bergens Tidende" (9 September 2017) and one online report made by "Forskerforum", a journal for the members of the labour union Forskerforbundet (19 November 2015). These articles mainly described the content of the two investigation reports.

9.5. Organisational responses

In this section, we analyse the organisational responses based on the framework of sensemaking, sensegiving and sensehiding described earlier. The findings are summarised in the following table:

	Dutch case 1	Dutch case 2	Norwegian case 1	Norwegian case 2
Sensemaking	- Several committees installed - Committees came to different conclusions	- Several committees installed - Ad hoc committees and national committee came to different conclusions	- One in-depth investigation committee involving 60 researchers from different countries	- Two committees installed - They came to different conclusions
Sensegiving	- New guidelines on how to handle misconduct allegations - New guidelines on text recycling - Enhancing (social) control	- Sanctioning involved actors - New guidelines to enhance (social) control - Explanatory framing: lack of supervision	- New control systems at the University - Larger research groups - New national legislation on scientific misconduct	- New ethical guidelines - New guideline for handling and processing alleged scientific misconduct - New ethics courses and workshops to employees
Sensehiding	- Focus on individual (Nijkamp)	- Focus on specific characteristics (external doctoral student) - Discussion on quality rather than integrity	- Focus on individual (Sudbø)	- Discussions of scientific misconduct were inflicted with a workplace conflict and cultural differences due to a recent merger of former separate research organisations

Table 9.1 Overview of sensemaking, sensegiving and sensehiding processes in the cases studied

9.5.1. Sensemaking: Ad hoc committees

As shown in table 9.1, sensemaking in all of the cases involved establishing committees tasked to investigate the allegations of misconduct. These committees were established on an ad-hoc basis, and generally had little experience regarding such procedures and went through relatively short learning curves. The cases thus demonstrate the existence of an ad hoc nature of investigating and making sense of allegations of scientific misconduct.

On the one hand, this ad-hoc nature showcases the novelty that allegations of scientific misconduct may impose on universities, as they have no such established structures in place. Hence, they can be regarded as fair attempts of dealing with complex cases. On the other hand, the ad-hoc committees might create several problems and difficulties. These include: disregarding (fair) procedure and law; being unaware of antecedents and (reports from) earlier cases; or failing to provide disclosure or coming to a final conclusion. In the latter situation, this can lead to the installation of new (ad hoc) committees. In several of the cases, we witness these new committees coming to diverging or even contrasting conclusions to those of prior committees, resulting in general confusion and uncertainty about the acceptability of (questionable) research practices. The usage of different definitions of misconduct by different committees might even give the appearance of committees being less than perfectly impartial or building to pre-desired conclusion (Interviewee 6).

The investigations of the ad hoc committees are thus a prime endeavour in the sensemaking process concerning the cases. When allegations arrive, the 'normal' is interrupted and investigation reports serve as a verbal account making sense of the situation and explicating the tensions, potential dilemmas and solutions. The ad hoc-ness of the committees seems to underscore the processual and unfolding nature of the alleged misconduct's sensemaking processes. This is exemplified in diverging assessments by different integrity committees, on the basis of different legislative frameworks (as in the Norwegian case 2) or new emerging information. In addition, both the Netherlands and Norway have a national research integrity office, which was involved in all of the described cases. These committees, which are permanent rather than ad hoc, demonstrate more in-depth knowledge of procedures and guidelines and henceforth often come to contrasting conclusions than their local counterparts. These contrasts further highlight the sensemaking processes.

These unfolding series of sensemaking events are processual as they fail to provide disclosure (Weick et al., 2005). This lack of closure leads to general confusion and particularly hinders the learning process of both involved actors as well as other organisations. This is particularly problematic since scholars have identified clarity of norms as crucial to a work atmosphere that promotes integrity in organisations: "Shared understandings of what is right and wrong, allowed and forbidden, desirable or undesirable set the normative context in which members of an organisation interact" (Palazzo, 2007).

Part of the sensemaking endeavour concerns the articulation of what should or should not be considered a problem of integrity. Whereas questions of integrity and quality were addressed simultaneously in Dutch case 2 and the Norwegian case 2, the involved committees processing the allegations in other cases specifically stated that they would 'only assess those aspects of the allegations concerning scientific integrity'. This requires the committees to explicitly demarcate between issues of quality and issues of integrity, which is in itself a highly problematic distinction in the light of contemporary discussion about integrity in science (Horbach & Halffman, 2017b; Bart Penders, 2017; B. Penders et al., 2009).

9.5.2. Sensegiving: Articulation of procedures and guidelines

All of the discussed cases led to the articulation of new procedures or guidelines, either on a local, national or even international level. The formulation of these procedures and guidelines is a clear example of sensegiving processes based on preventative actions (Gangloff, 2014). Regarding the institutional level, these new guidelines or procedures often aim to prevent similar cases to happen in the future (e.g. requiring more supervision on PhD-students, new guidelines to prevent work floor conflicts, etc.) or to describe how future allegations of misconduct should be handled (e.g. no longer accepting anonymous allegations, installing research integrity officers, etc.). In addition, in some cases (Dutch cases 1 & 2, and Norwegian case 2) the level of social control was actively elevated. For instance in Dutch case 1, measures were taken to making both ongoing and completed research more systematically accessible across colleagues through sharing publications via newsletters and organising departmental seminars more systematically and at higher frequency. These changes were not exclusively made because of the Nijkamp case, but lessons learned from the case were taken into account.

Through the articulation of new procedures, the cases provide a basis for organisational and institutional learning. Whereas new organisational procedures and guidelines were prominent in all cases, the Norwegian case 1 is explicitly regarded as a central precursor to the national legislation on scientific misconduct. The alterations in policy are in line with recommendations made in literature about organisational characteristics nourishing misconduct. In particular, social control, or the lack thereof, is put forward as one of the most prominent factors playing a role at the intersection of the micro- and meso-level of organisational structures. Lack of social control creates opportunities, required to engage in dubious behaviour (Faria, 2015; Hackett, 1994; Murphy & Dacin, 2011; Vaughan, 2002). This is particularly applicable to our cases, since the organisational characteristics and settings show poor signs of social control, either due to the specific (high hierarchical) status of the accused or the general work practices within the involved departments. Hence, novel policy, such as requiring more supervision for PhD-candidates, aiming at increasing (formal) social control might be effective ways of reducing the extent of dubious behaviour in the affected organisation (Faria, 2015; Mishkin, 1988; Murphy & Dacin, 2011; Trinkle et al., 2017; Vaughan, 1998).

In addition to aiming at the prevention of future cases of misconduct, the corrective actions serve as a sensegiving mechanism by conveying a zero-tolerance attitude regarding misconduct in the affected organisation. The involved universities express not to tolerate misconduct by taking a stand against the behaviour shown.

On top, Dutch case 2 showcased an example of explanatory framing in which the role of the supervisors was put forward as a main reason for the alleged behaviour of external doctoral student X. Here again, we see that a lack of social control is framed as the cause of trouble, whereas simultaneously, individuals are held responsible for this task. This framing may serve as a sensegiving mechanism, pointing attention in desired directions (Gangloff, 2014). However, the focus of specific individuals is more effectively described as a sensehiding mechanism, to which we will point our attention in the next section.

All of these organisational responses serve as sensegiving mechanisms; they help to (re)create a certain image of the situation or the organisation, such as that of zero-tolerance or immediate organisational retaliation. However, as we will discuss in more depth in the next section, the images did not necessarily conform to reality or actual practices. Indeed, various involved actors acknowledge that the organisational responses were in fact only rarely put in practice. Therefore, the responses serve as formal (sensegiving) reactions, but whose practical implications seem to be limited.

9.5.3. Sensehiding: Containment and individualisation

A reoccurring pattern in the cases concerns the (apparent) goal or rationale of the universities' responses to misconduct. The organisational response generally involves a risk that the problem with respect to part of the issue or allegation may be contained (e.g., through argumentation such as 'this is a debate about quality rather than integrity', 'this is a work floor conflict, not an issue of scientific integrity'). In addition, the response may also contain the problem to the individual scientist or a few colleagues. The alleged behaviour is thus framed along the 'rotten-apple argument' (or 'special person argument') which may distance it from the common or representative practices within the organisation. These types of responses are examples of sensehiding processes to the extent that they conceal the potentially more structural issues and thereby close or minimise discussion – and consequently learning – of the disputed cases.

In the Dutch case 1 no explicit sanctions were put forward against actors. But measures in general were taken: both within the department, for example by raising levels of social control, and on a national level, through the KNAW (The Royal Netherlands Academy of Arts and Sciences) guidelines on text recycling, new initiatives were put forward to regulate scientific practices regarding 'self-plagiarism'. As one of the informants said: "Therefore, there was no reason to formulate, on top of these general measures to discourage incentives for self-plagiarism, additional rules or regulations, or to put more emphasis on the commitment to the existing guidelines regarding self-plagiarism than the emphasis that was already induced by discussions on the ongoing case and on the newly formulated rules." (Interviewee 3). In the Norwegian case 2, the discussions of integrity and misconduct were interwoven in complex ways with other discussions, such as a work conflict between several of the involved researchers, an ongoing merger of two previously separate research institutions, and different cultural assessments of research integrity in the new merged organisation. In this complexity, it was difficult for the university college to make sense of the issue of integrity in particular.

Further, sensehiding also involve a diminishing of the need for further measures, especially due to the rotten-apple argument:

"On top of measures to eliminate possible incentives for the undesirable re-use of own texts, additional measures specifically directed to individuals' numbers of publications per se, ultimately the key undesirable possible consequence of self-plagiarism, did not seem very useful. Professor Nijkamp's way of working, with very many publications and many co-authors was unique and not representative for others at the department. To install new rules now that are especially targeted to possible risks of that way of working, and to prevent cases exactly like this one, is not very useful. There are currently no people at the department who work in that way. Nijkamp was in that respect a very unique colleague, in the most objective and neutral sense of the word" (Interviewee 3).

This reasoning, indicating the unicity of a specific person's work practices, reoccurs in several of the other cases. For example, in Dutch case 2, it was argued that little general measures needed to be taken, because the alleged practices where supposedly due to the specific background of the individual actors involved (an external doctoral student, with very limited contact with the rest of the department). Similar arguments and reasoning were used in the Norwegian Sudbø case. These examples, which stress the individual behaviour and characteristics of the alleged researchers, may potentially involve sensehiding processes to the extent that they obfuscate organisational responsibility for the issue. Even though the statement that work practices were (highly) specific to a certain individual might be legitimate, the question may be raised whether stressing this is an effective way of dealing with a case. Indeed, the rotten-apple metaphor stretches further: a rotten apple may contaminate its neighbours. An effective learning process may prevent such contaminations. Contrary to the formulation of guidelines as described in the previous section, stressing the individuality of specific behaviour may hinder the learning process and lead to limited change within the organisation.

9.6. Discussion

There are some key characteristics of the cases that seem to drive the sensemaking processes and outcomes. We highlight five characteristics. First, we note that allegations of misconduct accusation relate in complex ways to various forms of social tension. Therefore, processing allegations frequently involves discussion on aspects of science not directly related to scientific integrity. Within the cases described above, this includes aspects related to personal disagreements, work floor conflicts (such as in Norwegian case 2) and indignation over standards or quality of scientific work. Most notably, both Dutch cases involved allegations that targeted not only the integrity, but to a large extent also the quality of the alleged work.

Second, all cases demonstrated a lack of existing guidelines and institutionalised systems for handling the cases – as seen not only in the number of ad-hoc committees established, but also the variety of interpretations. Even in cases apparently involving clear-cut questions related to integrity in science, such as allegations of plagiarism, the standards against which these allegations are to be judged are often unclear and contentious. An example of this is Norwegian case 2 where the national research committee strongly disagreed with the local investigating committee on standards of plagiarism and the sanctionability of their transgressions. This potentially gives rise to further discussions about non-integrity related aspects of science and publication.

Third, a common feature of the universities' handling of the cases was a fear of reputational damage and the (potential) involvement of the media. Actors persuasively argue that most severe consequences arrived in the form of reputational and emotional damage. For all actors involved, this seems to be the threat that is most feared. This holds both for individuals, who see their career opportunities and personal contacts being damaged, as well as for organisations, for which the fear of reputational damage seemed to be one of the driving forces behind their willingness to take (seemingly strong) measures. In this respect, the role of the media should not be underestimated.

Frequently, actors acknowledged that the fact that the case would (or could) be discussed in the media, and hence be publicly visible and known, was a major trigger for the organisations' responses. In many cases, action in such cases would be primarily tailored at damage control and image management (e.g. Interviewee 1, 6 and 7). From the organisational perspective, individualising the case serves as an effective strategy to channel the reputational damage from the institution to the individual scientist. On the contrary, individuals often felt unprotected by their institution and hence reported on a diminished level of trust in their employer (e.g. Interviewee 2, 6, 8). Additionally, it may have led individuals to seek media attention in order to engage in sensegiving practices through explanatory framing. This can for instance be witnessed in the Nijkamp and Sudbø cases, where both actors gave newspaper interviews in order to defend their case (Gjerding & Utheim, 2013; Nijkamp, 2015). The importance or threat of media coverage for the involved organisations suggests that organisational responses are primarily tailored at sensegiving processes of explanatory framing (Gangloff, 2014). The next section elaborates on this perspective.

Fourth, in all cases, the organisation's response includes corrective actions in the form of several policy alterations and the installation of new procedures or guidelines. While this is laudable from a perspective of increased social control and quality assurance, the organisational responses apparently did not always meet up to the official statements communicated through press releases or official reports. Therefore, the responses have a high sensegiving content, i.e. as images of corrective actions, but arguably only limitedly serve as effective mechanisms in preventing future cases scientific misbehaviour. Actors within the cases acknowledge that in practice 'little has changed' since the case. One of the main reasons for this should be sought in the usage of the rotten-apple argument, containment and individualisation of the cases. While the organisational responses on the one hand provide a statement demonstrating the institution's zero-tolerance policy regarding scientific misconduct, it is also acknowledged that, since the case involved a 'particularly uncommon set of circumstances, not representative of the institution's common practices', there is little need to make drastic alterations to the everyday research practices or environment. The lack of changes taking place after the cases was acknowledged by several actors and subsequently deemed "disappointing". "I had hoped that more would have changed afterwards." "I am not so much disappointed because of a lack of official sanctions, but because of a lack of discussion. In my opinion, too little discussion has been going on. This could have been much more elaborate" (interviewee 2).

As such, we see a clear discrepancy between how the organisations talk about the cases and how they act upon them, i.e. between the explanatory framings and the corrective actions. The announcement of new procedures and the simultaneous lack of changes in practice clearly point to an amplified gap between formal structures and actual work practices. Due to these mixed processes of sensegiving through explanatory framing or corrective action and processes of sensehiding, a process of decoupling is exposed (MacLean & Behnam, 2010; Meyer & Rowan, 1977). In other words, the organisations fail to 'walk the talk'.

Fifth, the failure to 'walk the talk' point to organisational responses apparently mainly serve to restore symbolic order. Restoring symbolic and channelling reputational damage point to processes of legitimation (Maitlis & Lawrence, 2007; Monin et al., 2013; Thurman, 2001), and to the focus on universities to engage in forms of image management and damage control that are very similar to that of the private sector (Coombs, 2007). This raises questions on the rationale and motives of universities to respond to misconduct in the particular ways witnessed in our case studies, especially in the light of discussion on the changing financial incentives for universities and the introduction of new public management strategies to higher education and research (Andersen & Pallesen, 2008; De Boer, Enders, & Schimank, 2007; Halffman & Radder, 2015).

In the process of restoring symbolic order, institutions regularly tend to shift the focus from institutional or organisational causes of alleged misconduct to individual, micro-level causes. This should be considered a technique of neutralisation, redirecting the responsibility of an organisation towards an individual (Agnew, 1994; Minor, 1981; Thurman, 2001). Misconduct is hence framed as an issue of individual psychopathology, a theory generally regarded as the least satisfying to explain scientific misconduct by social scientists (Hackett, 1994; Mishkin, 1988), contrary to theories pointing to anomy and alienation as systemic causes of misconduct (Merton, 1938; Zuckerman, 1977). Accordingly, framing misconduct as rooted in individual causes distorts the institutional responsibility to act upon it.

Concluding, we witness processes of sensemaking, sensegiving and sensehiding in play at various stages of processing and responding to alleged cases of scientific misconduct. These processes mutually reinforce each other, resulting in a complex constellation: the sensemaking activities by ad-hoc committees serve as the base for the sensegiving mechanisms of announcing sanctions and new policies, which on their turn feed into the sensehiding mechanisms of channelling reputational damage. Conversely, the identified sensegiving processes potentially nourish (future) processes of sensemaking by suggesting boundaries of which issues are to be thought of as issues of integrity and which are not.

9.7. Conclusions

Our study analyses the organisational responses to allegations of scientific misconduct, thereby addressing a knowledge gap in the literature on scientific integrity, which tends to focus on either micro- or macro-level processes. Our analysis adds to the literature in two ways. First, by providing an empirical analysis of organisational responses and their outcomes in four influential cases of alleged misconduct. Second, by providing a theoretical model comprising of the interrelations between sensemaking (the committees), sensegiving (e.g. the new regulations) and sensehiding (e.g. the individualistic focus) in how universities handle allegations of misconduct (Degn, 2018; Faria, 2015; Weick, 1995).

As suggested in the theory on organisational misconduct, several aspects of the sensemaking, sensegiving and sensehiding processes identified in our analysis hinder effective learning processes within organisations and the involved individuals. First, the ad hoc committees tasked to investigate the allegations of misconduct, thereby engaging in the prime sensemaking activity, suffer from a lack of experience. Occasionally this leads to repetition of the sensemaking process, resulting in confusion over standards, definitions and good or acceptable practices. Second, the discrepancy between the announced establishment of novel procedures and the perceived consequences of their implementation contributes to decoupling speech and action, which has been asserted as a hurdle to effective learning in response to misconduct (Gangloff & Connelly, 2015; MacLean & Behnam, 2010). Third, the process of sensehiding through containment and individualisation of integrity issues distorts the organisation's responsibility of handling the case and incentivises actors to stick to current practices.

Similar processes of sensemaking, sensegiving and sensehiding have been identified in commercial or private organisations (Gangloff, 2014; Gangloff & Connelly, 2015; Thurman, 2001; Vaughan, 1999), as well as some of the intentions and consequences of these processes (Coombs, 2007; Davies & Olmedo-Cifuentes, 2016; MacLean & Behnam, 2010). Our analysis thereby provides an initial step in the endeavour of extending the theory on responses to misconduct from a commercial context to universities and the academic context.

9.8. References

Agnew, R. (1994). The techniques of neutralization and violence. *Criminology, 32*(4), 555-580.

Andersen, L. B., & Pallesen, T. (2008). "Not just for the money?" How financial incentives affect the number of publications at Danish research institutions. *International Public Management Journal, 11*(1), 28-47.

Anderson, M. S., Ronning, E. A., De Vries, R., & Martinson, B. C. (2007). The perverse effects of competition on scientists' work and relationships. *Science and Engineering Ethics, 13*(4), 437-461.

Ashforth, B. E., Gioia, D. A., Robinson, S. L., & Trevino, L. K. (2008). Re-viewing orgnizational corruption. *Academy of Management Review, 33*, 670-684.

Basten, F. (2013). Sorry, wetenschap, ik heb verzaakt #plagiaat. Retrieved from https://www.floorbasten.nl/sorry-wetenschap-ik-heb-verzaakt-plagiaat

Ben-Yehuda, N., & Oliver-Lumerman, A. (2017). *Fraud and Misconduct in Research: Detection, Investigation, and Organizational Response*: University of Michigan Press.

Bonito, A. J., Titus, S. L., & Wright, D. E. (2012). Assessing the preparedness of Research Integrity Officers (RIOs) to appropriately handle possible research misconduct cases. *Science and Engineering Ethics, 18*(4), 605-619.

Bouter, L. M., Tijdink, J., Axelsen, N., Martinson, B. C., & ter Riet, G. (2016). Ranking major and minor research misbehaviors: results from a survey among participants of four World Conferences on Research Integrity. *Research Integrity and Peer Review, 1*(1), 17.

Bozeman, B., & Youtie, J. (2016). Trouble in Paradise: Problems in Academic Research Co-authoring. *Science and Engineering Ethics, 22*(6), 1717-1743. Retrieved from http://dx.doi.org/10.1007/s11948-015-9722-5. doi:10.1007/s11948-015-9722-5

Callahan, J. L. (2017). The retrospective (im)moralization of self-plagiarism: Power interests in the social construction of new norms for publishing. *Organization, 25*(3), 305-319. Retrieved from https://doi.org/10.1177/1350508417734926. doi:10.1177/1350508417734926

Chrousos, G. P., Kalantaridou, S. N., Margioris, A. N., & Gravanis, A. (2012). The 'self-plagiarism'oxymoron: can one steal from oneself? *European journal of clinical investigation, 42*(3), 231-232.

College van Bestuur EUR. (2014). *Besluit CvB naar aanleiding van melding plagiaat – 12 juni 2014*. Retrieved from Rotterdam:

College van Bestuur EUR. (2015). *Definitief besluit College van Bestuur naar aanleiding van melding plagiaat*. Retrieved from Rotterdam: http://www.eur.nl/fileadmin/ASSETS/press/2015/Juli/Besluit_CvB_plagiaat_proefschrift_29062015.pdf

Collins, R. (1968). Competition and social control in science: An essay in theory-construction. *Sociology of Education*, 123-140.

Coombs, W. T. (2007). Protecting Organization Reputations During a Crisis: The Development and Application of Situational Crisis Communication Theory. *Corporate Reputation Review, 10*(3), 163-176. Retrieved from https://doi.org/10.1057/palgrave.crr.1550049. doi:10.1057/palgrave.crr.1550049

Curfman, G. D., Morrissey, S., & Drazen, J. M. (2006a). EDITORIALS: Expression of Concern: Sudbø J et al. DNA Content as a Prognostic Marker in Patients with Oral Leukoplakia. N Engl J Med 2001;344:1270-8 and Sudbø J et al. The Influence of Resection and Aneuploidy on Mortality in Oral Leukoplakia. N Engl J Med 2004;350:1405-13. *The New England Journal of Medicine, 354*(6), 638. Retrieved from http://www.nejm.org/doi/pdf/10.1056/NEJMe068020.

Curfman, G. D., Morrissey, S., & Drazen, J. M. (2006b). Retraction: Sudbø J et al. DNA Content as a Prognostic Marker in Patients with Oral Leukoplakia. N Engl J Med 2001;344:1270-8 and Sudbø J et al. The Influence of Resection and Aneuploidy on Mortality in Oral Leukoplakia. N Engl J Med 2004;350:1405-13. *The New England Journal of Medicine, 355*(18), 1927-1927. Retrieved from http://www.nejm.org/doi/pdf/10.1056/NEJMe068247. doi:10.1056/NEJMe068247

Davies, G., & Olmedo-Cifuentes, I. (2016). Corporate misconduct and the loss of trust. *European Journal of Marketing, 50*(7-8), 1426-1447. Retrieved from <Go to ISI>://WOS:000386019100014. doi:10.1108/ejm-11-2014-0729

De Boer, H., Enders, J., & Schimank, U. (2007). On the way towards new public management? The governance of university systems in England, the Netherlands, Austria, and Germany. *New forms of governance in research organizations*, 137-152.

Degn, L. (2018). Academic sensemaking and behavioural responses – exploring how academics perceive and respond to identity threats in times of turmoil. *Studies in Higher Education, 43*(2), 305-321. Retrieved from https://doi.org/10.1080/03075079.2016.1168796. doi:10.1080/03075079.2016.1168796

Doelder, H. d., Bensing, J. M., & Bosch, F. A. J. v. d. (2014). *Rapport Commissie Wetenschappelijke Integriteit: Melding plagiaat in proefschrift - april 2014*. Retrieved from Rotterdam:

Eisenhardt, K. M., & Graebner, M. E. (2007). Theory building from cases: Opportunities and challenges. *Academy of management Journal, 50*(1), 25-32.

Ekbom, A. (2006). Report from the Investigation Commision appointed by rikshospitalet - Radiumhospitalet MC and the University of Oslo January 18, 2006.

ESF/ALLEA. (2011). *The European Code of Conduct for Research Integrity*. Retrieved from Strasbourg:

EUR. (2013). *Klachtenregeling Wetenschappelijke Integriteit EUR (KWI-EUR)*. Retrieved from Rotterdam: http://www.eur.nl/fileadmin/ASSETS/abd/Integriteitscode/KWI_EUR_121210_NL.pdf

Fanelli, D. (2009). How Many Scientists Fabricate and Falsify Research? A Systematic Review and Meta-Analysis of Survey Data. *Plos One, 4*(5), 11. Retrieved from <Go to ISI>://WOS:000266490000014. doi:10.1371/journal.pone.0005738

Fanelli, D., Costas, R., Fang, F. C., Casadevall, A., & Bik, E. M. (2017). Why Do Scientists Fabricate And Falsify Data? A Matched-Control Analysis Of Papers Containing Problematic Image Duplications. *bioRxiv*. doi:10.1101/126805

Fanelli, D., Costas, R., & Lariviere, V. (2015). Misconduct Policies, Academic Culture and Career Stage, Not Gender or Pressures to Publish, Affect Scientific Integrity. *Plos One, 10*(6), 18. Retrieved from <Go to ISI>://WOS:000356567400015. doi:10.1371/journal.pone.0127556

Faria, R. (2014). Science under pressure: problematic behaviors and social harms. *Revista Crítica Penal y Poder*(7), 64-84.

Faria, R. (2015). Scientific misconduct: how organizational culture plays its part. *Tijdschrift over Cultuur & Criminaliteit, 5*(1), 38-54.

Forskerforum. (2015). *anklager om plagiat på kristen høgskole*. Retrieved from Olso:

Gangloff, K. A. (2014). *Misconduct and the Media: Organizational Sensegiving, Media Coverage, and Investor Reaction in the Aftermath of Wrongdoing.* (Doctor of Philosophy), Auburn University, Auburn, Alabama.

Gangloff, K. A., & Connelly, B. L. (2015). *Walking the Talk: Organizational Sensegiving in the Aftermath of Misconduct.* Paper presented at the Academy of Management Proceedings.

Gioia, D. A., & Chittipeddi, K. (1991). Sensemaking and sensegiving in strategic change initiation. *Strategic management journal, 12*(6), 433-448.

Gjerding, M. L., & Utheim, E. B. (2013, 25.01.2013 (Accessed 04.09.2017)). Avslørt for forskningssvindel - nå vil han forske igjen. *Verdens Gang (Internet version)*. Retrieved from http://www.vg.no/nyheter/innenriks/helse/avsloert-for-forskningssvindel-naa-vil-han-forske-igjen/a/10106865/

Greve, H. R., Palmer, D., & Pozner, J. E. (2010). Organizations Gone Wild: The Causes, Processes, and Consequences of Organizational Misconduct. *The Academy of Management Annals, 4*(1), 53-107. Retrieved from https://doi.org/10.1080/19416521003654186. doi:10.1080/19416521003654186

Guston, D. (1999). *Between politics and science: assuring the productivity and integrity of research*. Cambridge: Cambridge University Press.

Hackett, E. J. (1994). A social control perspective on scientific misconduct. *The Journal of Higher Education, 65*(3), 242-260.

Halffman, W., & Radder, H. (2015). The Academic Manifesto: From an Occupied to a Public University. *Minerva, 53*(2), 165-187. Retrieved from https://doi.org/10.1007/s11024-015-9270-9. doi:10.1007/s11024-015-9270-9

Harboe, M. (2006). Dette er ikke vår medisinske etikk. *Tidsskrift For Den Norske Laegeforening: Tidsskrift For Praktisk Medicin, Ny Raekke, 126*(17), 2283.

Haug, C. (2007). Erfaringer, forventninger og forsetter. *Tidsskrift For Den Norske Laegeforening: Tidsskrift For Praktisk Medicin, Ny Raekke, 127*(1), 13.

Hesselmann, F., Graf, V., Schmidt, M., & Reinhart, M. (2017). The visibility of scientific misconduct: A review of the literature on retracted journal articles. *Current Sociology, 65*(6), 814-845. Retrieved from <Go to ISI>://WOS:000411555000002. doi:10.1177/0011392116663807

Hiney, M. (2015). *What it Means, Why it Is Important and How we Might Protect it*. Retrieved from Brussels:

Horbach, S. P. J. M., & Halffman, W. (2017a). The extent and causes of academic text recycling or 'self-plagiarism'. *Research Policy*. Retrieved from http://www.sciencedirect.com/science/article/pii/S0048733317301543. doi:https://doi.org/10.1016/j.respol.2017.09.004

Horbach, S. P. J. M., & Halffman, W. (2017b). Promoting Virtue or Punishing Fraud: Mapping Contrasts in the Language of 'Scientific Integrity'. *Science and Engineering Ethics, 23*(6), 1461-1485. Retrieved from https://doi.org/10.1007/s11948-016-9858-y. doi:10.1007/s11948-016-9858-y

Horton, R. (2006a). Expression of concern: non-steroidal anti-inflammatory drugs and the risk of oral cancer. *The Lancet, 367*(9506), 196-196. Retrieved from http://ac.els-cdn.com/S0140673606680148/1-s2.0-S0140673606680148-main.pdf?_tid=5bc02dca-99d8-11e6-902f-00000aacb35e&acdnat=1477306701_dd2350bec979b0cb68d578f5ae5acb4a. doi:10.1016/S0140-6736(06)68014-8

Horton, R. (2006b). Retraction—Non-steroidal anti-inflammatory drugs and the risk of oral cancer: a nested case-control study. *The Lancet, 367*(9508), 382.

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling. *Psychological Science, 23*(5), 524-532. Retrieved from <Go to ISI>://WOS:000314464500014. doi:10.1177/0956797611430953

Kluge, L., Hafstad, A., & Torp, I. S. (2006, 27 January 2006). Jon Sudbø ble avslørt som juksemaker. Men han har selv avslørt hvordan forskningen kan lure en hel verden. *Aftenposten A-magasin*.

KNAW. (2014). *CORRECT CITATION PRACTICE, Academy Advisory Memorandum*. Retrieved from https://www.knaw.nl/en/news/publications/correct-citation-practice?set_language=en

Kolfschooten, F. v. (2015, 30-06-2015). Erasmus Universiteit in Rotterdam vraagt psycholoog doctorsbul terug. *NRC Handelsblad*.

Komite for vurdering av klagesak ved NLA Høgskolen. (2014). *Uttalelse fra komite for vurdering av klagesak ved NLA Høgskolen*. Retrieved from Bergen:

Krause, M. (2016). Comparative research: Beyond linear-causal explanation. In J. Deville, M. Guggenheim, & Z. Hrdlickova (Eds.), *Practising Comparison* (pp. 45-67). Manchester: Mattering Press.

Krippendorff, K., & Bock, M. A. (2009). *The content analysis reader*: Sage.

Leahey, E., & Montgomery, K. (2011). The meaning of regulation in a changing academic profession. *The American Academic Profession: Changing forms and functions*, 295-311.

Leeflang, P. S. H., Dijk, H. K. v., & Bunt, H. G. v. d. (2015). *Rapportage adviescommissie beoordeling plagiaat*. Retrieved from Rotterdam: http://www.eur.nl/fileadmin/ASSETS/press/2015/Juli/Geanonimiseerd_Eindrapport_adviescommissie_plagiaat_RSM_24_april_2015_02.pdf

LOWI. (2014). *LOWI Advies 2014, nr. 10*. Retrieved from The Hague:

LOWI. (2015). *LOWI Advies 2015, nr. 2*. Retrieved from Amsterdam:

LOWI. (2016). *LOWI Advies 2016, nr.1*. Retrieved from Amsterdam:

MacLean, T. L., & Behnam, M. (2010). The dangers of decoupling: The relationship between compliance programs, legitimacy perceptions, and institutionalized misconduct. *Academy of management Journal, 53*(6), 1499-1520. Retrieved from <Go to ISI>://WOS:000286642700013. doi:10.5465/amj.2010.57319198

Maitlis, S., & Lawrence, T. B. (2007). Triggers and enablers of sensegiving in organizations. *Academy of management Journal, 50*(1), 57-84.

Martin, B. (1992). Scientific fraud and the power structure of science. *Prometheus, 10*(1), 83-98.

Martinson, B. C., Anderson, M. S., & de Vries, R. (2005). Scientists behaving badly. *Nature, 435*(7043), 737-738. Retrieved from <Go to ISI>://WOS:000229638700024. doi:10.1038/435737a

Martinson, B. C., Thrush, C. R., & Lauren Crain, A. (2013). Development and Validation of the Survey of Organizational Research Climate (SORC). *Science and Engineering Ethics, 19*(3), 813-834. Retrieved from https://doi.org/10.1007/s11948-012-9410-7. doi:10.1007/s11948-012-9410-7

Mazur, A. (1989). Allegations of dishonesty in research and their treatment by American universities. *Minerva, 27*(2-3), 177-194.

Merton, R. K. (1938). Social structure and anomie. *American Sociological Review, 3*(5), 672-682.

Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigations*: University of Chicago press.

Meyer, J. W., & Rowan, B. (1977). Institutionalized Organizations: Formal Structure as Myth and Ceremony. *American Journal of Sociology, 83*(2), 340-363. Retrieved from http://www.jstor.org/stable/2778293.

Minor, W. W. (1981). Techniques of neutralization: A reconceptualization and empirical examination. *Journal of Research in Crime and Delinquency, 18*(2), 295-318.

Mishkin, B. (1988). Responding to scientific misconduct due process and prevention. *Jama, 260*(13), 1932-1936.

Monin, P., Noorderhaven, N., Vaara, E., & Kroon, D. (2013). Giving sense to and making sense of justice in postmerger integration. *Academy of management Journal, 56*(1), 256-284.

Montgomery, K., & Oliver, A. L. (2009). Shifts in Guidelines for Ethical Scientific Conduct: How Public and Private Organizations Create and Change Norms of Research Integrity1. *Social Studies of Science, 39*(1), 137-155.

Montgomery, K., & Oliver, A. L. (2017). Conceptualizing Fraudulent Studies as Viruses: New Models for Handling Retractions. *Minerva, 55*(1), 49-64. Retrieved from http://dx.doi.org/10.1007/s11024-016-9311-z. doi:10.1007/s11024-016-9311-z

Murphy, P. R., & Dacin, M. T. (2011). Psychological pathways to fraud: Understanding and preventing fraud in organizations. *Journal of business ethics, 101*(4), 601-618.
Nasjonal utvalg for granskning av redelighet i forskning. (2015). *Uttalelse i sak om mulig vitenskapelig uredelighet ved NLA Høgskolen AS*. Retrieved from
Nijkamp, P. (2015). *'Zelfplagiaat'-rapport van de Commissie-Zwemmer: Broddelwerk over fictieve peanuts*. Retrieved from Amsterdam: http://www.advalvas.vu.nl/sites/default/files/dossiers/Peter%20Nijkamp%20uitgebreide%20verantwoording%20rapport%20Zwemmer%2017%20maart%202015.pdf
Nylenna, M. (2007). Research misconduct: lessons to be learned? *Michael, 4*, 7-9.
Nylenna, M. (2016). Ti år etter Sudbø-saken. *Tidsskrift For Den Norske Laegeforening: Tidsskrift For Praktisk Medicin, Ny Raekke, 136*(17), 1420.
Palazzo, G. (2007). Organizational Integrity — Understanding the Dimensions of Ethical and Unethical Behavior in Corporations. In W. C. Zimmerli, M. Holzinger, & K. Richter (Eds.), *Corporate ethics and corporate governance* (pp. 113-128). Berlin, Heidelberg: Springer Berlin Heidelberg.
Penders, B. (2017). Beyond Trust: Plagiarism and Truth. *Journal of Bioethical Inquiry*. Retrieved from https://doi.org/10.1007/s11673-017-9825-6. doi:10.1007/s11673-017-9825-6
Penders, B., Vos, R., & Horstman, K. (2009). A question of style: method, integrity and the meaning of proper science. *Endeavour, 33*(3), 93-98. doi:10.1016/j.endeavour.2009.07.001
Remmie, M. (2014). Vier voorbeelden van het (zelf)plagiaat van topeconoom Nijkamp. *NRC Handelsblad*.
Rouleau, L. (2005). Micro-practices of strategic sensemaking and sensegiving: How middle managers interpret and sell change every day. *Journal of Management studies, 42*(7), 1413-1441.
Sævik, T. H. (2015a, 21 November 2015). Arbeidstilsynet til kristen høgskole etter plagiatstrid. *Dagen*.
Sævik, T. H. (2015b, 19 November 2015). Plagiat-strid på kristen høgskole. *Dagen*.
Steneck, N. H. (1994). Research universities and scientific misconduct: History, policies, and the future. *The Journal of Higher Education, 65*(3), 310-330.
Steneck, N. H. (2006). Fostering integrity in research: Definitions, current knowledge, and future directions. *Science and Engineering Ethics, 12*(1), 53-74. Retrieved from <Go to ISI>://WOS:000235044700006. doi:10.1007/pl00022268
Struiksma, J., van Mil, J., & Widdershoven, J. (2015). *Adviesrapport CWI, Casus 2014-04 FEWEB, NN3 versus Nijkamp, Kourtit, Baycan*. Retrieved from Amsterdam:
Sudbø, J., Kildal, W., Risberg, B., Koppang, H. S., Danielsen, H. E., & Reith, A. (2001). DNA content as a prognostic marker in patients with oral leukoplakia. *New England Journal of Medicine, 344*(17), 1270-1278.
Sudbø, J., Lee, J., Lippman, S., Mork, J., Sagen, S., Flatner, N., . . . Zhou, X. (2005). RETRACTED: Non-steroidal anti-inflammatory drugs and the risk of oral cancer: a nested case-control study. *The Lancet, 366*(9494), 1359–1366.
Sudbø, J., Lippman, S. M., Lee, J. J., Mao, L., Kildal, W., Sudbø, A., . . . Risberg, B. (2004). The influence of resection and aneuploidy on mortality in oral leukoplakia. *New England Journal of Medicine, 350*(14), 1405-1413.

Sudbø, J., Samuelsson, R., Risberg, B. r., Heistein, S., Nyhus, C., Samuelsson, M., . . . Reith, A. (2005). Risk markers of oral cancer in clinically normal mucosa as an aid in smoking cessation counseling. *Journal of clinical oncology, 23*(9), 1927-1933.
Tavare, A. (2011). Managing research misconduct: is anyone getting it right? *BMJ : British Medical Journal, 343*. Retrieved from http://www.bmj.com/content/bmj/343/bmj.d8212.full.pdf. doi:10.1136/bmj.d8212
Thurman, A. (2001). Institutional responses to medical mistakes: ethical and legal perspectives. *Kennedy Institute of Ethics Journal, 11*(2), 147-156.
Trinkle, B. S., Phillips, T., Hall, A., & Moffatt, B. (2017). Neutralising fair credit: factors that influence unethical authorship practices. *Journal of Medical Ethics, 43*(6), 368-373. doi:10.1136/medethics-2015-103365
van Vuuren, M. (2012). Merger, She Wrote: Improvising on the Script of Identification Processes During Mergers. In *Advances in Mergers and Acquisitions* (pp. 39-53): Emerald Group Publishing Limited.
Vaughan, D. (1998). Rational choice, situated action, and the social control of organizations. *Law and Society Review*, 23-61.
Vaughan, D. (1999). The dark side of organizations: Mistake, misconduct, and disaster. *Annual review of sociology, 25*(1), 271-305.
Vaughan, D. (2002). Criminology and the sociology of organizations. *Crime, law and social change, 37*(2), 117-136.
Verbon, H. (2016). Peter Nijkamp (VU): verzon feitenvrij migratiebeleid. Retrieved from http://harrieverbon.blogspot.nl/2016/01/peter-nijkamp-vu-verzon-feitenvrij.html
VSNU. (2012). *De Nederlandse Gedragscode Wetenschapsbeoefening: Principes van goed wetenschappelijk onderwijs en onderzoek*. Retrieved from http://www.vsnu.nl/files/documenten/Domeinen/Onderzoek/Code_wetenschapsbeoefening_2004_(2012).pdf
VSNU. (2015). *VU Plagiaat - ongegrond 2*. Retrieved from http://www.vsnu.nl/wi-2015.html
VU. (2017). Department of Spatial Economics. Retrieved from http://www.feweb.vu.nl/nl/afdelingen-en-instituten/spatial-economics/index.aspx
Wagner, P. D., & Bates, J. H. T. (2016). Maintaining the integrity of peer review. *Journal of Applied Physiology, 120*(5), 479-480. Retrieved from <Go to ISI>://WOS:000371442900001. doi:10.1152/japplphysiol.00067.2016
Weick, K. E. (1995). *Sensemaking in organizations* (Vol. 3): Sage.
Weick, K. E., Sutcliffe, K. M., & Obstfeld, D. (2005). Organizing and the process of sensemaking. *Organization science, 16*(4), 409-421.
Williams, R. H., & Benford, R. D. (2000). Two faces of collective action frames: A theoretical consideration. *Current perspectives in social theory, 20*, 127-152.
Yin, R. K. (2013). *Case study research: Design and methods*: Sage publications.
Zuckerman, H. (1977). Deviant behavior and social control in science. *Deviance and social change, 1*, 87-138.
Zuckerman, H. (1984). Norms and deviant behavior in science. *Science, Technology, & Human Values, 9*(1), 7-13.
Zwemmer, J., Gunning, J. W., & Grobbee, R. (2015). *Rapport over verwijzingen in het werk van prof. dr. P. Nijkamp*. Retrieved from Amsterdam:

CONCLUSIONS & REFLECTION

Chapter 10 - Conclusions, discussion and open questions

10

In this dissertation, I have discussed the issue of research integrity from multiple perspectives. In three more or less independent parts, I set out to study problematic research practices and their consequences for the scientific literature. These parts describe how problematic research spills through the literature, how the editorial process at academic journals may filter problematic research, and how research performing organisations may clean-up after alleged cases of irresponsible practices have come to light.

Part I – *To Spill* elaborated on two forms of problematic or contentious research practices, those of text recycling and the usage of misidentified or contaminated research materials. Using scientometric techniques to trace articles and citations, and text similarity software to check for text recycling, it finds considerable levels of both types of problematic research practices in the academic literature. In particular, it demonstrated that the occurrence of these practices varies between research disciplines, with text recycling being more common in economics and psychology than in history or biochemistry, and between geographical locations, with articles on misidentified cells most commonly originating from countries with well-established research cultures, such as the US and Germany. Moreover, it finds indications for other factors influencing the amount of problematic research ending up in the academic literature, such as the number of co-authors of a paper and the productivity of these authors.

In part II – *To Filter* we turned our attention to the editorial peer review system in academic journals. First, chapter 4 described the emergence of peer review's current procedures by reviewing the scientific literature and by adding recent developments based on information from editors and publishers. It analysed the rationale for developing new review formats and discussed how they have been implemented in the current system, resulting in a systematisation of peer review forms. It showed how expectations of the editorial system have been in constant flux demonstrating that this leads to tensions in the current academic debate about the responsibilities and abilities of the peer review system.

Chapter 5 continued to assess the ability of different peer review formats to live up to one of the system's expectations, by correlating review format with retraction rates. Using the systemisation of review formats from chapter 4, survey information from journal editors, and the Retraction Watch database, it demonstrated that several review procedures are associated with significantly lower retraction rates. In particular, journals blinding author identities to reviewers, journals using pre-submission review and journals using digital tools such as plagiarism or statistics scanners had lower retraction rates. Since several peer review procedures were introduced with the express intent to prevent problematic publications, this suggests that in these journals lower retraction rates show enhanced peer review, not an unwillingness to correct errors.

In the two subsequent chapters, part II – *To Filter* specifically focused on innovation and implementation of editorial procedures. Using both quantitative data from a survey among journal editors and ethnographic data of field visits to the editorial offices of commercial publishers, it aimed to clarify the innovation process. These chapters addressed open questions regarding the implementation of new review procedures, the occurrence rate of various peer review procedures and their distribution over scientific disciplines or academic publishers, as well as the motivations for editors or publishers to engage in novel review procedures. They demonstrated that, in spite of enthusiastic innovation, the adoption of new peer review procedures is in fact slow, with the exception of text similarity scanners and some adoptions in specific niches of academic publishing. The ethnographic data subsequently identified several factors that are important in transforming editorial practices of commercial publishers. Both the hierarchical structure within the publishers and the strong focus on commercial interest were shown to be decisive factors in understanding why some innovations get implemented while others remain largely unused.

Last, part III – *To Clean*, presented studies on reporting and dealing with cases of alleged research misconduct. It continued on the role of hierarchy and power to suggest that power relations – in particular research seniority, temporary vs. permanent work appointments, and gender – may impact on the likelihood that cases are reported and that reporting leads to constructive organisational changes. Thereby it elucidated the processes that affect researchers' ability and willingness to report research misconduct, and the likelihood that research organisations constructively respond to this. Chapter 9 completed this part by studying organisational responses to cases of alleged misconduct. It demonstrated that these responses are characterised by processes of sensehiding, sensemaking and sensegiving. Specifically, universities responses commonly show a lack of formal guidelines and procedures to deal with misconduct cases, they are characterised by a fear of reputational damage, are influenced by (the possibility of) media coverage, and may include corrective action that seems to aim for restoration of symbolic order. Thereby, universities' responses closely resemble those identified in commercial or private organisations and arguably fail to provide learning potential from cases of alleged misconduct.

The conclusions from this last chapter were recently underlined by the work of Grey, Bolland, Gamble, and Avenell (2019) who find important deficiencies in universities' investigation rapports of academic misconduct as well. These authors studied four universities' investigation rapports, one in the US and three in Japan, on a single large case involving over 200 publications co-authored by an overlapping set of authors. Hence, while our study revealed shortcomings of integrity investigations by European universities, similar problems seem to be present in universities on other continents.

10.1. Recurring themes

The results and conclusions from the individual chapters present several recurring themes. These themes arguably constitute important facets of the answers to the research questions I posed in chapter 1 of this dissertation.

10.1.1. Complex interplay of values and practices

First and foremost, the chapters on editorial innovations and organisational responses to integrity issues clearly show that academic arguments, aiming to improve science and its ability to produce reliable knowledge, are not the sole arguments at play in debates on integrity. The studies in these chapters demonstrate that commercial interests of publishers, new public management sentiments in universities, and career considerations of individual scientists, are at least as important in making decisions about integrity issues. This demonstrates the perhaps unsurprising, though important lesson that not merely scientific values are at stake when dealing with integrity. Instead science is embedded in a large spectrum of complex practices, stakes and stakeholders. To deal with research integrity issues is hence not a question of addressing science alone, as if it acts in a vacuum, but rather requires solutions that address these wider interactions.

The studies in this dissertation explicate some of the values and interests that are at stake in integrity challenges. Besides, scientific values, these firstly include matters of reputation, which we have come across at several occasions. Chapters 8 and 9 discuss the reputation of organisations, while chapters 6 and 7 indicate the importance of journal reputation in issues of integrity. In addition, the reputation of individual researchers is constantly at stake throughout the various integrity issues discussed. In the interplay with other practices and values, these reputational arguments constitute a major factor determining how actors respond to specific cases. Stakeholders' perception of the potential damage due to integrity challenges, as well as their perception of the potential benefits of (seemingly) responding to or proactively acting against them, has repeatedly turned out to be of decisive influence on actors' behaviour. For example, this became clear in publishers' considerations about establishing research integrity teams and the tasks such teams should perform. In addition, organisational responses to cases of misconduct were typically characterised by damage control mechanisms, aimed at restoring symbolic order and protecting or repairing the organisation's reputation. This sometimes even turned out to be a more important factor in shaping their responses than the alignment of measures with current practices within the organisation or the protection of the individuals involved. Other non-scientific values turning out to be relevant in integrity matters include commercial values, equality and fairness among researchers with different backgrounds, and openness and transparency.

I should note here that some of these competing values have been discussed in the research literature before, mainly in the context of entrepreneurial science. This branch of science, following market logics and commercial motives on top of intellectual and epistemic motives of enhancing science or society, has been discussed for some decades now (Etzkowitz, 1998; Shimshoni, 1970). Often it is either represented as all virtue, with reference to helping the economy, making science relevant to human well-being, or tackling climate change; or it is represented as all vice, referring to the corruption of universities, distortion of science's objectivity, or ethical violations (Shapin, 2008). Rarely does the discussion on entrepreneurial or commercial aspects in science escape the vocabularies of moral value. In this dissertation, while being aware of the non-value-neutral metaphors used in both its title and section headings, I aimed to be more descriptive, showing how commercial interests may shape science and its institutions, rather than judging them on moral grounds.

Interestingly, integrity has not only been influenced by commercial arguments that seem to hinder the effective implementation of integrity initiatives, it has in itself also become prone to commercialisation. Chapter 7 already indicated how efforts to foster integrity might contribute to publishers' business model, hence allowing them to use it as a way of creating added value. In addition, new stakeholders are emerging that crucially depend on integrity issues for their existence and make those challenges into an effective business model. Among others, this includes the developers of text similarity scanners and new platforms providing peer review tailored to integrity issues (Research Square, 2017). Last, even some researchers and universities themselves are increasingly considering integrity as a 'commodity', reflected by their use of this term in blogs about authors requesting the retraction of one of their own papers after spotting errors in their analyses (Marcus, 2019).

Not only do alternative values come into play when discussing integrity issues, the studies in this dissertation demonstrate that those issues are commonly tied to other concerns, either intentionally or unintentionally. Most prominently, the demarcation between concerns of integrity and those of quality is commonly problematic. This became apparent in our discussions of text recycling in chapter 2, the usage of misidentified cells in chapter 3, efforts of peer review models to detect integrity issues in chapter 4, and the reporting and handling of alleged misconduct cases in chapters 8 and 9. Some of these discussions further clarified that the boundary between poor research quality on the one hand and research misconduct on the other might differ between research disciplines or epistemic cultures, as well as between different moments in history.

Besides their entanglement with concerns over quality, integrity issues commonly go hand-in-hand with wider issues such as personal disagreements and workplace conflicts. The discussions in part III – *To Clean*, indicate the difficulty of disentangling these issues and show the danger of misconduct allegations being used as a tool to settle other conflicts. This signals the need for fair and responsible procedures to handle misconduct allegations, both in research performing organisations and academic journals. However, our studies suggest that much is to be gained. Specifically, chapters 8 and 9 indicate that organisations commonly adhere to strategies of partitioning and demarcation, isolating research integrity issues from other matters. This may serve to avoid conflict escalation and provide containment of the issue. However, while this may be effective strategies to handle individual cases and minimize reputational damage, it commonly leaves systematic causes of integrity issues unaddressed.

10.1.2. Diversity of norms and standards

The sensitivity to diversity in integrity challenges and norms constitutes another major theme throughout this dissertation. The results of chapter 3 demonstrate how perceptions about proper and improper research behaviour vary between research disciplines. Inherent to the different forms of knowledge production, various research disciplines seem to value and assess similar behaviour differently. While reusing parts of one's own work hardly occurs in disciplines within the humanities we studied, it is much more common in economics and psychological research. As one commentator noticed: 'wording is the essence of the new' within the humanities, thereby suggesting why text recycling is deemed inappropriate in this discipline. Similarly, the results of part II – *To Filter* indicate that notions of quality assurance, expectations of peer review and evaluation mechanisms, as well as retraction policies differ across disciplines.

While variations between the disciplines of science have always existed and may not be problematic in and of themselves, they do create specific challenges within the contemporary research system. With a growing international expectation of interdisciplinary research and transdisciplinary collaborations, these differences present additional challenges to research integrity matters. For example, the strong focus on productivity and output related indicators in contemporary research evaluation and funding mechanisms, may skew credit and funding allocation when notions about the permissibility of text recycling strongly differ between disciplines. In addition, the inherent differences in standards of quality and relevance among researchers in various disciplines combined with an urge to standardise norms and ways of assessing research merits by others, including science policymakers and evaluators, may hamper interdisciplinarity and transdisciplinary collaborations.

Moreover, part III – *To Clean* indicated how differences in academic rank and seniority, gender, national background, and employment models may influence how cases of alleged misconduct are dealt with. It indicated that cases of several less clear-cut forms of research misbehaviour are less likely to be reported to relevant institutions. Similarly, cases witnessed by less powerful members of an organisation seem to be less likely to be reported, as well as that they less often lead to constructive (organisational) changes. Handling of cases by institutional or national integrity committees has repeatedly led to the clarification and articulation of integrity norms and standards, as well as formed the starting point for addressing systematic causes of misbehaviour. Therefore, the fact that some cases are less likely to be reported hinders the potential of effectively tackling integrity issues. Combined, these factors further stress the need to address integrity challenges in their context, being sensitive to cultural, epistemic and methodological variations across science.

10.1.3. Status of the academic literature

Another recurring theme comprises a trend in the view on, and position of, the scientific publication system in the research landscape. Arguably, interpretations of the academic literature are shifting from a logbook or dialogue model into a database of verified facts. As we outlined in chapter 4, there seems to be a growing expectation that published research articles contain universal, standardised facts. This is among others reflected in the number of retractions in science, as well as the emergence of novel reasons to retract articles, which we outlined in chapter 5. Furthermore, new research methods such as meta analyses further enhance this perception of academic research output. With novel innovations in publishing technologies, such as the desire to establish systems of automated peer review described in chapters 1 and 4, as well as the arrival of computer written academic content, the end of this trajectory of change does not seem to be in sight yet.

The changing conception of the academic literature's status is potentially problematic for several reasons, not the least because the output of scientific endeavours has never been certainty. To go with Harry Collins' analyses of replication in science: "Variation of expert opinion is to be treated as natural and ordinary rather than as an area of eradicable weakness or bias. So far as certainty is concerned, it is an impression given by social and temporal distance from the seat of knowledge creation" (Collins, 1992).

10.2. Limitations

As with most research endeavours, the studies in this dissertation suffer from several limitations. Most chapters include a section on limitations of that particular study, but this naturally leaves some limitations relating to the entirety of this dissertation unaddressed. I will explicate some of them in this section.

The first limitations of our studies comprise a limitation of scope. In this dissertation I have now mainly focused on the 'back-end' of the research process by studying the publication system. This has resulted in several new insights in the way in which 'bad science' may be filtered from the literature. Even though some of the findings relate to proactive measures against misconduct such as taking away incentives to obtain and publish high impact, spectacular results, most findings relate to detecting and subsequently filtering records that have already been prone to problematic research practices. Arguably, such behaviour would preferably be prevented in an earlier stage of the research process. This might include interventions at the stage of proposal writing and grant application, the stage of collecting data by performing experiments or doing field work, of analysing or writing-up results, or in occasions of supervision and collaboration that carry through the entire process. Even though some such interventions have already been implemented, such as registered reports or the obligation to have multiple supervisors (Center for Open Science, 2018; College van Bestuur EUR, 2015), only limited research has been done to study the effects and consequences of such interventions. Especially the first stages of the research process represent an understudied topic, leaving many questions regarding responsible practices in writing research proposals and grant applications, as well as the effect of different funding instruments on responsible research practices in the remainder of the research process.

A second limitation to the scope of our project relates to our study of the peer review system, in which we have predominantly focussed on the editorial perspective, rather than the perspective of an individual reviewer. This has led to a study in which we could analyse large amounts of journals and different review procedures, yielding recommendations on how to structure the system at the level of a journal or publisher. However, we acknowledge that highly relevant variations occur at the level of the individual reviewer, as has been shown for instance by Michele Lamont in her study on review boards of grant applications (2009). Future research should focus on this reviewer perspective as well, aiming to bridge the gap between formal review procedures and actual review practices. This includes questions such as how blinding or disclosing author or reviewer identities affects the way in which review is performed or manuscripts are judged, as well as the way in which digital tools assisting in review are operated by reviewers or editorial staff.

Third, this dissertation’s focus on research and publication practices overlooks research funding and rewards, among the decisive factors that structure researchers’ behaviour. Relating issues regarding misconduct to novel research evaluation practices could be a particularly fruitful avenue for further investigations. Whether quantitative performance indicators have directly led to misconduct or questionable research practice, is still difficult to assess. However, at the very minimum they have resulted in different publication strategies (Bloch & Schneider, 2016; Butler, 2003), with a strong focus on quantity and impact indicators, as well as resulted in more fundamental shifts, including a shift in alleged objectives (i.e. ‘goal displacement’ using Mertonian terms) and task reduction (Müller & de Rijcke, 2017). Besides leading to potential forms of ‘gaming’ the scientific reward system, this focus on performance metrics may potentially have more epistemically intrusive effects by impacting on the choice of research topics, the design of research projects and social organisations in which they are performed, as well as researchers willingness to share research results, even among collaborators (Müller & de Rijcke, 2017; Sigl, 2016). These truly important considerations have been relatively untouched in this dissertation, but arguably make up key pillars for efforts to foster research integrity.

Fourth, research integrity has proven itself to be a fast-moving target, which has created some limitations in studying it. Interest in research integrity has mounted over the past years, now leading to an almost innumerable set of initiatives to study and foster the responsible conduct of research (Horbach, Nuijten, Tijdink, & O'Neil, 2019). In addition, wider concerns over scientific publishing and the role of academic publishers have further sparked the debate and led to various initiatives that have the potential to drastically alter the publishing landscape. Occasionally, new initiatives were introduced while planning, performing or analysing our field work. This may have caused us to not include important initiatives in our study, which might turn out to be crucial to integrity discussions in the future. In other occasions, initiatives may have been so recent that they did not yet allow for proper analysis, such as in the case of registered reports and open peer review as novel editorial procedures. Hence, repeating some of the studies in this dissertation might allow for more optimal inclusion of these very recent initiatives.

In particular, various initiatives regarding Open Science have only recently emerged. More and more, transparency and openness are considered as important mechanisms to foster integrity and allow for higher scrutiny, thereby deterring researchers from misconduct. This includes the publication of open access research articles (cOAlition S, 2018; Nguyen, 2012), increased sharing of data and research materials (Hardwicke & Ioannidis, 2018) and opening up the peer review process (C. J. Lee & Moher, 2017; Mirowski, 2018). In addition, norms regarding research integrity have been in flux, with new guidelines and codes of conduct being published while carrying out our research (e.g. ALLEA, 2017; KNAW et al., 2018). For example, these include more explicit norms regarding the reuse of textual material in later publications by the same authors.

The observation that the research integrity landscape is changing so rapidly has caused some difficulties in performing our studies, but above all are obviously laudable as they aim and help to improve trustworthy science. Hence, it is a trend that we are happily applauding. Ultimately, this dissertation even hopes to contribute to it and help facilitate a research culture in which integrity is core pillar of good research.

10.3. Recommendations

This dissertation studied research integrity from various perspectives. In doing so, it identified several integrity challenges or ways in which they might (not) be properly dealt with. First, it showed that integrity challenges emerge differently in diverse academic contexts. Research discipline, epistemic culture, geographical location and organisational research climate all seem to influence the way in which integrity challenges are perceived and how they can be effectively addressed. These differences are most likely to be further mediated by differences in funding schemes and perceived work pressure. Second, while several stakeholders in the academic arena claim to highly value a strong research integrity climate, other, competing considerations might equally well influence their policy and strategy. These include efforts to obtain business value in commercial publishers, as well as reputational arguments in research performing organisations and even individual researchers. Third, hierarchies and power structures may complicate effective handling of integrity issues, as well as hinder transparency and effective learning of such issues.

These findings have implications for researchers, science policy makers, research performing organisations, science funders and academic publishers.

First and foremost, these findings indicate the urgent need to work towards a system of shared and distributed responsibility for research integrity. This entails more than simply claiming that ‘we should do this all together’, but actually demands that we make clear how different actors may contribute to integrity in science. It also demands that all actors are aware of their own and each other’s responsibilities. To start, research performing organisations may be expected to contribute to a culture of integrity in multiple ways. Partly based on the results of this dissertation, we listed these expectations in a joint statement with research integrity experts across Europe. In this statement we outline how research performing organisations may contribute through: Providing information about research integrity, providing education, training and mentoring; strengthening a research integrity culture; facilitating open dialogue; wise incentive management; implementing quality assurance procedures, improving the work environment and work satisfaction; increasing transparency of misconduct cases; opening up research; implementing safe and effective whistle-blowing channels; protecting alleged perpetrators; establishing a research integrity committee and appointing an ombudsperson; and by making explicit the applicable standards for research integrity (Forsberg et al., 2018).

Second, academic journals and publishers could foster research integrity through facilitating effective models of peer review and by adequately addressing problematic research that slipped through the review system. This includes among others warning readers through post-publication comments for caveats in a study’s design or potentially misleading conclusions, such as we indicated in research on misidentified cells. In addition, journals should acknowledge the role they have in dealing with misconduct cases. During my visits to editorial offices, employees repeatedly indicated that they were not interested in punishing misbehaving researchers, but merely intended to “make sure that the literature contains valid knowledge”. However, our studies in part III – *To Clean* demonstrate that both individual researchers as well as their university managers consider retracted journal articles one of the most severe punishments. Hence, journals and publishers should acknowledge their role in this and implement fair and effective procedures to retract or correct published articles. This also requires transparency about why articles are or will (not) be retracted.

Explicating these responsibilities in how to deal with integrity issues and understanding different actors' concerns and stakes may prevent finger pointing between stakeholders, such as occurred for example between journal editors and research performing organisations (e.g. Bauchner, Fontanarosa, Flanagin, & Thornton, 2018). This is not to say that actors should not publicly point out each other's responsibilities, but rather aims to anticipate these discussions.

Third, while this dissertation aimed to divert the common focus in research integrity debates away from individual researchers, I of course acknowledge that they have a crucial role in the academic arena. The research climate, organisational culture or institutional procedures can be as favourable as they get, ultimately it are the scientists themselves who should responsibly conduct research. Moreover, they have a crucial role in shaping initiatives at various levels: journals are very sensitive to authors' and reviewers' desires, research organisations commonly involve researchers in setting their policy, and a research culture is obviously largely determined by those carrying out this research. Moreover, researchers have a responsibility in mentoring junior colleagues towards becoming virtuous scientists, and they should correct or report colleagues when witnessing problematic research behaviour.

Finally, funders have a prominent role in crediting and rewarding researchers and research performing organisations. In this role, they have a responsibility to provide fair and wise incentives for researchers to conduct responsible research. In addition, funders have the authority to set conditions to the research they fund. They may hence use this authority to demand responsible research practices such as the appropriate use of research materials (e.g. testing for misidentified cell lines), transparency in research practices, and openness of research data. As I mentioned earlier, these conditions should always take into account the specific context in which research takes place, with appropriate standards for diverse research settings. Last, research funders are in a prime position to support studies that increase our knowledge about responsible research practices.

This urge for a more distributed and explicated system of responsibilities to strengthen research integrity is not an isolated call. Recent work by Berggren and Karabag (2019) contains similar elements, introduced by the notion of *institutional fields.* They use this notion to describe the different but related perspectives, values and stakes of actors in the integrity debate and show how different institutional fields are at play in covering, uncovering, legitimising and immoralising research misconduct.

Hitherto, the academic debate has been too preoccupied with the academic perspective, neglecting other relevant institutional logics, including those of research performing organisations, academic journals and publishers, funding organisations, and the media. In this dissertation we demonstrate that those perspectives are similarly important and that they do not necessarily align with the academic perspective. The existence of these diverse sets of motives and values has been identified before, when Shapin (2008, p. 87) noted that in the 1980's it became commonplace to assert that "scientists are driven by the same range of motives as businessmen, politicians, and churchmen". He used this as a major explanation of several cases of misconduct, when he asked: "Why do you think that the devotees of Newton's Laws will be more saintly than those ruled by Cardinal Law?" (Richard Lewontin, quoted by Shapin).

We also note that these institutional fields are not only forced upon academia by external parties, but are equally maintained from within, through what Roumbanis (2018), after Bourdieu, calls *symbolic violence*. In an ethnography of Swedish universities, he shows how junior researchers are educated into a system of career uncertainty that stresses the need for personal achievements and a strong reputation. He shows how senior researchers exercise subtle forms of power upon junior scholars when passing on the values, norms and criteria of the academic elite and the funding agencies. This further highlights the need for shared action towards fostering research integrity. If all actors reinforce and influence each other, then only combined efforts of all actors will lead to effective change.

10.4. Open questions

The discussions identified and raised in this dissertation might have direct impact on policy, funding, education or individuals, but, as is common with scientific research, it predominantly leads to many more new questions.

While this study identified several differences between disciplinary fields, geographical locations and organisational cultures, more systematic research is needed to identify how integrity challenges emerge differently in diverse research contexts. This includes not only a systematic study of diverse norms and formal guidelines in these contexts, but also a detailed study of research and publishing practices. Among others, these studies could continue on our analysis of power structures and hierarchical relations in research. Some evidence exists, showing that these phenomena apply differently across research disciplines and cultures, but more research is needed to clarify this and to understand how such power structures affect research integrity. This may shed further light on the diversity in integrity challenges and should ultimately contribute to more effective strategies to foster integrity across the entire spectrum of research contexts.

A similar reasoning applies to the diversity in institutional logics identified in the previous subsection. In this dissertation I demonstrated that multiple motives and value frameworks are at play and that the discussion about research integrity has hitherto too much focussed on the academic discourse. Despite these findings, questions remain about how the different motives and values are co-produced, how they influence each other and how they may be effectively aligned to strengthen an integrity culture.

The results in chapters 4 and 5 in this dissertation suggest additional directions for future research regarding editorial and peer review practices. While showing differences in diverse review procedures' effectiveness in preventing retractions, more research is needed to clarify underlying mechanisms. Future research should elucidate the causal connections behind the identified strong correlations. What specific mechanism, for instance, makes double-blind review better capable of detecting errors in research records than single-blind review? And what causes augmented interaction between authors and reviewers to be related to more retracted journal articles? A closer look at the practice of peer review, and particularly in those cases that led to retractions, could clarify this.

Last, this dissertation's conclusions raise future questions about the applicability of deviance theories to the issues of integrity and misconduct in science. While initiating this connection, mainly in part III – *To Clean*, there is abundant room for further progress in this direction. Some laudable efforts have been made in the work of Faria (2015) and Van Buggenhout and Christiaens (2016), but questions remain about the potential for learning about prevention that these approaches provide. While I do not aim to frame text recycling or reporting on misidentified cell lines as criminal behaviour, we do spot an interesting resemblance between responses to our study and theories of deviance (Agnew, 1994). This may provide fruitful avenues for future research about contexts in which researchers legitimise questionable research practices and ultimately assist in maintaining such practices.

In line with this, it is noteworthy that little dialogue has occurred between scholars of integrity and misconduct in science and those of similar issues in other domains, such as corporate fraud, integrity in public administration, or even doping in sports. While research on scientific integrity is relatively recent (Horbach & Halffman, 2017), studies of comparable issues in other domains have a much longer history. Even though some research in this direction has been done (Sacco, Bruton, & Brown, 2018), much remains to be elucidated. Hence, I argue that much could be learned from a constructive dialogue between these, yet separated, research fields.

10.5. References

Agnew, R. (1994). The techniques of neutralization and violence. *Criminology, 32*(4), 555-580.
ALLEA. (2017). The European Code of Conduct for Research Intregrity - Revised Edition. In A.-A. E. Academies (Ed.). Berlin, Germany.
Bauchner, H., Fontanarosa, P. B., Flanagin, A., & Thornton, J. (2018). Scientific misconduct and medical journals. *Jama*. doi: 10.1001/jama.2018.14350
Berggren, C., & Karabag, S. F. (2019). Scientific misconduct at an elite medical institute: The role of competing institutional logics and fragmented control. *Research Policy, 48*(2), 428-443. doi: 10.1016/j.respol.2018.03.020
Bloch, C., & Schneider, J. W. (2016). Performance-based funding models and researcher behavior: An analysis of the influence of the Norwegian Publication Indicator at the individual level. *Research Evaluation, 25*(4), 371-382. doi: 10.1093/reseval/rvv047
Butler, L. (2003). Modifying publication practices in response to funding formulas. *Research Evaluation, 12*(1), 39-46. doi: 10.3152/147154403781776780
Center for Open Science. (2018). Registered Reports: Peer review before results are known to align scientific values and practices - Participating Journals.   Retrieved October 4, 2018, from https://cos.io/rr/
cOAlition S. (2018). Accelerating the transition to full and immediate Open Access to scientific publications. from https://www.coalition-s.org/about/
College van Bestuur EUR. (2015). Definitief besluit College van Bestuur naar aanleiding van melding plagiaat. In EUR (Ed.). Rotterdam.
Collins, H. (1992). *Changing order: Replication and induction in scientific practice*: University of Chicago Press.
Etzkowitz, H. (1998). The norms of entrepreneurial science: cognitive effects of the new university–industry linkages. *Research Policy, 27*(8), 823-833. doi: https://doi.org/10.1016/S0048-7333(98)00093-6
Faria, R. (2015). Scientific misconduct: how organizational culture plays its part. *Tijdschrift over Cultuur & Criminaliteit, 5*(1), 38-54.
Forsberg, E.-M., Anthun, F. O., Bailey, S., Birchley, G., Bout, H., Casonato, C., . . . Zöller, M. (2018). Working with Research Integrity—Guidance for Research Performing Organisations: The Bonn PRINTEGER Statement. *Science and Engineering Ethics, 24*(4), 1023-1034. doi: 10.1007/s11948-018-0034-4
Grey, A., Bolland, M., Gamble, G., & Avenell, A. (2019). Quality of reports of investigations of research integrity by academic institutions. *Research Integrity and Peer Review, 4*(1), 3. doi: 10.1186/s41073-019-0062-x
Hardwicke, T. E., & Ioannidis, J. P. A. (2018). Populating the Data Ark: An attempt to retrieve, preserve, and liberate data from the most highly-cited psychology and psychiatry articles. *Plos One, 13*(8), e0201856. doi: 10.1371/journal.pone.0201856
Horbach, S. P. J. M., & Halffman, W. (2017). Promoting Virtue or Punishing Fraud: Mapping Contrasts in the Language of 'Scientific Integrity'. *Science and Engineering Ethics, 23*(6), 1461-1485. doi: 10.1007/s11948-016-9858-y

Horbach, S. P. J. M., Nuijten, M., Tijdink, J., & O'Neil, G. (2019). Position Paper: Fostering Responsible Research Practices - Start document for ZonMw's BVO program. The Hague: ZonMw.
KNAW, NFU, NWO, TO2-federatie, Verneniging Hogescholen, & VSNU. (2018). Nederlandse gedragscode wetenschappelijke integriteit.
Lamont, M. (2009). *How professors think: inside the curious world of academic judgment*. Cambridge, Massachusetts: Harvard University Press.
Lee, C. J., & Moher, D. (2017). Promote scientific integrity via journal peer review data. *Science, 357*(6348), 256-257. doi: 10.1126/science.aan4141
Marcus, A. (2019). Scientific integrity "a rare commodity these days:" Cancer journal makes its first retraction. Retrieved from https://retractionwatch.com/2019/03/29/scientific-integrity-a-rare-commodity-these-days-cancer-journal-makes-its-first-retraction/
Mirowski, P. (2018). The future(s) of open science. *Social Studies of Science, 48*(2), 171-203. doi: 10.1177/0306312718772086
Müller, R., & de Rijcke, S. (2017). Exploring the epistemic impacts of academic performance indicators in the life sciences. *Research Evaluation, 26*(3), 157-168. doi: 10.1093/reseval/rvx023
Nguyen, N. T. (2012). Can open access journals help uphold the value of research integrity? *Jornal of Business and Financial Affairs, 1*(1).
Research Square. (2017). Rubriq. from https://www.rubriq.com/
Roumbanis, L. (2018). Symbolic Violence in Academic Life: A Study on How Junior Scholars are Educated in the Art of Getting Funded. *Minerva*. doi: 10.1007/s11024-018-9364-2
Sacco, D. F., Bruton, S. V., & Brown, M. (2018). In Defense of the Questionable: Defining the Basis of Research Scientists' Engagement in Questionable Research Practices. *Journal of Empirical Research on Human Research Ethics, 13*(1), 101-110. doi: 10.1177/1556264617743834
Shapin, S. (2008). *The Scientific life: a moral history of a late modern vocation*. Chicago: The University of Chicago Press.
Shimshoni, D. (1970). The Mobile Scientist in the American Instrument Industry. *Minerva, 8*(1), 59-89.
Sigl, L. (2016). On the Tacit Governance of Research by Uncertainty:How Early Stage Researchers Contribute to the Governance of Life Science Research. *Science, Technology, & Human Values, 41*(3), 347-374. doi: 10.1177/0162243915599069
Van Buggenhout, M., & Christiaens, J. (2016). Deviance in science: a criminological analysis - PRINTEGER deliverable D II.5. from http://printeger.eu/wp-content/uploads/2016/12/D2.5.pdf

Epilogue – Reflections on researching a sensitive topic

Doing research, writing up scientific articles, finishing a doctoral thesis, all come with considerable struggles and have henceforth pushed many to their limits. However, doing so in the end is, hopefully, a rewarding process all worth the efforts and sacrifices that it took. Seeing your research making an impact, contributing to a wider knowledge base and actually being part of a process of change, is what most researchers ultimately aim for. In my case, performing research on the sensitive topic of research integrity and misconduct has magnified all these experiences: It both led to more difficulties, demanding considerable additional care and caution, but also to an overwhelming amount of responses, both very positive and critical ones. In this chapter, I will briefly reflect on the process of researching this highly interesting and timely, but, as we learned on the way, also very sensitive topic. I will discuss some of the issues that came up, ways in which we dealt with this and give an overview of the impact that our research had, the responses that it invoked, and what these responses may teach us.

Sensitive research topics have been theorised and have been the subject of empirical research for several decades now. R. M. Lee and Renzetti (1990) describe sensitive topics as those which: "(a) intrude into the private sphere or delve into some deeply personal experience, (b) concern with deviance and social control; (c) impinge on the vested interests of powerful persons or the exercise of coercion or domination; [or] (d) deal with things sacred to those being studied which they do not wish profaned." During our studies we ended up in situations reflecting multiple of these characteristics. Evidently, we dealt with issues of deviance and social control, but on top of that we also intruded into the private sphere or personal experiences of some actors, including for instance the interviewees in our study on cases of alleged misconduct in chapter 9. In fact, we might also have stepped on the toes of some (powerful) people seeing something 'sacred' to them threatened, such as their reputation or career. Hence, it seems reasonable to conclude that we have been dealing with a sensitive topic. Sieber and Stanley (1988), furthermore, note that sensitive research addresses some of society's most vital concerns. Ignoring the ethical dilemmas in such research would be irresponsible; however, refraining from studying them, just because they are controversial, would be negligent as well.

E.1. Challenges in researching sensitive topics

Not surprisingly, performing research on sensitive and controversial topics requires additional care, as such endeavours might meet with "methodological, technical, ethical, political, and legal problems" as well as have potential effects on the personal life of the researcher (R. M. Lee & Renzetti, 1990). We have indeed met with several of these challenges at various stages of the research process.

First and foremost, and obviously not surprising, studying research integrity suffers from methodological issues in obtaining data. As with other forms of deviant or questionable behaviour, obtaining data requires some creativity and care, because data is not readily available, those involved in it are not willing to share data, and even witnesses of deviance might be reluctant to share their knowledge. Criminologists refer to this as the ‘dark number issue’ (Biderman & Reiss, 1967). This hence excludes several forms of data collection and forces researchers to resort to other, more indirect, approaches (Thomas, 1993).

We indeed noticed that carrying out a project on research integrity may cause difficulties in data sampling. This became apparent in several parts of our study, including our survey among journal editors: “I would like to inform you that we prefer not to participate in your survey, because of the nature of your study object”, was one of the responses we got from editors. It was also reflected in our attempts to visit editorial offices of academic journals and publishers. In the latter case, strict negotiations had to be made to grant access to these research sites. However, challenges in obtaining access to data were most pressing in our attempts to study specific cases of alleged misconduct. Interview requests were commonly ignored, denied or only provisionally accepted under strict conditions, even though from the start it was made clear that participants would be fully anonymised and would only appear as ‘interviewee 1, 2, 3 …’ in any project output. The conditions interviewees requested included interviews not to be recorded (out of fear of someone using the freedom of information-act to request access to the recordings and subsequently recognising the interviewee’s voice) and interviews to be held in public places rather than the interviewee’s office (out of fear of colleagues seeing me enter the office and asking for explanations). Also, one interview was denied because the envisaged interviewee was declared unfit for work after suffering heavily from (the stress caused by) one of the cases we aimed to investigate.

These challenges are not uncommon in ethnographic studies, where one always has to deal with issues of getting access and gaining trust from interviewees, participants or research subjects to obtain reliable data. However, they may be particularly relevant when questioning people about situations which they experienced as markedly traumatic or for which they fear retaliation. Following Sieber (1993), this required us to gain insight into the way others perceive the risk of participating in our research and design the research accordingly, minimalizing potential sensitivities and risks. We chose to use what Fielding has called an ‘intercalary role’ in performing our fieldwork (Fielding, 1993). This entails a method holding a balance between the appreciative stance (Matza & Blomberg, 2010) and investigative research or conflict methodologies (Douglas, 1976). It places the researcher in a position between passive recipient and sceptical investigator. This allows a researcher to both build a relation of trust with the research participant, as well as be critical of the presented material. In our case, the latter involved, among others, the continual verification, as far as possible, of whether participants were engaging in acts of libel or defamation.

A second set of challenges emerged during the stage of writing and reporting the findings of our studies. We were very much aware of some of the potential implications of our research and findings, which repeatedly required great care in formulating them: What sticks are we presenting to whom? And who could be hit with them? This particularly applies to the findings of the studies in part I – *To Spill*. Following our conviction that the overall discourse on scientific integrity focuses too much on the individual scientist or perpetrator, we systematically chose not to follow-up on this rhetoric. Rather, we chose to report our findings only at the aggregate level on which we were able to study systemic and organisational factors contributing to the publication practices on which we reported. This not only applies to reporting our findings in research articles or this dissertation, but also (or even mainly) to the news reports and interviews that succeeded the academic output. I will discuss the challenges related to the responses and media attention for our study in more depth later in this chapter.

Other social scientists studying sensitive topics, including drug abuse and the criminal justice system, have discussed the issues with reporting their results, both in academic outlets and news channels (Adler & Adler, 1993; Channels, 1993). Adler and Adler (1993) discuss the act of *self-censorship* in reporting results, which might occur for several reasons: researchers might develop loyalties to their research participants, hindering them from fully disclosing their results; certain institutional or political factors might hinder full disclosure, as well as fear for losing sponsorship or funding. Last, an attempt to avoid contagion of stigma might lead researchers to be restricted in publishing their work (Adler & Adler, 1993). Such incomplete reporting of data and results, the Adlers argue, may be perfectly understandable but also creates ethical challenges that researchers should not be blind to. Others add to this that issues may be particularly pressing when media attention might result from disseminating research findings. Especially the media's tendency to sensationalise or trivialise results may pose difficulties (Channels, 1993; Renzetti & Lee, 1993). We encountered this ourselves when dealing with journalist's questions about 'how many patients died as a result of cell line contamination?' and 'what to do with economists who stop writing new stuff, but keep on recycling their old work?'. As a response, we continually tried to nuance our findings, refrained from individualising reporting, and aimed to keep a focus on organisational and systemic factors contributing to the reported findings. I want to thank all journalists who have supported us in disseminating our findings in such ways.

Last, the nature of the topics we aimed to investigate, as well as the overwhelming responses to the publication of our findings in chapters 2 and 3, initiated some challenges regarding the choice of future research topics and priorities. Both studies at the base of part I – *To Spill* were quite successful in drawing media attention as well as several critical or even offensive responses. The positive vibe of the media attention may spark willingness to follow up on these studies, at the risk of getting bogged in a witch hunt with little academic merit but high personal visibility and satisfaction. Although I do not want to point at individual behaviour, it is exactly this pattern of aiming for increased attention and fame that let some prominent researchers into their eventual derailment (Abma, 2013; Stapel, 2012). Conversely, the critical and hostile responses, which admittedly let to some nights of reduced sleep, might lead to sentiments of shying away from such controversial topics and potentially highly visible outputs. In the end we decided to continue our research on these topics, with studies on the effect of journal guidelines on the reporting of misidentified cell lines (Hepkema, 2019) and more in-depth and wide scale analyses of text recycling in academic articles.

E.2. Responses to results on sensitive topics

Besides leading to various challenges in carrying out our research, the nature of our study and the resulting findings evoked a range of responses. Over the past years, using Altmetrics has become a standard method to measure the impact of scientific research by counting mentions on social media, news platforms, blog posts and other digital outlets. Even though it is debatable whether these metrics present an adequate indication of a studies impact or merit (Thelwall, Haustein, Lariviere, & Sugimoto, 2013), there seems to be little controversy about the fact that Altmetrics give a fair attention-score. Judging from these scores, we conclude that our studies indeed received a fair amount of attention, which might teach us interesting lessons about how research on this sensitive topic is interpreted and judged.

On top of these mentions and responses in public outlets, we received quite some feedback through personal communication (emails, letters, etc.). The responses typically varied in tone, scope and meaning, but they may be meaningfully grouped in several clusters. Far from being a rigid categorisation of the responses, this clustering might give some indication of what our study evoked among various stakeholders.

The first set of responses can be broadly summarised as: ‘Good that you study this, but I know of another case which is far worse and more important’. Repeatedly, we received comments, usually through personal communication, from people who felt that our work was important in helping to uncover problematic behaviour in science, but who pointed us to other cases that, in their opinion, merited even more scrutiny and (public) attention. Although, I do not want to disclose any of these comments or potential cases, it suffices to say that critically and systematically examining all the suggestions could easily fill another few chapters of a doctoral dissertation. These responses mainly demonstrate that, apparently, a substantial number of people have witnessed a case of dubious academic behaviour while feeling that these cases are not, or not sufficiently, acted upon. The fact that these commentators turn to us, i.e. relative strangers, just because we happen to have studied other forms of academic (mis)behaviour, suggests these people felt their concerns had not been taken sufficiently seriously. Also, the wording of some of the responses in this category reveals a substantial level of frustration with this perceived lack of support:

“So you obtained a grant and you want information on responsible research. I'd like to suggest you have a look at …. […] I contacted X, director of A, he did virtually nothing. […] which is an absolute shame. […] Professor Y says X is a crackpot, that is how bad it all is.”

A second set of responses consists of ‘calls-for-help’. Somewhat similar to the previous set of responses, this set consists of people commenting on our work by indicating that they have witnessed, or been part of a case of potential scientific misbehaviour. But rather than insisting that we should investigate their cases, these responses are characterised by an explicit call for support or advice. For instance, I was contacted by the editor-in-chief of an academic journal, who recently received an allegation of plagiarism in one of his journal’s articles. He/she was unsure about how to deal with this situation. The editor wrote me that he/she was “unfamiliar with this situation and not sure about which procedures to follow”, hence “I would like to ask your advice”. Again, these responses highlight the sensitivity of the topic that we are dealing with. It also points to the relatively low level of standardisation and common knowledge on this subject. While several national and international committees and advisory boards have been established - including COPE, ORI, and LOWI – they apparently have not yet become established platforms to go to when soliciting advice on matters relating integrity. In the words of the editor-in-chief: “the beaten path seems to be far less beaten than you might think”.

A third and fourth category of responses comprises critical comments to our work. I distinguish two groups of such comments because they arguably originate from a different issue or aim to achieve different outcomes. The first set of comments mainly emerged after publication of our study on text recycling and consisted of critical assessments of our study’s methodology. Perhaps not surprisingly, the comments were mainly brought forward by (Dutch) economists, some privately, others more publicly, for instance in the form of a letter to the editor of *Research Policy* (Lukkezen, 2019). These comments mainly centred on choices made with respect to the inclusion of certain individuals in our sample and our sampling strategy in general. Commentators both asked questions about them and frankly criticised them. Also, the threshold for labelling a paper to contain ‘problematic amounts of text recycling’ was called into question. Because we replied to these comments elsewhere (Horbach & Halffman, 2019), I will focus on the implications of these comments and the lessons to be learned from them, rather than go into the details of our rejoinder.

Arguably, some of the comments indicate a fear of being included in our sample and of being ‘caught’ engaging in dubious behaviour. One of the comments, sent to us privately, explicitly mentioned this: “I think I may be part of your sample”. While some framed their responses as a question and opportunity to reflect on their own practices: “how may I/we learn from this in the future?”, others critically asserted that our methodology misrepresented the ‘actual world’ by bringing their discipline in disrepute. In general, the responses underscore the sensitivity and controversy of text recycling and research misconduct, showcasing one of Lee and Renzetti’s definitions of sensitive topics as “deal[ing] with things sacred to those being studied which they do not wish profaned” (1990).

More or less the same applies to the fourth category of responses: the critics asserting that what we label as 'problematic' is not that problematic at all. This was brought forward after our text-recycling study, but was mainly voiced in response to our contaminated-cell-line study in chapter 3. As we mentioned in chapter 2, text recycling is a contentious practice. In some occasions, there may be very legitimate reasons to re-use parts of one's own work. Hence, some commentators stressed that our study misleadingly addressed text recycling as problematic behaviour. The same occurred on a larger scale when commentators, almost exclusively biomedical researchers, questioned our framing of studies contaminating the scientific literature by using misidentified cell lines. Researchers working with cell lines repeatedly contended that the use of misidentified cells did not (necessarily) harm the scientific literature. In fact, some publicly claimed to know about a misidentification, but nevertheless continued their study because the precise characteristics of the cell line were deemed irrelevant (Keulemans, 2017). In addition, researchers claimed that the problem of cell line contamination was one that "all researchers are familiar with" and "we have learned to deal with", even though our results suggested quite the opposite. These responses clearly resemble what criminologists have described as one of the ***techniques of** *neutralisation (Sykes & Matza, 1957)*. Denial of injury, i.e. stating that no (real) harm has been done, is one of the ways in which people legitimise actions which they might otherwise be concerned of. A more radical form of this technique was also voiced: "This is indeed a problem, but other problems are far more severe, so we should not bother too much about it".

Another such neutralisation technique can also be found in the responses to our cell line article: the denial of responsibility (Sykes & Matza, 1957). Several (biomedical) researchers stated that indeed they were aware of potential issues with their cell lines, and indeed they knew that they should do something about it (e.g. by better testing and validating their research materials). In fact, they even wanted to do so, if not for the fact that work pressure was so high and competition for (fast) results so fierce. "As long as others do not do it, and as long as we do not need it to get our work published, additional testing puts myself in a position of competitive disadvantage", one of the researchers claimed in an email. Hence, the researcher him-/herself would arguably not be to blame because he/she is not in the position to make proper choices. These choices are instead forced upon him/her by the system. Without making normative judgments about the validity of this claim, it presents anecdotal support for this dissertation's focus on systemic and organisational causes of problematic research practices.

A last response I would like to mention does not neatly fit into one of the categories mentioned above. It comprises a single message, that initially slightly shocked me, later made me laugh a bit, and ultimately makes me understand that issues concerning integrity might arise in all sorts of unintended ways. In the past years, there has been growing concern about fake peer reviews, where authors manage to get appointed as reviewer of their own article by intentionally misleading journal editors. Several attempts have now been made to prevent and detect such undesirable attempts (e.g. Gao & Zhou, 2017; Stern, 2017). I was therefore thoroughly surprised when I opened my email on the train returning from a conference, and saw an invitation to be a reviewer of one of the articles in this dissertation, i.e. an article of which I am the first and corresponding author. After declining the invitation and alerting the (associate) editor about this mistake, he/she responded by apologising and mentioning that: "The irony of having an author review his own paper with a topic of integrity and questionable practices has not escaped me!"

Arguably, this is just a minor anecdote about my research, merely involving simple administrative mistakes that one will come across in every professional setting. Still, this case again made me realise that research and publishing are human endeavours in which mistakes are made and challenges may arise totally unexpected. Indeed, even simple administrative mistakes could ultimately lead to situations of questionable research practices. This also reminded me about the vulnerability of our systems and the trust required in the well-functioning of all actors in the research process.

E.3. Valorisation

Clearly, our research evoked a substantial set of responses, comments and public attention. While this arguably indicates that we touched upon topics considered relevant by researchers, policy makers and the wider society alike, it tells us little about the actual impact that our research may have had.

Tackling the issue of research misconduct and problematic research entering the academic literature may be classified as a wicked problem, i.e. one that is complex, requires multiple actors to change their practices, and is characterised by a plurality of values or knowledge bases. When working on such problems and actively engaging in solving them, one may expect or aim for various forms of impact. De Rijcke and Wilsdon (Wilsdon, 2018) in their plea for the responsible use of quantitative indicators distinguish between the following: (i) Create advocacy coalitions, (ii) lead to cosmetic adjustments, (iii) corroborate the machine, (iv) act as tin opener (to wider issues and concerns), or (v) create institutional culture change. I would like to add to this list at least two items: (vi) creating awareness, and (vii) provoke discussion and debate.

Judging from the first and second set of responses to our research, at least the fourth item on the above list was attained by our initiatives. To some extent, our work has fuelled discussion on wider topics related to research integrity and proved to be fertile ground for other actors to voice their concerns. Specifically, during our research on peer review procedures in academic journals, we concluded that finding out about these procedures was notoriously difficult. For the majority of journals, we could not even figure out some pretty obvious features, such as whether a journal performs single- of double-blind review. This may be because editors consider their procedures obvious, or there may be a particular procedure that is obvious in a particular field. Still, it seems fair that prospective authors are informed about what will happen to their paper, that reviewers know what is expected of them and that researchers get access to study the advantages and disadvantages of specific editorial procedures. After discussion at a workshop with stakeholders in scientific publishing, these observations led to the publication of the *Leiden Declaration on Transparent Editorial Policies (Horbach, Halffman, De Rijcke, Reyes Elizondo, & Van Leeuwen, 2018)*. Hence, our work acted as a tin-opener to wider concerns of transparency in academic publishing.

This latter effort also constitutes an example of the first item on the above list: creating advocacy coalitions. In addition to the publication of the Leiden Declaration, we aimed to create such coalitions in other ways as well. In another effort to combine various stakeholders' perspectives, we published a call for research organisations to take their responsibility in creating an integrity culture. In the *PRINTEGER-Bonn Statement* we not only ask such organisations to foster research integrity, but also explain what steps this might involve (Forsberg et al., 2018).

On top of the forms of impact identified by de Rijcke and Wilsdon, I think that our work has had an impact in at least two additional ways. First, it may have simply added to awareness of certain issues among certain stakeholders. The media attention for our text recycling and contaminated cell line articles constitute a prominent example of this. But also on a more mundane level, editors and publishers may have benefitted from our work on novel peer review procedures. For instance, some of the responses to our survey among journal editors indicate a heightened level of awareness due to our work: "I did not know tools for checking references exist. We should be doing that as well!" Similarly, the arguments in our studies have evoked considerable discussion and debate, for instance on the permissibility of text recycling or the responsibility of research performing organisations to foster a research integrity culture (Horbach, 2018). In addition, some editors indicated that the results of our studies immediately led them to discuss novel ways of organising their journal's editorial process. In the current research integrity landscape, such discussions may be a decisively important step in creating a culture in which it is normal to talk about research integrity challenges, to openly discuss norms and collectively set implicit and explicit standards. I therefore hope this dissertation has contributed to such debates, ultimately leading to more trustworthy research.

Revisiting the conversation I had at the beginning of chapter 1, this dissertation has at least contributed to our understanding of some of the basic questions arising in the aftermath of misconduct cases. While it is difficult to assess whether our studies' conclusions have already influenced specific research cultures, I hope they will do so in the future. For now, at the very minimum, they have at least provided me with ways of responding in such everyday conversations. And who knows, perhaps having such conversations might be of decisive importance in transforming research cultures.

E.4. References

Abma, R. (2013). *De publicatiefabriek: Over de betekenis van de affaire Stapel*. Nijmegen: Uitgeverij Vantilt.

Adler, P. A., & Adler, P. (1993). Ethical Issues in Self-Censorship: Ethnographic Research on Sensitive Topics. In C. M. Renzetti & R. M. Lee (Eds.), *Researching Sensitive Topics* (pp. 249-266). Newbury Park, California: Sage.

Biderman, A. D., & Reiss, A. J. (1967). On Exploring the "Dark Figure" of Crime. *The ANNALS of the American Academy of Political and Social Science, 374*(1), 1-15. doi: 10.1177/000271626737400102

Channels, N. L. (1993). Anticipating Media Coverage: Methodological Decisions in Criminal Justice Research. In C. M. Renzetti & R. M. Lee (Eds.), *Researching Sensitive Topics* (pp. 267-280). Newbury Park, California: Sage.

Douglas, J. D. (1976). *Investigative Social Research: Individual and team field research*. Oxford: Sage.

Fielding, N. G. (1993). Mediating the Message: Affinity and Hostility in Research on Sensitive Topics. In C. M. Renzetti & R. M. Lee (Eds.), *Researching Sensitive Topics* (pp. 146-159). Newbury Park, California: Sage.

Forsberg, E.-M., Anthun, F. O., Bailey, S., Birchley, G., Bout, H., Casonato, C., . . . Zöller, M. (2018). Working with Research Integrity—Guidance for Research Performing Organisations: The Bonn PRINTEGER Statement. *Science and Engineering Ethics, 24*(4), 1023-1034. doi: 10.1007/s11948-018-0034-4

Gao, J., & Zhou, T. (2017). Stamp out fake peer review. *Nature, 546*(7656), 33-33. doi: 10.1038/546033a

Hepkema, W. (2019). *The effect of journal guidelines on the use of misidentified cell lines in scientific literature.* (Master of Science), Radboud University, Nijmegen.

Horbach, S. P. J. M. (2018, 22-10-2018). Universiteiten, geef het goede voorbeeld en stimuleer integer onderzoek. *Trouw*. Retrieved from www.trouw.nl/opinie/universiteiten-geef-het-goede-voorbeeld-en-stimuleer-integer-onderzoek~beee5ffa/

Horbach, S. P. J. M., & Halffman, W. (2019). Rejoinder to 'Is there a Peter Nijkamp effect?' by Jasper Lukkezen. *Research Policy, 48*(2), 505. doi: https://doi.org/10.1016/j.respol.2018.10.024

Horbach, S. P. J. M., Halffman, W., De Rijcke, S., Reyes Elizondo, A., & Van Leeuwen, T. (2018). Leiden declaration on transparent editorial policies. from www.ru.nl/transparencydeclaration

Keulemans, M. (2017). Voor dit onderzoek zijn de verkeerde cellen gebruikt. *de Volkskrant*.

Lee, R. M., & Renzetti, C. M. (1990). The Problems of Researching Sensitive Topics: An Overview and Introduction. *American Behavioral Scientist, 33*(5), 510-528. doi: 10.1177/0002764290033005002

Lukkezen, J. H. (2019). Is there a Peter Nijkamp effect?: A comment on The extent and causes of academic text recycling or 'self-plagiarism' by Serge Horbach and Willem Halffman. *Research Policy, 48*(2), 503-504. doi: https://doi.org/10.1016/j.respol.2018.10.023

Matza, D., & Blomberg, T. G. (2010). *Becoming Deviant*. New York: Routledge.

Renzetti, C. M., & Lee, R. M. (1993). *Researching Sensitive Topics*. Newbury Park, California: Sage.

Sieber, J. E. (1993). The Ethics and Politics of Sensitive Research. In C. M. Renzetti & R. M. Lee (Eds.), *Researching Sensitive Topics* (pp. 14-26). Newbury Park, California: Sage.

Sieber, J. E., & Stanley, B. (1988). Ethical and professional dimensions of socially sensitive research. *American Psychologist, 43*(1), 49.
Stapel, D. (2012). *Ontsporing*. Amsterdam: Prometheus.
Stern, V. (2017). Publisher issues first retractions for fake peer review, starts new checking policy. Retrieved from http://retractionwatch.com/2017/11/21/frontiers-issues-first-retractions-fake-peer-review-starts-new-checking-policy/
Sykes, G. M., & Matza, D. (1957). Techniques of neutralization: A theory of delinquency. *American Sociological Review, 22*(6), 664-670.
Thelwall, M., Haustein, S., Lariviere, V., & Sugimoto, C. R. (2013). Do Altmetrics Work? Twitter and Ten Other Social Web Services. *Plos One, 8*(5). doi: 10.1371/journal.pone.0064841
Thomas, J. J. (1993). Measuring the Underground Economy: A Suitable Case for Interdisciplinary Treatment? In C. M. Renzetti & R. M. Lee (Eds.), *Researching Sensitive Topics* (pp. 52-70). Newbury Park, California: Sage.
Wilsdon, J. (2018, 14-12-2018). *The role of researchers in valueing science.* Paper presented at the On Quality - Symposium on the occasion of Paul Wouters' farewell as director of CWTS, Leiden.

Reference list

Aagaard, K. (2016). New and persistent gender equality challenges in academia. *Scandinavian Journal of Public Administration, 20*(1), 87-90.

Aagaard, K., Bloch, C., & Schneider, J. W. (2015). Impacts of performance-based research funding systems: The case of the Norwegian Publication Indicator. *Research Evaluation, 24*(2), 106-117. doi: 10.1093/reseval/rvv003

Abma, R. (2013). *De publicatiefabriek: Over de betekenis van de affaire Stapel*. Nijmegen: Uitgeverij Vantilt.

Adler, P. A., & Adler, P. (1993). Ethical Issues in Self-Censorship: Ethnographic Research on Sensitive Topics. In C. M. Renzetti & R. M. Lee (Eds.), *Researching Sensitive Topics* (pp. 249-266). Newbury Park, California: Sage.

Agnew, R. (1994). The techniques of neutralization and violence. *Criminology, 32*(4), 555-580.

Al-Marzouki, S., Evans, S., Marshall, T., & Roberts, I. (2005). Are these data real? Statistical methods for the detection of data fabrication in clinical trials. *BMJ, 331*(7511), 267-270. doi: 10.1136/bmj.331.7511.267

ALLEA. (2017). The European Code of Conduct for Research Intregrity - Revised Edition. In A.-A. E. Academies (Ed.). Berlin, Germany.

Allison, D. B., Brown, A. W., George, B. J., & Kaiser, K. A. (2016). A tragedy of errors. *Nature, 530*(7588), 27-29.

Altman, D. G. (1998). Statistical reviewing for medical journals. *Statistics in Medicine, 17*(23), 2661-2674. doi: 10.1002/(sici)1097-0258(19981215)17:23<2661::aid-sim33>3.0.co;2-b

American Sociological Review. (1955). Notice to contributors. *American Sociological Review, 20*(3), 341.

American Type Culture Collection Standards Development Organization Workgroup ASN-0002. (2010). Cell line misidentification: the beginning of the end. *Nat Rev Cancer, 10*(6), 441-448. doi: 10.1038/nrc2852

Amsen, E. (2014). What is open peer review. Retrieved from https://blog.f1000.com/2014/05/21/what-is-open-peer-review/

Andersen, L. B., & Pallesen, T. (2008). “Not just for the money?” How financial incentives affect the number of publications at Danish research institutions. *International Public Management Journal, 11*(1), 28-47.

Anderson, M. S., Ronning, E. A., De Vries, R., & Martinson, B. C. (2007). The perverse effects of competition on scientists’ work and relationships. *Science and Engineering Ethics, 13*(4), 437-461.

Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E., . . . Stockinger, H. (2012, Jul). ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res.* 2012/06/05. Retrieved Web Server issue, 40, from http://web.expasy.org/cellosaurus/

Ashforth, B. E., Gioia, D. A., Robinson, S. L., & Trevino, L. K. (2008). Re-viewing orgnizational corruption. *Academy of Management Review, 33*, 670-684.

Auer, N. J., & Krupar, E. M. (2001). Mouse click plagiarism: The role of technology in plagiarism and the librarian's role in combating it. *Library Trends, 49*(3), 415.

Azoulay, P., Bonatti, A., & Krieger, J. L. (2017). The career effects of scandal: Evidence from scientific retractions. *Research Policy, 46*(9), 1552-1569. doi: https://doi.org/10.1016/j.respol.2017.07.003

Babic, Z., Capes-Davis, A., Martone, M. E., Bairoch, A., Ozyurt, I. B., Gillespie, T. H., & Bandrowski, A. E. (2019). Incidences of problematic cell lines are lower in papers that use RRIDs to identify cell lines. *Elife, 8*, e41676. doi: 10.7554/eLife.41676

Bachrach, P., & Baratz, M. S. (1962). The Two Faces of Power. *American Political Science Review, 56*, 941-952.

Bakker, M., & Wicherts, J. M. (2011). The (mis)reporting of statistical results in psychology journals. *Behavior Research Methods, 43*(3), 666-678. doi: 10.3758/s13428-011-0089-5

Baldwin, M. (2013). 'Keeping in the race': physics, publication speed and national publishing strategies in Nature, 1895–1939. *The British Journal for the History of Science, 47*(2), 257-279. doi: 10.1017/S0007087413000381

Baldwin, M. (2015). Credibility, peer review, and Nature, 1945–1990. *Notes and Records: the Royal Society journal of the history of science, 69*(3), 337-352. doi: 10.1098/rsnr.2015.0029

Baldwin, M. (2017). In referees we trust? *Physics Today, 70*(2), 44.

Baldwin, M. (2018). Scientific Autonomy, Public Accountability, and the Rise of "Peer Review" in the Cold War United States. *Isis, 109*(3), 538-558. doi: 10.1086/700070

Barroga, E. F. (2013). Cascading peer review for open-access publishing. *Eur Sci Ed, 39*, 90-91.

Basten, F. (2013). Sorry, wetenschap, ik heb verzaakt #plagiaat. from https://www.floorbasten.nl/sorry-wetenschap-ik-heb-verzaakt-plagiaat

Bauchner, H., Fontanarosa, P. B., Flanagin, A., & Thornton, J. (2018). Scientific misconduct and medical journals. *Jama*. doi: 10.1001/jama.2018.14350

Bazdaric, K., Bilic-Zulle, L., Brumini, G., & Petrovecki, M. (2012). Prevalence of Plagiarism in Recent Submissions to the Croatian Medical Journal. *Science and Engineering Ethics, 18*(2), 223-239. doi: 10.1007/s11948-011-9347-2

Begley, C. G., & Ioannidis, J. P. (2015). Reproducibility in science. *Circulation research, 116*(1), 116-126.

Ben-Yehuda, N., & Oliver-Lumerman, A. (2017). *Fraud and Misconduct in Research: Detection, Investigation, and Organizational Response*: University of Michigan Press.

Benedek, E. P. (1976). Editorial practices of psychiatric and related journals: implications for women. *American Journal of Psychiatry, 133*(1), 89-92.

Benya, F. (2019). Treating Sexual Harassment as a Violation of Research Integrity. *Issues in Science and Technology, 35*(2), 56-59.

Berggren, C., & Karabag, S. F. (2019). Scientific misconduct at an elite medical institute: The role of competing institutional logics and fragmented control. *Research Policy, 48*(2), 428-443. doi: 10.1016/j.respol.2018.03.020

Biagioli, M. (2002). From book censorship to academic peer review. *Emergences: Journal for the Study of Media & Composite Cultures, 12*(1), 11-45.

Biagioli, M., Kenney, M., Martin, B. R., & Walsh, J. P. (2019). Academic misconduct, misrepresentation and gaming: A reassessment. *Research Policy, 48*(2), 401-413. doi: 10.1016/j.respol.2018.10.025

Biderman, A. D., & Reiss, A. J. (1967). On Exploring the "Dark Figure" of Crime. *The ANNALS of the American Academy of Political and Social Science, 374*(1), 1-15. doi: 10.1177/000271626737400102

Bijker, W. E., Hughes, T. P., & Pinch, T. J. (1987). *The social construction of technological systems: New directions in the sociology and history of technology*. Cambridge, Massachusetts: MIT press.

Binder, C. (1990). SELF-PLAGIARISM AND MULTIPLE PUBLICATIONS - NEGLIGENCE OR EPIDEMIC DISEASE. *Acta Endocrinologica, 122*(1), 7-7.

BioMed Central. (2017). What might peer review look like in 2030? In B. Central (Ed.). figshare.

Bjørkelo, B., & Matthiesen, S. B. (2012). Preventing and Dealing with Retaliation Against Whistleblowers. In W. Vandekerckhove & D. Lewis (Eds.), *Whistleblowing and Democratic Values*: The international whistleblowing research network.

Bloch, C., & Schneider, J. W. (2016). Performance-based funding models and researcher behavior: An analysis of the influence of the Norwegian Publication Indicator at the individual level. *Research Evaluation, 25*(4), 371-382. doi: 10.1093/reseval/rvv047

Bloor, D. (1991). *Knowledge and social imagery*: University of Chicago Press.

BMJ Open. (2018). Reviewer guidelines. from http://bmjopen.bmj.com/pages/reviewerguidelines/

Bohannon, J. (2013). Who's Afraid of Peer Review? *Science, 342*(6154), 60-65.

Bohlin, I. (2004). Communication Regimes in Competition. *Social Studies of Science, 34*(3), 365-391. doi: doi:10.1177/0306312704041522

Bonito, A. J., Titus, S. L., & Wright, D. E. (2012). Assessing the preparedness of Research Integrity Officers (RIOs) to appropriately handle possible research misconduct cases. *Science and Engineering Ethics, 18*(4), 605-619.

Boonstra, J. J., van Marion, R., Beer, D. G., Lin, L., Chaves, P., Ribeiro, C., . . . Altorki, N. K. (2010). Verification and unmasking of widely used human esophageal adenocarcinoma cell lines. *Journal of the National Cancer Institute, 102*(4), 271-274.

Bornmann, L. (2011). Scientific peer review. *Annual Review of Information Science and Technology, 45*(1), 197-245. doi: 10.1002/aris.2011.1440450112

Bouter, L. M., Tijdink, J., Axelsen, N., Martinson, B. C., & ter Riet, G. (2016). Ranking major and minor research misbehaviors: results from a survey among participants of four World Conferences on Research Integrity. *Research Integrity and Peer Review, 1*(1), 17.

Bowie, N. E. (2010). Organizational integrity and moral climates *The oxford handbook of business ethics*: Oxford University Press.

Bowker, G. C., & Star, S. L. (2000). Invisible Mediators of Action: Classification and the Ubiquity of Standards. *Mind, Culture, and Activity, 7*(1-2), 147-163. doi: 10.1080/10749039.2000.9677652

Bowker, G. C., & Star, S. L. (2000). *Sorting things out: Classification and its consequences*: MIT press.

Bozeman, B., & Youtie, J. (2016). Trouble in Paradise: Problems in Academic Research Co-authoring. *Science and Engineering Ethics, 22*(6), 1717-1743. doi: 10.1007/s11948-015-9722-5

Bravo, G., Grimaldo, F., López-Iñesta, E., Mehmani, B., & Squazzoni, F. (2019). The effect of publishing peer review reports on referee behavior in five scholarly journals. *Nature Communications, 10*(1), 322. doi: 10.1038/s41467-018-08250-2
Breedveld, P. (2015). Leidse hoogleraar wil internationale discussie over Nijkamp. from http://www.advalvas.vu.nl/nieuws/leidse-hoogleraar-wil-internationale-discussie-over-nijkamp
Bretag, T., & Carapiet, S. (2007). A Preliminary Study to Identify the Extent of Self-Plagiarism in Australian Academic Research. Plagiary: cross-disciplinary studies in plagiarism, fabrication, and falsification. *MPublishing, University of Michigan Library, Ann Arbor, 2*.
Bretag, T., & Mahmud, S. (2009). Self-Plagiarism or Appropriate Textual Re-use? *Journal of Academic Ethics, 7*(3), 193-205. doi: 10.1007/s10805-009-9092-1
Broome, M. E., Pryor, E., Habermann, B., Pulley, L., & Kincaid, H. (2005). The Scientific Misconduct Questionnaire—Revised (SMQ-R): Validation and Psychometric Testing. *Accountability in Research, 12*(4), 263-280. doi: 10.1080/08989620500440253
Brown, N., & Michael, M. (2003). A Sociology of Expectations: Retrospecting Prospects and Prospecting Retrospects. *Technology Analysis & Strategic Management, 15*(1), 3-18. doi: 10.1080/0953732032000046024
Burnham, J. C. (1990). The evolution of editorial peer review. *Jama, 263*(10), 1323-1329. doi: 10.1001/jama.1990.03440100023003
Butler, L. (2003). Modifying publication practices in response to funding formulas. *Research Evaluation, 12*(1), 39-46. doi: 10.3152/147154403781776780
Callahan, J. L. (2014). Creation of a Moral Panic? Self-Plagiarism in the Academy. *Human Resource Development Review, 13*(1), 3-10. doi: 10.1177/1534484313519063
Callahan, J. L. (2017). The retrospective (im)moralization of self-plagiarism: Power interests in the social construction of new norms for publishing. *Organization, 25*(3), 305-319. doi: 10.1177/1350508417734926
Callaway, E. (2015). Faked peer reviews prompt 64 retractions. *Nature News*.
Callon, M. (1995). Four Models for the Dynamics of Science. In S. Jasanoff, G. E. Markle, J. C. Petersen, & T. Pinch (Eds.), *Handbook of Science and Technology Studies* (pp. 29-63). Thousand Oaks: SAGE.
Campbell, D. T. (1979). Assessing the impact of planned social change. *Evaluation and Program Planning, 2*(1), 67-90. doi: https://doi.org/10.1016/0149-7189(79)90048-X
Cao, H. H., Han, B., Hirshleifer, D., & Zhang, H. H. (2009). Fear of the unknown: Familiarity and economic decisions. *Review of Finance, 15*(1), 173-206.
Capes-Davis, A., & Neve, R. M. (2016). Authentication: A Standard Problem or a Problem of Standards? *PLOS Biology, 14*(6), e1002477. doi: 10.1371/journal.pbio.1002477
Capes-Davis, A., Theodosopoulos, G., Atkin, I., Drexler, H. G., Kohara, A., MacLeod, R. A., . . . Freshney, R. I. (2010). Check your cultures! A list of cross-contaminated or misidentified cell lines. *Int J Cancer, 127*(1), 1-8. doi: 10.1002/ijc.25242
Carlisle, J. B. (2017). Data fabrication and other reasons for non-random sampling in 5087 randomised, controlled trials in anaesthetic and general medical journals. *Anaesthesia, 72*(8), 944-952. doi: 10.1111/anae.13938
Casadevall, A., Steen, R. G., & Fang, F. C. (2014). Sources of error in the retracted scientific literature. *Faseb Journal, 28*(9), 3847-3855. doi: 10.1096/fj.14-256735

Cassematis, P. G., & Wortley, R. (2013). Prediction of Whistleblowing or Non-reporting Observation: The Role of Personal and Situational Factors. *Journal of business ethics, 117*(3), 615-634. doi: 10.1007/s10551-012-1548-3
Center for Open Science. (2018). Registered Reports: Peer review before results are known to align scientific values and practices - Participating Journals. Retrieved October 4, 2018, from https://cos.io/rr/
Chambers, C. D. (2013). Registered reports: a new publishing initiative at Cortex. *Cortex, 49*(3), 609-610.
Chambers, C. D., Feredoes, E., Muthukumaraswamy, S. D., & Etchells, P. (2014). Instead of" playing the game" it is time to change the rules: Registered Reports at AIMS Neuroscience and beyond. *AIMS Neuroscience, 1*(1), 4-17.
Channels, N. L. (1993). Anticipating Media Coverage: Methodological Decisions in Criminal Justice Research. In C. M. Renzetti & R. M. Lee (Eds.), *Researching Sensitive Topics* (pp. 267-280). Newbury Park, California: Sage.
Chrousos, G. P., Kalantaridou, S. N., Margioris, A. N., & Gravanis, A. (2012). The ‘self-plagiarism’oxymoron: can one steal from oneself? *European journal of clinical investigation, 42*(3), 231-232.
Cintas, P. (2016). Peer review: from recognition to improved practices. *Fems Microbiology Letters, 363*(12). doi: 10.1093/femsle/fnw115
Clair, J. A. (2015). Procedural Injustice in the System of Peer Review and Scientific Misconduct. *Academy of Management Learning & Education, 14*(2), 159-172. doi: 10.5465/amle.2013.0243
Clark, A. (2017). Results-free review: impressions from the first published articles. Retrieved from http://blogs.biomedcentral.com/bmcseriesblog/2017/06/12/results-free-review-impressions-from-the-first-published-article/
Claxton, L. D. (2005). Scientific authorship: Part 1. A window into scientific fraud? *Mutation Research/Reviews in Mutation Research, 589*(1), 17-30. doi: http://dx.doi.org/10.1016/j.mrrev.2004.07.003
cOAlition S. (2018). Accelerating the transition to full and immediate Open Access to scientific publications. from https://www.coalition-s.org/about/
Cokol, M., Ozbay, F., & Rodriguez-Esteban, R. (2008). Retraction rates are on the rise. *Embo Reports, 9*(1), 1. doi: 10.1038/sj.embor.7401143
College van Bestuur EUR. (2014). Besluit CvB naar aanleiding van melding plagiaat – 12 juni 2014. Rotterdam.
College van Bestuur EUR. (2015). Definitief besluit College van Bestuur naar aanleiding van melding plagiaat. In EUR (Ed.). Rotterdam.
Collins, H. (1992). *Changing order: Replication and induction in scientific practice*: University of Chicago Press.
Collins, H., & Pinch, T. (1993). *The Golem: What Everyone Should Know About Science* (Vol. 2nd). Croyden: Cambridge University Press.
Collins, R. (1968). Competition and social control in science: An essay in theory-construction. *Sociology of Education*, 123-140.
Consoli, L. (2006). Scientific misconduct and science ethics: A case study based approach. *Science and Engineering Ethics, 12*(3), 533-541.

Coombs, W. T. (2007). Protecting Organization Reputations During a Crisis: The Development and Application of Situational Crisis Communication Theory. *Corporate Reputation Review, 10*(3), 163-176. doi: 10.1057/palgrave.crr.1550049

Cronin, B. (2005). *The hand of science: Academic writing and its rewards*: Scarecrow Press.

Csiszar, A. (2016). Peer Review: Troubled from the start. *Nature, 532,* 306-308.

Culiberg, B., & Mihelic, K. K. (2017). The Evolution of Whistleblowing Studies: A Critical Review and Research Agenda. *Journal of business ethics, 146*(4), 787-803. doi: 10.1007/s10551-016-3237-0

Curfman, G. D., Morrissey, S., & Drazen, J. M. (2006a). EDITORIALS: Expression of Concern: Sudbø J et al. DNA Content as a Prognostic Marker in Patients with Oral Leukoplakia. N Engl J Med 2001;344:1270-8 and Sudbø J et al. The Influence of Resection and Aneuploidy on Mortality in Oral Leukoplakia. N Engl J Med 2004;350:1405-13. *The New England Journal of Medicine, 354*(6), 638.

Curfman, G. D., Morrissey, S., & Drazen, J. M. (2006b). Retraction: Sudbø J et al. DNA Content as a Prognostic Marker in Patients with Oral Leukoplakia. N Engl J Med 2001;344:1270-8 and Sudbø J et al. The Influence of Resection and Aneuploidy on Mortality in Oral Leukoplakia. N Engl J Med 2004;350:1405-13. *The New England Journal of Medicine, 355*(18), 1927-1927. doi: 10.1056/NEJMe068247

CWTS. (2018). CWTS Leiden Ranking - Fields. from http://www.leidenranking.com/information/fields

Daston, L., & Galison, P. (2007). *Objectivity*. Cambridge: MIT Press.

Davies, G., & Olmedo-Cifuentes, I. (2016). Corporate misconduct and the loss of trust. *European Journal of Marketing, 50*(7-8), 1426-1447. doi: 10.1108/ejm-11-2014-0729

Davis, P. (2010). Cascading Peer-Review: The future of open access publishing? Retrieved from https://scholarlykitchen.sspnet.org/2010/10/12/cascading-peer-review-future-of-open-access/

Davis, P. (2017). PLOS ONE Output Drops Again In 2016. Retrieved from https://scholarlykitchen.sspnet.org/2017/01/05/plos-one-output-drops-again-in-2016/

De Boer, H., Enders, J., & Schimank, U. (2007). On the way towards new public management? The governance of university systems in England, the Netherlands, Austria, and Germany. *New forms of governance in research organizations*, 137-152.

De Bruijn, H. (2002). Performance measurement in the public sector: strategies to cope with the risks of performance measurement. *International Journal of Public Sector Management, 15*(7), 578-594.

de Oliveira, D. E., Marques, C. S., & Losi, V. C. (2016). "Cell identity" crisis: Another call for immediate action. *Cancer Letters, 381*(1), 122-123. doi: 10.1016/j.canlet.2016.07.031

de Rijcke, S. (2019). *Wetenschap, op waarde geschat.* (Inaugural lecture for professorship), Leiden University, Leiden. Retrieved from https://www.universiteitleiden.nl/binaries/content/assets/sociale-wetenschappen/cwts/oratie-de-rijcke-extra.pdf

de Solla Price, D. J. (1963). *Little science, big science*: Columbia University Press New York.

de Vasconcelos, S. M. R., & Roig, M. (2015). Prior Publication and Redundancy in Contemporary Science: Are Authors and Editors at the Crossroads? *Science and Engineering Ethics, 21*(5), 1367-1378. doi: 10.1007/s11948-014-9599-8

de Wit, O., van den Ende, J., Schot, J., & van Oost, E. (2002). Innovation Junctions: Office Technologies in the Netherlands, 1880-1980. *Technology and Culture, 43*(1), 50-72.
Degn, L. (2018). Academic sensemaking and behavioural responses – exploring how academics perceive and respond to identity threats in times of turmoil. *Studies in Higher Education, 43*(2), 305-321. doi: 10.1080/03075079.2016.1168796
Dickersin, K. (1990). The existence of publication bias and risk factors for its occurrence. *Jama, 263*(10), 1385-1389.
Directory of Open Access Journals. (2018). Search. Retrieved October 10, 2018, from https://doaj.org/search#.W74NsPkzaUk
Doelder, H. d., Bensing, J. M., & Bosch, F. A. J. v. d. (2014). Rapport Commissie Wetenschappelijke Integriteit: Melding plagiaat in proefschrift - april 2014. In EUR (Ed.). Rotterdam: EUR.
Douglas, J. D. (1976). *Investigative Social Research: Individual and team field research*. Oxford: Sage.
Douglas, M. (1966). *Purity and Danger: An Analysis of Concepts of Pollution and Taboo*. London: Routledge.
Dozier, J. B., & Miceli, M. P. (1985). Potential predictors of whistleblowing - a pro-social behavior perspective. *Academy of Management Review, 10*(4), 823-836. doi: 10.2307/258050
Drenth, P. J. D. (2006). Responsible conduct in research. *Science and Engineering Ethics, 12*(1), 13-21. doi: 10.1007/pl00022265
Drexler, H. G., Dirks, W. G., Matsuo, Y., & MacLeod, R. A. F. (2003). False leukemia-lymphoma cell lines: an update on over 500 cell lines. *Leukemia, 17*(2), 416-426. doi: 10.1038/sj.leu.2402799
Easterbrook, P. J., Gopalan, R., Berlin, J., & Matthews, D. R. (1991). Publication bias in clinical research. *The Lancet, 337*(8746), 867-872.
Eckberg, D. L. (1991). When nonreliability of reviews indicates solid science. *Behavioral and Brain Sciences, 14*(1), 145-146. doi: 10.1017/S0140525X00065791
Edwards, P. N. (2010). *A vast machine: Computer models, climate data, and the politics of global warming*: Mit Press.
Eisenhardt, K. M., & Graebner, M. E. (2007). Theory building from cases: Opportunities and challenges. *Academy of management Journal, 50*(1), 25-32.
Ekbom, A. (2006). Report from the Investigation Commision appointed by rikshospitalet - Radiumhospitalet MC and the University of Oslo January 18, 2006.
Elizondo, A. R., De Rijcke, S., & Van Leeuwen, T. (2017). Handling publishing misconduct: tools used by publishing houses and editors. In PRINTEGER (Ed.). Leiden: Leiden University.
EMBO Press. (2017). About: Cross Peer Review. *EMBO Press.* from http://emboj.embopress.org/about#Cross_peer_review
Enders, & Hoover, G. A. (2004). Whose Line Is It? Plagiarism in Economics. *Journal of Economic Literature, 42*(2), 487-493. doi: doi: 10.1257/0022051041409066
Enserink, M. (2018). Research on research. *Science, 361*(6408), 1178-1179. doi: 10.1126/science.361.6408.1178
Epskamp, S., & Nuijten, M. (2014). statcheck: Extract statistics from articles and recompute p values (R package version 1.0. 0.).

Reference list

Ercegovac, Z., & Richardson, J. V. (2004). Academic dishonesty, plagiarism included, in the digital age: A literature review. *College & Research Libraries, 65*(4), 301-318.

Erren, T. C. (2009). On establishing priority of ideas: Revisiting the “pli cacheté” (deposition of a sealed envelope). *Medical Hypotheses, 72*(1), 8-10. doi: http://dx.doi.org/10.1016/j.mehy.2008.08.013

ESF/ALLEA. (2011). The European Code of Conduct for Research Integrity. Strasbourg: European Science Foundation & All-European Academies.

Etzkowitz, H. (1998). The norms of entrepreneurial science: cognitive effects of the new university–industry linkages. *Research Policy, 27*(8), 823-833. doi: https://doi.org/10.1016/S0048-7333(98)00093-6

EUR. (2013). Klachtenregeling Wetenschappelijke Integriteit EUR (KWI-EUR). Rotterdam.

Faidhi, J. A., & Robinson, S. K. (1987). An empirical approach for detecting program similarity and plagiarism within a university programming environment. *Computers & Education, 11*(1), 11-19.

Fanelli, D. (2009). How Many Scientists Fabricate and Falsify Research? A Systematic Review and Meta-Analysis of Survey Data. *Plos One, 4*(5), 11. doi: 10.1371/journal.pone.0005738

Fanelli, D. (2016). Set up a 'self-retraction' system for honest errors. *Nature, 531*(7595), 415-415.

Fanelli, D., Costas, R., Fang, F. C., Casadevall, A., & Bik, E. M. (2017). Why Do Scientists Fabricate And Falsify Data? A Matched-Control Analysis Of Papers Containing Problematic Image Duplications. *bioRxiv*. doi: 10.1101/126805

Fanelli, D., Costas, R., & Lariviere, V. (2015). Misconduct Policies, Academic Culture and Career Stage, Not Gender or Pressures to Publish, Affect Scientific Integrity. *Plos One, 10*(6), 18. doi: 10.1371/journal.pone.0127556

Fang, F. C., Steen, R. G., & Casadevall, A. (2012). Misconduct accounts for the majority of retracted scientific publications. *Proceedings of the National Academy of Sciences of the United States of America, 109*(42), 17028-17033. doi: 10.1073/pnas.1212247109

Faria, R. (2014). Science under pressure: problematic behaviors and social harms. *Revista Crítica Penal y Poder*(7), 64-84.

Faria, R. (2015). Scientific misconduct: how organizational culture plays its part. *Tijdschrift over Cultuur & Criminaliteit, 5*(1), 38-54.

Fernandez, E., Vicente, A., Zapata, A., Brera, B., Lozano, J. J., Martinez, C., & Toribio, M. L. (1994). Establishment and Characterization of Cloned Human Thymic Epithelial - Cell Lines - Analysis of Adhesion Mulecule Expression and Cytokine Production *Blood, 83*(11), 3245-3254.

Fielding, N. G. (1993). Mediating the Message: Affinity and Hostility in Research on Sensitive Topics. In C. M. Renzetti & R. M. Lee (Eds.), *Researching Sensitive Topics* (pp. 146-159). Newbury Park, California: Sage.

Fitzpatrick, K. (2009). The future of peer review. In K. Fitzpatrick (Ed.), *Planned Obsolescene: Publishing, Technology, and the Future of the Academy*. New York: NYU Press.

Fitzpatrick, K. (2010). Peer-to-peer Review and the Future of Scholarly Authority. *Social Epistemology, 24*(3), 161-179. doi: 10.1080/02691728.2010.498929

Fletcher, R., Hall, S., Corbera, E., Leung, M. W. H., MacLeavy, J., Neo, H., . . . Turner, S. (2019). The future of scholarly publishing: Paywalls and profits or a new plan? *Geoforum, 102*, 1-4. doi: https://doi.org/10.1016/j.geoforum.2019.03.005

Forsberg, E.-M., Anthun, F. O., Bailey, S., Birchley, G., Bout, H., Casonato, C., . . . Zöller, M. (2018). Working with Research Integrity—Guidance for Research Performing Organisations: The Bonn PRINTEGER Statement. *Science and Engineering Ethics, 24*(4), 1023-1034. doi: 10.1007/s11948-018-0034-4

Forskerforum. (2015). anklager om plagiat på kristen høgskole. In Forskerforum (Ed.). Olso.

Franke, N., & Shah, S. (2003). How communities support innovative activities: an exploration of assistance and sharing among end-users. *Research Policy, 32*(1), 157-178. doi: https://doi.org/10.1016/S0048-7333(02)00006-9

Franke, R. H., & Kaul, J. D. (1978). The Hawthorne Experiments: First Statistical Interpretation. *American Sociological Review, 43*(5), 623-643. doi: 10.2307/2094540

Freedman, L. P., Cockburn, I. M., & Simcoe, T. S. (2015). The Economics of Reproducibility in Preclinical Research. *PLOS Biology, 13*(6), e1002165. doi: 10.1371/journal.pbio.1002165

French, J., Raven, B., & Cartwright, D. (1959). The bases of social power. *Classics of organization theory, 7*, 311-320.

Freshney, R. I. (2002). Cell line provenance. *Cytotechnology, 39*(2), 55-67.

Frontiers. (2014). Frontiers is a community-rooted, open-access academic publisher. *About Frontiers.* from https://www.frontiersin.org/about/reviewsystem

Funtowicz, S., & Ravetz, J. R. (2001). Peer review and quality control. *International Encyclopaedia of the Social and Behavioural Sciences, Elsevier*, 11179-11183.

Fyfe, A., Coate, K., Curry, S., Lawson, S., Moxham, N., & Rostvik, C. M. (2017). Untangling academic publishing: a history of the relationship between commercial interests, academic prestige and the circulation of research.

Fyfe, A., McDougall-Waters, J., & Moxham, N. (2015). 350 years of scientific periodicals. *Notes and Records: the Royal Society journal of the history of science, 69*(3), 227-239. doi: 10.1098/rsnr.2015.0036

Gangloff, K. A. (2014). *Misconduct and the Media: Organizational Sensegiving, Media Coverage, and Investor Reaction in the Aftermath of Wrongdoing.* (Doctor of Philosophy), Auburn University, Auburn, Alabama.

Gangloff, K. A., & Connelly, B. L. (2015). *Walking the Talk: Organizational Sensegiving in the Aftermath of Misconduct.* Paper presented at the Academy of Management Proceedings.

Gao, J., & Zhou, T. (2017). Stamp out fake peer review. *Nature, 546*(7656), 33-33. doi: 10.1038/546033a

Gao, J. Y., Greenberg, R., & Wong-On-Wing, B. (2015). Whistleblowing Intentions of Lower-Level Employees: The Effect of Reporting Channel, Bystanders, and Wrongdoer Power Status. *Journal of business ethics, 126*(1), 85-99. doi: 10.1007/s10551-013-2008-4

Garfield, E. (1983). More on the ethics of scientific publication: abuses of authorship attribution and citation amnesia undermine the reward system of science. *Essays of an Information Scientist, 5*, 1981-1982.

Garfield, E., & Welljams-Dorof, A. (1990). The impact of fraudulent research on the scientific literature: The Stephen E. Breuning case. *Jama, 263*(10), 1424-1426.

Gartler, S. M. (1967). Genetic markers as tracers in cell culture. *National Cancer Institute Monograph, 26*, 167-195.

Gartler, S. M. (1968). Apparent HeLa Cell Contamination of Human Heteroploid Cell Lines. *Nature, 217*(5130), 750-750. doi: 10.1038/217750a0

George, S. L. (1985). STATISTICS IN MEDICAL JOURNALS - A SURVEY OF CURRENT POLICIES AND PROPOSALS FOR EDITORS. *Medical and Pediatric Oncology, 13*(2), 109-112. doi: 10.1002/mpo.2950130215

Gilbert, D. A., Reid, Y., Gail, M. H., Pee, D., White, C., Hay, R. J., & O'Brien, S. J. (1990). Application of DNA fingerprints for cell-line individualization. *American journal of human genetics, 47*(3), 499.

Gioia, D. A., & Chittipeddi, K. (1991). Sensemaking and sensegiving in strategic change initiation. *Strategic management journal, 12*(6), 433-448.

Gjerding, M. L., & Utheim, E. B. (2013, 25.01.2013 (Accessed 04.09.2017)). Avslørt for forskningssvindel - nå vil han forske igjen. *Verdens Gang (Internet version)*. Retrieved from http://www.vg.no/nyheter/innenriks/helse/avsloert-for-forskningssvindel-naa-vil-han-forske-igjen/a/10106865/

Godecharle, S., Nemery, B., & Dierickx, K. (2014). Heterogeneity in European Research Integrity Guidance: Relying on Values or Norms? *Journal of Empirical Research on Human Research Ethics, 9*(3), 79-90. doi: 10.1177/1556264614540594

Godlee, F. (2002). Making reviewers visible: openness, accountability, and credit. *Jama, 287*(21), 2762-2765.

Gold, M. (1986). *A Conspiracy of Cells: One Woman's Immortal Legacy-And the Medical Scandal It Caused*: SUNY Press.

Goodman, S. N. (2017). *Statistical Review in Biomedical Journals: Can We Get SMARTA?* Paper presented at the Eight International Congress on Peer Review and Scientific Publication, Chicago.

Graham, K. (2014). Thanking Our Peer Reviewers. Retrieved from http://blogs.plos.org/everyone/2014/01/06/thanking-peer-reviewers/

Greenhalgh, T., Robert, G., Macfarlane, F., Bate, P., & Kyriakidou, O. (2004). Diffusion of innovations in service organizations: Systematic review and recommendations. *Milbank Quarterly, 82*(4), 581-629. doi: 10.1111/j.0887-378X.2004.00325.x

Grens, K. (2015). The Great Big Clean-Up. *Scientist, 29*(9), 50-55.

Grens, K. (2016, 31-08-2016). Popular Tumor Cell Line Contaminated. *The Scientist.* from http://www.the-scientist.com/?articles.view/articleNo/46929/title/Popular-Tumor-Cell-Line-Contaminated/

Greve, H. R., Palmer, D., & Pozner, J. E. (2010). Organizations Gone Wild: The Causes, Processes, and Consequences of Organizational Misconduct. *The Academy of Management Annals, 4*(1), 53-107. doi: 10.1080/19416521003654186

Grey, A., Bolland, M., Gamble, G., & Avenell, A. (2019). Quality of reports of investigations of research integrity by academic institutions. *Research Integrity and Peer Review, 4*(1), 3. doi: 10.1186/s41073-019-0062-x

Grilli, J., & Allesina, S. (2017). Last name analysis of mobility, gender imbalance, and nepotism across academic systems. *Proceedings of the National Academy of Sciences, 114*(29), 7600-7605.

Guédon, J.-C. (2001). *In Oldenburg's long shadow: Librarians, research scientists, publishers, and the control of scientific publishing*: Association of Research Libr.
Gunnarsdottir, K. (2005). Scientific journal publications: On the role of electronic preprint exchange in the distribution of scientific literature. *Social Studies of Science, 35*(4), 549-579.
Guston, D. (1999). *Between politics and science: assuring the productivity and integrity of research*. Cambridge: Cambridge University Press.
Guston, D. H. (2007). *Between politics and science: Assuring the integrity and productivity of reseach*: Cambridge University Press.
Habermann, B., Broome, M., Pryor, E. R., & Ziner, K. W. (2010). Research Coordinators Experiences with Scientific Misconduct and Research Integrity. *Nursing research, 59*(1), 51-57. doi: 10.1097/NNR.0b013e3181c3b9f2
Hackett, E. J. (1994). A social control perspective on scientific misconduct. *The Journal of Higher Education, 65*(3), 242-260.
Halffman, W. (2019). What Is Science? (And Why Does This Matter?). In E. Turnhout, W. Halffman, & W. Tuinstra (Eds.), *Environmental Expertise: Connecting Science, Policy and Society* (pp. 11-35). Cambridge: Cambridge University Press.
Halffman, W., & Radder, H. (2015). The Academic Manifesto: From an Occupied to a Public University. *Minerva, 53*(2), 165-187. doi: 10.1007/s11024-015-9270-9
Hall, E. (2016). RGC-5 is not rat retinal ganglion. *Case studies.* from http://iclac.org/case-studies/rgc-5/
Hames, I. (2014). The changing face of peer review. *Sci Ed, 1*(1), 9-12. doi: 10.6087/kcse.2014.1.9
Hammersley, M., & Atkinson, P. (2007). *Ethnography: Principles in practice*: Routledge.
Hansoti, B., Langdorf, M. I., & Murphy, L. S. (2016). Discriminating Between Legitimate and Predatory Open Access Journals: Report from the International Federation for Emergency Medicine Research Committee. *Western Journal of Emergency Medicine, 17*(5), 497-507. doi: 10.5811/westjem.2016.7.30328
Harboe, M. (2006). Dette er ikke vår medisinske etikk. *Tidsskrift For Den Norske Laegeforening: Tidsskrift For Praktisk Medicin, Ny Raekke, 126*(17), 2283.
Hardwicke, T. E., & Ioannidis, J. P. A. (2018). Populating the Data Ark: An attempt to retrieve, preserve, and liberate data from the most highly-cited psychology and psychiatry articles. *Plos One, 13*(8), e0201856. doi: 10.1371/journal.pone.0201856
Harriman, S., & Patel, J. (2014). Text recycling: acceptable or misconduct? *Bmc Medicine, 12*, 148. doi: 10.1186/s12916-014-0148-8
Harris, J., Graue, M., Dunning, T., Haltbakk, J., Austrheim, G., Skille, N., . . . Kirkevold, M. (2015). Involving people with diabetes and the wider community in diabetes research: a realist review protocol. *Systematic Reviews, 4*(1), 146. doi: 10.1186/s13643-015-0127-y
Haug, C. (2007). Erfaringer, forventninger og forsetter. *Tidsskrift For Den Norske Laegeforening: Tidsskrift For Praktisk Medicin, Ny Raekke, 127*(1), 13.
Haven, T. L., Tijdink, J. K., Martinson, B. C., & Bouter, L. M. (2019). Perceptions of research integrity climate differ between academic ranks and disciplinary fields: Results from a survey among academic researchers in Amsterdam. *Plos One, 14*(1), e0210599. doi: 10.1371/journal.pone.0210599

He, T. W. (2013). Retraction of global scientific publications from 2001 to 2010. *Scientometrics, 96*(2), 555-561. doi: 10.1007/s11192-012-0906-3

Hepkema, W. (2019). *The effect of journal guidelines on the use of misidentified cell lines in scientific literature.* (Master of Science), Radboud University, Nijmegen.

Hesselmann, F., Graf, V., Schmidt, M., & Reinhart, M. (2017). The visibility of scientific misconduct: A review of the literature on retracted journal articles. *Current Sociology, 65*(6), 814-845. doi: 10.1177/0011392116663807

Hessels, L. K., Franssen, T., Scholten, W., & de Rijcke, S. (2019). Variation in Valuation: How Research Groups Accumulate Credibility in Four Epistemic Cultures. *Minerva, 57*(2), 127-149. doi: 10.1007/s11024-018-09366-x

Hicks, D., Wouters, P., Waltman, L., De Rijcke, S., & Rafols, I. (2015). Bibliometrics: the Leiden Manifesto for research metrics. *Nature, 520*, 429-431.

Hiney, M. (2015). What it Means, Why it Is Important and How we Might Protect it. Brussels: Science Europe, working group research integrity.

Hirschauer, S. (2010). Editorial Judgments: A Praxeology of 'Voting' in Peer Review. *40*(1), 71-103. doi: 10.1177/0306312709335405

Honig, B., & Bedi, A. (2012). The fox in the hen house: A critical examination of plagiarism among members of the Academy of Management. *Academy of Management Learning & Education, 11*(1), 101-123.

Hoover, G. A. (2006). A game-theoretic model of plagiarism. *Atlantic Economic Journal, 34*(4), 449-454.

Hopp, C., & Hoover, G. A. (2017). How prevalent is academic misconduct in management research? *Journal of Business Research, 80*, 73-81. doi: https://doi.org/10.1016/j.jbusres.2017.07.003

Horbach, S. P. J. M. (2018, 22-10-2018). Universiteiten, geef het goede voorbeeld en stimuleer integer onderzoek. *Trouw*. Retrieved from www.trouw.nl/opinie/universiteiten-geef-het-goede-voorbeeld-en-stimuleer-integer-onderzoek~beee5ffa/

Horbach, S. P. J. M., & Halffman, W. (2017a). The ghosts of HeLa: How cell line misidentification contaminates the scientific literature. *Plos One, 12*(10), e0186281. doi: 10.1371/journal.pone.0186281

Horbach, S. P. J. M., & Halffman, W. (2017b). Promoting Virtue or Punishing Fraud: Mapping Contrasts in the Language of 'Scientific Integrity'. *Science and Engineering Ethics, 23*(6), 1461-1485. doi: 10.1007/s11948-016-9858-y

Horbach, S. P. J. M., & Halffman, W. (2018). The changing forms and expectations of peer review. *Research Integrity and Peer Review, 3*(1), 8. doi: 10.1186/s41073-018-0051-5

Horbach, S. P. J. M., & Halffman, W. (2019a). The extent and causes of academic text recycling or 'self-plagiarism'. *Research Policy, 48*(2), 492-502. doi: https://doi.org/10.1016/j.respol.2017.09.004

Horbach, S. P. J. M., & Halffman, W. (2019b). Rejoinder to 'Is there a Peter Nijkamp effect?' by Jasper Lukkezen. *Research Policy, 48*(2), 505. doi: https://doi.org/10.1016/j.respol.2018.10.024

Horbach, S. P. J. M., & Halffman, W. (under review). Journal Peer Review: Cautious Innovator or Sleepy Giant? *Minerva*.

Horbach, S. P. J. M., Halffman, W., De Rijcke, S., Reyes Elizondo, A., & Van Leeuwen, T. (2018). Leiden declaration on transparent editorial policies. from www.ru.nl/transparencydeclaration

Horbach, S. P. J. M., Nuijten, M., Tijdink, J., & O'Neil, G. (2019). Position Paper: Fostering Responsible Research Practices - Start document for ZonMw's BVO program. The Hague: ZonMw.

Horner, J., & Minifie, F. D. (2011). Research Ethics II: Mentoring, Collaboration, Peer Review, and Data Management and Ownership. *Journal of Speech Language and Hearing Research, 54*(1), S330-S345. doi: 10.1044/1092-4388(2010/09-0264)

Horton, R. (2006a). Expression of concern: non-steroidal anti-inflammatory drugs and the risk of oral cancer. *The Lancet, 367*(9506), 196-196. doi: 10.1016/S0140-6736(06)68014-8

Horton, R. (2006b). Retraction—Non-steroidal anti-inflammatory drugs and the risk of oral cancer: a nested case-control study. *The Lancet, 367*(9508), 382.

Huang, Y., Liu, Y., Zheng, C., & Shen, C. (2017). Investigation of Cross-Contamination and Misidentification of 278 Widely Used Tumor Cell Lines. *Plos One, 12*(1), e0170384. doi: 10.1371/journal.pone.0170384

ICLAC. (2016). Definitions. 2017, from http://iclac.org/resources/definitions/

Identity crisis. (2009). *Nature, 457*(7232), 935-936. doi: 10.1038/457935b

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Med, 2*(8), e124.

Jacob, M.-A. (2019). Under repair: A publication ethics and research record in the making. *Social Studies of Science, 49*(1), 77-101. doi: 10.1177/0306312718824663

Jellreys, A. (1985). Individual-specific ‘fingerprints’ of human DNA. *Nature, 316*, 4.

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling. *Psychological Science, 23*(5), 524-532. doi: 10.1177/0956797611430953

Joob, B., & Wiwanitkit, V. (2016). Scientific misconduct and self-plagiarism. *American Journal of Obstetrics and Gynecology, 214*(4), 549-550. doi: http://dx.doi.org/10.1016/j.ajog.2015.12.011

Kallio, K. P. (2017). Suble radical moves in scientific publishing. *Fennia - International Journal of Geography, 195*(1), 1-4. doi: https://doi.org/10.11143/fennia.63678

Karabag, S. F., & Berggren, C. (2012). Retraction, dishonesty and plagiarism: analysis of a crucial issue for academic publishing, and the inadequate responses from leading journals in economics and management disciplines. *Journal of Applied Economics and Business Research, 2*(3), 172-183.

Karabag, S. F., & Berggren, C. (2016). Misconduct, Marginality and Editorial Practices in Management, Business and Economics Journals. *Plos One, 11*(7), e0159492. doi: 10.1371/journal.pone.0159492

Kaushal, R., Bates, D. W., Landrigan, C., McKenna, K. J., Clapp, M. D., Federico, F., & Goldmann, D. A. (2001). Medication errors and adverse drug events in pediatric inpatients. *Jama, 285*(16), 2114-2120.

Kennefick, D. (2005). Einstein Versus the Physical Review. *Physics Today, 58*(9), 43-48. doi: 10.1063/1.2117822

Keulemans, M. (2017). Voor dit onderzoek zijn de verkeerde cellen gebruikt. *de Volkskrant*.

Kharasch, E. D., & Houle, T. T. (2018). Seeking and reporting apparent research misconduct: errors and integrity. *Anaesthesia, 73*(1), 125-126. doi: 10.1111/anae.14147

Kluge, L., Hafstad, A., & Torp, I. S. (2006, 27 January 2006). Jon Sudbø ble avslørt som juksemaker. Men han har selv avslørt hvordan forskningen kan lure en hel verden. *Aftenposten A-magasin*.

KNAW. (2014). CORRECT CITATION PRACTICE, Academy Advisory Memorandum: KNAW, Committee on Citation Practice.

KNAW, NFU, NWO, TO2-federatie, Verneniging Hogescholen, & VSNU. (2018). Nederlandse gedragscode wetenschappelijke integriteit.

Knoepfler, P. (2015). Reviewing post-publication peer review. *Trends in Genetics, 31*(5), 221-223.

Kochan, C. A., & Budd, J. M. (1992). The persistence of fraud in the literature: the Darsee case. *Journal of the American Society for Information Science, 43*(7), 488.

Kolfschooten, F. v. (2015, 30-06-2015). Erasmus Universiteit in Rotterdam vraagt psycholoog doctorsbul terug. *NRC Handelsblad*.

Komite for vurdering av klagesak ved NLA Høgskolen. (2014). Uttalelse fra komite for vurdering av klagesak ved NLA Høgskolen. Bergen: NLA Høgskolen.

Koshland, D. E. (1987). Fraud in Science. *Science, 235*(4785), 141-141. doi: 10.1126/science.3798097

Kovanis, M., Porcher, R., Ravaud, P., & Trinquart, L. (2016). The Global Burden of Journal Peer Review in the Biomedical Literature: Strong Imbalance in the Collective Enterprise. *Plos One, 11*(11), e0166387. doi: 10.1371/journal.pone.0166387

Krause, M. (2016). Comparative research: Beyond linear-causal explanation. In J. Deville, M. Guggenheim, & Z. Hrdlickova (Eds.), *Practising Comparison* (pp. 45-67). Manchester: Mattering Press.

Krippendorff, K., & Bock, M. A. (2009). *The content analysis reader*: Sage.

Kronick, D. A. (1976). A history of scientific and technical periodicals: the origins and development of the scientific and technological press, 1665-1790.

Laakso, M., & Björk, B.-C. (2012). Anatomy of open access publishing: a study of longitudinal development and internal structure. *Bmc Medicine, 10*(1), 124.

Lacroix, M. (2008). Persistent use of "false" cell lines. *Int J Cancer, 122*(1), 1-4. doi: 10.1002/ijc.23233

LaFollette, M. C. (1992). *Stealing into print: fraud, plagiarism, and misconduct in scientific publishing*: Univ of California Press.

LaFollette, M. C. (1994). The Politics of Research Misconduct: Congressional Oversight, Universities, and Science. *The Journal of Higher Education, 65*(3), 261-285. doi: 10.2307/2943967

Lamont, M. (2009). *How professors think: inside the curious world of academic judgment*. Cambridge, Massachusetts: Harvard University Press.

Larivière, V., Gingras, Y., & Archambault, É. (2009). The decline in the concentration of citations, 1900–2007. *Journal of the Association for Information Science and Technology, 60*(4), 858-862.

Larivière, V., Haustein, S., & Mongeon, P. (2015). The Oligopoly of Academic Publishers in the Digital Era. *Plos One, 10*(6), e0127502. doi: 10.1371/journal.pone.0127502

Larivière, V., Sugimoto, C. R., Macaluso, B., Milojević, S., Cronin, B., & Thelwall, M. (2014). arXiv E-prints and the journal of record: An analysis of roles and relationships. *Journal of the Association for Information Science and Technology, 65*(6), 1157-1169. doi: 10.1002/asi.23044

Leahey, E., & Montgomery, K. (2011). The meaning of regulation in a changing academic profession. *The American Academic Profession: Changing forms and functions*, 295-311.

Lee, C. J., & Moher, D. (2017). Promote scientific integrity via journal peer review data. *Science, 357*(6348), 256-257. doi: 10.1126/science.aan4141

Lee, C. J., Sugimoto, C. R., Zhang, G., & Cronin, B. (2013). Bias in peer review. *Journal of the Association for Information Science and Technology, 64*(1), 2-17.

Lee, C. J., Sugimoto, C. R., Zhang, G., & Cronin, B. (2013). Bias in peer review. *Journal of the American Society for Information Science and Technology, 64*(1), 2-17. doi: 10.1002/asi.22784

Lee, R. M., & Renzetti, C. M. (1990). The Problems of Researching Sensitive Topics: An Overview and Introduction. *American Behavioral Scientist, 33*(5), 510-528. doi: 10.1177/0002764290033005002

Leeflang, P. S. H., Dijk, H. K. v., & Bunt, H. G. v. d. (2015). Rapportage adviescommissie beoordeling plagiaat. In EUR (Ed.). Rotterdam: Erasmus University.

Liang-Chu, M. M., Yu, M., Haverty, P. M., Koeman, J., Ziegle, J., Lee, M., . . . Neve, R. M. (2015). Human biosample authentication using the high-throughput, cost-effective SNPtrace(TM) system. *Plos One, 10*(2), e0116218. doi: 10.1371/journal.pone.0116218

Lins, M. P., Vieira, L. F. D., Rosa, A. A. M., & Smaniotto, S. (2016). Growth hormone in the presence of laminin modulates interaction of human thymic epithelial cells and thymocytes in vitro. *Biological Research, 49*. doi: 10.1186/s40659-016-0097-0

Lock, S. (1985). *A difficult balance: editorial peer review in medicine*: Nuffield Provincial Hospitals Trust.

Loop, S. M., Rozanski, T. A., & Ostenson, R. C. (1993). Human Primary Prostate Tumor-Cell Line, Alva 31: A New Model for Studying the Hormonal-Regulation of Prostate Tumor-Cell Growth. *Prostate, 22*(2), 93-108. doi: 10.1002/pros.2990220202

Lorsch, J. R., Collins, F. S., & Lippincott-Schwartz, J. (2014). Fixing problems with cell lines. *Science, 346*(6216), 1452-1453. doi: 10.1126/science.1259110

LOWI. (2014). LOWI Advies 2014, nr. 10. The Hague.

LOWI. (2015). LOWI Advies 2015, nr. 2. Amsterdam.

LOWI. (2016). LOWI Advies 2016, nr.1. Amsterdam.

Luhmann, N. (1992). *Die wissenschaft der gesellschaft* (Vol. 9): Suhrkamp Frankfurt am Main.

Lukes, S. (2005). *Power: A Radical View* (2nd edition ed.). Basingstoke and New York: Palgrave MacMillan.

Lukkezen, J. H. (2019). Is there a Peter Nijkamp effect?: A comment on The extent and causes of academic text recycling or 'self-plagiarism' by Serge Horbach and Willem Halffman. *Research Policy, 48*(2), 503-504. doi: https://doi.org/10.1016/j.respol.2018.10.023

MacIntyre, A. (1985). *After virtue : a study in moral theory* (2nd (corrected) ed. (with postscript). ed.). London: Duckworth.

MacKenzie, D., & Wajcman, J. (1999). *The social shaping of technology*: Open university press.

MacLean, T. L., & Behnam, M. (2010). The dangers of decoupling: The relationship between compliance programs, legitimacy perceptions, and institutionalized misconduct. *Academy of management Journal, 53*(6), 1499-1520. doi: 10.5465/amj.2010.57319198

MacLeod, R. A. F., Dirks, W. G., Matsuo, Y., Kaufmann, M., Milch, H., & Drexler, H. G. (1999). Widespread intraspecies cross-contamination of human tumor cell lines arising at source. *Int J Cancer, 83*(4), 555-563.

MacLeod, R. A. F., & Drexler, H. G. (2002). Authentication and Characterization. In J. R. W. Masters & B. O. Palsson (Eds.), *Human Cell Culture: Cancer Cell Lines Part 3: Leukemias and Lymphomas* (pp. 371-397). Dordrecht: Springer Netherlands.

Madlock-Brown, C. R., & Eichmann, D. (2015). The (lack of) Impact of Retraction on Citation Networks. *Science and Engineering Ethics, 21*(1), 127-137. doi: 10.1007/s11948-014-9532-1

Mahoney, M. J. (1985). Open exchange and epistemic progress. *American Psychologist, 40*(1), 29.

Maitlis, S., & Lawrence, T. B. (2007). Triggers and enablers of sensegiving in organizations. *Academy of management Journal, 50*(1), 57-84.

Mamelund, S.-E., Breit, E., & Forsberg, E.-M. (2018). A multinational survey on research misconduct and integrity: a workfloor perspective (DIV.2). In PRINTEGER (Ed.). Oslo.

Manske, P. R. (1997). A review of peer review. *The Journal of Hand Surgery, 22*(5), 767-771. doi: http://dx.doi.org/10.1016/S0363-5023(97)80067-6

Marcus, A. (2019). Scientific integrity "a rare commodity these days:" Cancer journal makes its first retraction. Retrieved from https://retractionwatch.com/2019/03/29/scientific-integrity-a-rare-commodity-these-days-cancer-journal-makes-its-first-retraction/

Marin-Spiotta, E. (2018). Harassment should count as scientific misconduct. *Nature, 557*(7704), 141-141.

Martin, B. (1992). Scientific fraud and the power structure of science. *Prometheus, 10*(1), 83-98.

Martin, B. R. (2013). Whither research integrity? Plagiarism, self-plagiarism and coercive citation in an age of research assessment. *Research Policy, 42*(5), 1005-1014. doi: http://dx.doi.org/10.1016/j.respol.2013.03.011

Martinson, B. C., Anderson, M. S., & de Vries, R. (2005). Scientists behaving badly. *Nature, 435*(7043), 737-738. doi: 10.1038/435737a

Martinson, B. C., Crain, A. L., De Vries, R., & Anderson, M. S. (2010). The importance of organisational justice in ensuring research integrity. *Journal of Empirical Research on Human Research Ethics, 5*(3), 67-83. doi: 10.1525/jer.2010.5.3.67

Martinson, B. C., Thrush, C. R., & Lauren Crain, A. (2013). Development and Validation of the Survey of Organizational Research Climate (SORC). *Science and Engineering Ethics, 19*(3), 813-834. doi: 10.1007/s11948-012-9410-7

Masters, J. R. W. (2002). HeLa cells 50 years on: the good, the bad and the ugly. *Nature Reviews Cancer, 2*(4), 315-319.

Masters, J. R. W. (2010). Cell line misidentification: the beginning of the end. *Nature Reviews Cancer, 10*(6), 441-448. doi: 10.1038/nrc2852

Masters, J. R. W., & Iclac. (2012). End the scandal of false cell lines. *Nature, 492*(7428), 186-186.

Masters, J. R. W., Thomson, J. A., Daly-Burns, B., Reid, Y. A., Dirks, W. G., Packer, P., . . . Arlett, C. F. (2001). Short tandem repeat profiling provides an international reference standard

for human cell lines. *Proceedings of the National Academy of Sciences, 98*(14), 8012-8017.
Matza, D., & Blomberg, T. G. (2010). *Becoming Deviant*. New York: Routledge.
Mazur, A. (1989). Allegations of dishonesty in research and their treatment by American universities. *Minerva, 27*(2-3), 177-194.
McElreath, R., & Smaldino, P. E. (2015). Replication, Communication, and the Population Dynamics of Scientific Discovery. *Plos One, 10*(8), e0136088. doi: 10.1371/journal.pone.0136088
Mellor, D. (2016). Registered Reports: Peer review before results are known to align scientific values and practices., from https://cos.io/rr/
Merton, R. K. (1938). Social structure and anomie. *American Sociological Review, 3*(5), 672-682.
Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigations*: University of Chicago press.
Mesmer-Magnus, J. R., & Viswesvaran, C. (2005). Whistleblowing in organizations: An examination of correlates of whistleblowing intentions, actions, and retaliation. *Journal of business ethics, 62*(3), 277-297. doi: 10.1007/s10551-005-0849-1
Meyer, J. W., & Rowan, B. (1977). Institutionalized Organizations: Formal Structure as Myth and Ceremony. *American Journal of Sociology, 83*(2), 340-363.
Minor, W. W. (1981). Techniques of neutralization: A reconceptualization and empirical examination. *Journal of Research in Crime and Delinquency, 18*(2), 295-318.
Mirowski, P. (2018). The future(s) of open science. *Social Studies of Science, 48*(2), 171-203. doi: 10.1177/0306312718772086
Mishkin, B. (1988). Responding to scientific misconduct due process and prevention. *Jama, 260*(13), 1932-1936.
Mody, C. C. M. (2001). A Little Dirt Never Hurt Anyone. *Social Studies of Science, 31*(1), 7-36. doi: doi:10.1177/030631201031001002
Mongeon, P., Waltman, L., & Rijcke, S. d. (2016). What do we know about journal citation cartels? A call for information. Retrieved from https://www.cwts.nl/blog?article=n-q2w2b4
Monin, P., Noorderhaven, N., Vaara, E., & Kroon, D. (2013). Giving sense to and making sense of justice in postmerger integration. *Academy of management Journal, 56*(1), 256-284.
Montgomery, K., & Oliver, A. L. (2009). Shifts in Guidelines for Ethical Scientific Conduct: How Public and Private Organizations Create and Change Norms of Research Integrity1. *Social Studies of Science, 39*(1), 137-155.
Montgomery, K., & Oliver, A. L. (2017). Conceptualizing Fraudulent Studies as Viruses: New Models for Handling Retractions. *Minerva, 55*(1), 49-64. doi: 10.1007/s11024-016-9311-z
Moskovitz, C. (2016). Self-Plagiarism, Text Recycling, and Science Education. *Bioscience, 66*(1), 5-6. doi: 10.1093/biosci/biv160
Moxham, N., & Fyfe, A. (2017). The Royal Society and the prehistory of peer review, 1665–1965. *The Historical Journal*, 1-27. doi: 10.1017/S0018246X17000334
Müller, R., & de Rijcke, S. (2017). Exploring the epistemic impacts of academic performance indicators in the life sciences. *Research Evaluation, 26*(3), 157-168. doi: 10.1093/reseval/rvx023

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., . . . Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour, 1*, 0021. doi: 10.1038/s41562-016-0021

Muraki, J., Addonizio, J. C., Choudhury, M. S., Fischer, J., Eshghi, M., Davidian, M. M., . . . Chiao, J. W. (1990). Establishment of New Human Prostatic-Cancer Cell-Line (JCA-1). *Urology, 36*(1), 79-84. doi: 10.1016/0090-4295(90)80319-i

Murphy, P. R., & Dacin, M. T. (2011). Psychological pathways to fraud: Understanding and preventing fraud in organizations. *Journal of business ethics, 101*(4), 601-618.

Nardone, R. M. (2007). An Open Letter Regarding the Misidentification and Cross-Contamination of Cell Lines: Significance and Recommendations for Correction. from http://www.cellbankaustralia.com/resources/Nardone-Open-Letter.pdf

Nasjonal utvalg for granskning av redelighet i forskning. (2015). Uttalelse i sak om mulig vitenskapelig uredelighet ved NLA Høgskolen AS.

National Institutes of Health. (2015). Enhancing Reproducibility through Rigor and Transparency, NOT-OD-15-103. Bethesda.

National Institutes of Health. (2016). Implementing Rigor and Transparency in NIH & AHRQ Research Grant Applications, NOT-OD-16-011. Bethesda.

Near, J. P., & Miceli, M. P. (1985). Organizational dissidence - The case of whistle-blowing. *Journal of business ethics, 4*(1), 1-16. doi: 10.1007/bf00382668

Near, J. P., & Miceli, M. P. (2005). Standing Up or Standing By: What Predicts Blowing the Whistle on Organizational Wrongdoing? *Research in Personnel and Human Resources Management* (pp. 95-136).

Near, J. P., & Miceli, M. P. (2016). After the wrongdoing: What managers should know about whistleblowing. *Business Horizons, 59*(1), 105-114.

Necker, S. (2014). Scientific misbehavior in economics. *Research Policy, 43*(10), 1747-1759. doi: 10.1016/j.respol.2014.05.002

Nelson-Rees, W. A., Daniels, D., & Flandermeyer, R. R. (1981). Cross-contamination of cells in culture. *Science, 212*(4493), 446-452.

Nelson-Rees, W. A., & Flandermeyer, R. R. (1976). HeLa cultures defined. *Science, 191*(4222), 96-98.

Nelson-Rees, W. A., Flandermeyer, R. R., & Hawthorne, P. K. (1974). Banded marker chromosomes as indicators of intraspecies cellular contamination. *Science, 184*(4141), 1093-1096.

Nguyen, N. T. (2012). Can open access journals help uphold the value of research integrity? *Jornal of Business and Financial Affairs, 1*(1).

Nguyen, V. M., Haddaway, N. R., Gutowsky, L. F. G., Wilson, A. D. M., Gallagher, A. J., Donaldson, M. R., . . . Cooke, S. J. (2015). How Long Is Too Long in Contemporary Peer Review? Perspectives from Authors Publishing in Conservation Biology Journals. *Plos One, 10*(8), 20. doi: 10.1371/journal.pone.0132557

Nijkamp, P. (2015). 'Zelfplagiaat'-rapport van de Commissie-Zwemmer: Broddelwerk over fictieve peanuts. In Advalvas (Ed.). Amsterdam.

Normile, D. (2014, Jul. 2, 2014). Nature retracts controversial stem cell papers. from http://www.sciencemag.org/news/2014/07/nature-retracts-controversial-stem-cell-papers

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., . . . Yarkoni, T. (2015). Promoting an open research culture. *Science, 348*(6242), 1422-1425. doi: 10.1126/science.aab2374

Nosek, B. A., & Lakens, D. (2014). Registered reports: A method to increase the credibility of published results. *Social Psychology, 45*(3), 137-141. doi: 10.1027/1864-9335/a000192

Nuijten, M. B., Hartgerink, C. H. J., van Assen, M. A. L. M., Epskamp, S., & Wicherts, J. M. (2016). The prevalence of statistical reporting errors in psychology (1985–2013). *Behavior Research Methods, 48*(4), 1205-1226. doi: 10.3758/s13428-015-0664-2

Nylenna, M. (2007). Research misconduct: lessons to be learned? *Michael, 4*, 7-9.

Nylenna, M. (2016). Ti år etter Sudbø-saken. *Tidsskrift For Den Norske Laegeforening: Tidsskrift For Praktisk Medicin, Ny Raekke, 136*(17), 1420.

O'Brien, S. J., Shannon, J. E., & Gail, M. H. (1980). A molecular approach to the identification and individualization of human and animal cells in culture: isozyme and allozyme genetic signatures. *In Vitro Cellular & Developmental Biology-Plant, 16*(2), 119-135.

O'Hair, R. A. J., & Neff, J. (2013). Is There "Self Plagiarism" and "Salami" Publishing in JASMS? *Journal of the American Society for Mass Spectrometry, 24*(7), 986-987. doi: 10.1007/s13361-013-0674-1

O'Toole, C. M., Povey, S., Hepburn, P., & Franks, L. (1983). Identity of some human bladder cancer cell lines.

Office of Research Integrity. (1993). Definition of Research Misconduct. from https://ori.hhs.gov/definition-misconduct

Okike, K., Hug, K. T., Kocher, M. S., & Leopold, S. S. (2016). Single-blind vs double-blind peer review in the setting of author prestige. *Jama, 316*(12), 1315-1316. doi: 10.1001/jama.2016.11014

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), aac4716.

Oransky, I. (2017). Stuck in limbo: What happens to papers flagged by journals as potentially problematic? *Retraction Watch.* Retrieved 28-02-2017, from http://retractionwatch.com/2017/02/28/stuck-limbo-happens-papers-flagged-journals-potentially-problematic/

Oudshoorn, N. E. J., & Pinch, T. (2007). User-technology relationships: Some Recent Developments. In E. J. Hackett, O. Amsterdamska, M. Lycnh, & J. Wajcman (Eds.), *Handbook for Social Studies of Science* (3 ed., pp. 541 - 567). London: MIT Press.

Overman, S., Akkerman, A., & Torenvlied, R. (2016). Targets for honesty: How performance indicators shape integrity in Dutch Higher Education. *Public Administration, 94*(4), 1140-1154. doi: 10.1111/padm.12282

Paine, L. S. (1994). Managing for organizational integrity. *Harvard business review, 72*(2), 106-117.

Palazzo, G. (2007a). Organizational integrity—understanding the dimensions of ethical and unethical behavior in corporations. *Corporate ethics and corporate governance*, 113-128.

Palazzo, G. (2007b). Organizational Integrity — Understanding the Dimensions of Ethical and Unethical Behavior in Corporations. In W. C. Zimmerli, M. Holzinger, & K. Richter (Eds.),

Corporate ethics and corporate governance (pp. 113-128). Berlin, Heidelberg: Springer Berlin Heidelberg.
Palmer, D. (2012). *Normal organizational wrongdoing: A critical analysis of theories of misconduct in and by organizations*. Oxford: Oxford University Press.
Park, H., & Lewis, D. (2018). The negative health effects of external whistleblowing: A study of some key factors. *The Social Science Journal*. doi: https://doi.org/10.1016/j.soscij.2018.04.002
Patel, J. (2014). Why training and specialization is needed for peer review: a case study of peer review for randomized controlled trials. *Bmc Medicine, 12*, 7. doi: 10.1186/s12916-014-0128-z
Pattinson, D., & Prater, C. (2017). *Assessment of the Prevalence of Integrity Issues in Submitted Manuscripts*. Paper presented at the Eigth International Congress on Peer Review and Scientific Publication, Chicago.
Peerage of Science. (2017). A Free service for scientific peer review and publishing. from https://www.peerageofscience.org/
Penders, B. (2017). Beyond Trust: Plagiarism and Truth. *Journal of Bioethical Inquiry*. doi: 10.1007/s11673-017-9825-6
Penders, B., Vos, R., & Horstman, K. (2009). A question of style: method, integrity and the meaning of proper science. *Endeavour, 33*(3), 93-98. doi: 10.1016/j.endeavour.2009.07.001
Peres, R., Muller, E., & Mahajan, V. (2010). Innovation diffusion and new product growth models: A critical review and research directions. *International Journal of Research in Marketing, 27*(2), 91-106. doi: 10.1016/j.ijresmar.2009.12.012
Perkel, J. M. (2011). Curing Cell Lines. *Biotechniques, 51*(2), 85-90.
Peters, D. P., & Ceci, S. J. (1982). Peer-review practices of psychological journals: The fate of published articles, submitted again. *Behavioral and Brain Sciences, 5*(2), 187-195. doi: 10.1017/S0140525X00011183
Phlippen, S. (2013). Economentop 40. In ESB (Ed.), (Vol. 4674). Amsterdam: ESB.
PLOS. (2018). Guidelines for Reviewers. 2018, from http://journals.plos.org/plosone/s/reviewer-guidelines
Pontille, D., & Torny, D. (2014). The Blind Shall See! The Question of Anonymity in Journal Peer Review. *Ada: A Journal of Gender, New Media, and Technology, 4*. doi: doi:10.7264/N3542KVW
Pontille, D., & Torny, D. (2015). From Manuscript Evaluation to Article Valuation: The Changing Technologies of Journal Peer Review. *Human Studies, 38*(1), 57-79. doi: 10.1007/s10746-014-9335-z
Pöschl, U. (2012). Multi-stage open peer review: scientific evaluation integrating the strengths of traditional peer review with the virtues of transparency and self-regulation. *Frontiers in computational neuroscience, 6*.
Prechelt, L., Graziotin, D., & Fernández, D. M. (2017). On the status and future of peer review in software engineering. *arXiv*.
PRINTEGER. (2016). Documents and Results.   Retrieved 20-04-2016, 2016, from https://printeger.eu/documents-results/
PubPeer Foundation. (2017). PubPeer. from https://pubpeer.com/

Puto, L. A., Brognard, J., & Hunter, T. (2015). Transcriptional Repressor DAXX Promotes Prostate Cancer Tumorigenicity via Suppression of Autophagy. *Journal of Biological Chemistry, 290*(25), 15406-15420. doi: 10.1074/jbc.M115.658765

Reckwitz, A. (2002). Toward a Theory of Social Practices:A Development in Culturalist Theorizing. *European Journal of Social Theory, 5*(2), 243-263. doi: 10.1177/13684310222225432

Remmie, M. (2014). Vier voorbeelden van het (zelf)plagiaat van topeconoom Nijkamp. *NRC Handelsblad*.

Rennie, D. (2003). Editorial peer review: its development and rationale. *Peer review in health sciences, 2*, 1-13.

Rennie, D. (2016). Make peer review scientific. *Nature, 535*(7610), 31-33.

Renzetti, C. M., & Lee, R. M. (1993). *Researching Sensitive Topics*. Newbury Park, California: Sage.

Research Square. (2017a). Editorial Checks & Badges. from https://www.researchsquare.com/publishers/badges

Research Square. (2017b). Rubriq. from https://www.rubriq.com/

Resnick, B. (2019). The costs of academic publishing are absurd. The University of California is fighting back. *Vox.* from https://www.vox.com/science-and-health/2019/3/1/18245235/university-of-california-elsevier-subscription-open-access

Retraction Watch. (2014, 02-07-2014). Retractions arrive in plagiarism scandal involving economist Nijkamp. from http://retractionwatch.com/2014/07/02/retractions-arrive-in-plagiarism-scandal-involving-economist-nijkamp/

Retraction Watch. (2018). The Retraction Watch Retraction Database. from http://retractiondatabase.org/RetractionSearch.aspx?

Rinkinen, J., Shove, E., & Smits, M. (2017). Cold chains in Hanoi and Bangkok: Changing systems of provision and practice. *Journal of Consumer Culture, 0*(0), 1469540517717783. doi: 10.1177/1469540517717783

Robin, E. D., & Burke, C. M. (1987). Peer review in medical journals. *CHEST Journal, 91*(2), 252-257.

Roig, M. (2010). Plagiarism and self-plagiarism: What every author should know. *Biochemia Medica, 20*(3), 295-300.

Rojas, F. (2007). A modest proposal: triple blind review. *Orgtheory.net.* Retrieved January 23, 2007, from https://orgtheory.wordpress.com/2007/01/23/a-modest-proposal-triple-blind-review/

Ross-Hellauer, T. (2017). *What is open peer review? A systematic review [version 2; referees: 1 approved, 3 approved with reservations]* (Vol. 6).

Ross, J. S., Gross, C. P., Desai, M. M., & et al. (2006). Effect of blinded peer review on abstract acceptance. *Jama, 295*(14), 1675-1680. doi: 10.1001/jama.295.14.1675

Rouleau, L. (2005). Micro-practices of strategic sensemaking and sensegiving: How middle managers interpret and sell change every day. *Journal of Management studies, 42*(7), 1413-1441.

Roumbanis, L. (2018). Symbolic Violence in Academic Life: A Study on How Junior Scholars are Educated in the Art of Getting Funded. *Minerva*. doi: 10.1007/s11024-018-9364-2

Russel, R. (2008). The business of academic publishing: A strategic analysis of the academic journal publishing industry and its impact on the future of scholarly publishing. *Electron J Acad Spec Librarianship. http://southernlibrarianship. icaap. org/content*.
Sacco, D. F., Bruton, S. V., & Brown, M. (2018). In Defense of the Questionable: Defining the Basis of Research Scientists' Engagement in Questionable Research Practices. *Journal of Empirical Research on Human Research Ethics, 13*(1), 101-110. doi: 10.1177/1556264617743834
Sævik, T. H. (2015a, 21 November 2015). Arbeidstilsynet til kristen høgskole etter plagiatstrid. *Dagen*.
Sævik, T. H. (2015b, 19 November 2015). Plagiat-strid på kristen høgskole. *Dagen*.
Sage Open. (2018). More information: Description. 2018, from https://uk.sagepub.com/en-gb/eur/journal/sage-open#description
Samuelson, P. (1994). SELF-PLAGIARISM OR FAIR-USE. *Communications of the Acm, 37*(8), 21-25.
Santoro, D., & Kumar, M. (2018). *Speaking Truth to Power: A Theory of Whistleblowing* (Vol. 6): Springer.
Sarewitz, D. (2016). The pressure to publish pushes down quality. *Nature, 533*(7602), 147-147.
Scanlon, P. M. (2007). Song From Myself: An Anatomy of Self-Plagiarism. *MPublishing, University of Michigan Library, Ann Arbor, 2*.
Schatzki, T. R. (1996). *Social practices: A Wittgensteinian approach to human activity and the social*: Cambridge University Press.
Schatzki, T. R. (2002). *The Site of the Social: A Philosophical Account of the Constitution of Social Life and Change*: Penn State University Press.
Schekman, R., Watt, F., & Weigel, D. (2013). The eLife approach to peer review. *Elife, 2*. doi: 10.7554/eLife.00799
Scheman, R., & Bennett, C. N. (2017). *Assessing the Outcomes of Introducing a Digital Image Quality Control Review Into the Publication Process for Research Articles in Physiology Journals.* Paper presented at the International Congress on Peer Review and Scientific Publication, Chicago.
Schmidt, M. (2018). An Analysis of the Validity of Retraction Annotation in PubMed and the Web of Science. *Journal of the Association for Information Science and Technology, 69*(2), 318-328. doi: 10.1002/asi.23913
Schneider, J. W., Aagaard, K., & Bloch, C. W. (2016). What happens when national research funding is linked to differentiated publication counts? A comparison of the Australian and Norwegian publication-based funding models. *Research Evaluation, 25*(3), 244-256. doi: 10.1093/reseval/rvv036
Schor, S., & Karten, I. (1966). Statistical evaluation of medical journal manuscripts. *Jama, 195*(13), 1123-1128.
Schulz, J. B., Cookson, M. R., & Hausmann, L. (2016). The impact of fraudulent and irreproducible data to the translational research crisis - solutions and implementation. *Journal of Neurochemistry, 139*, 253-270. doi: 10.1111/jnc.13844
Shapin, S. (2008). *The Scientific life: a moral history of a late modern vocation*. Chicago: The University of Chicago Press.

Shen, C., & Björk, B.-C. (2015). ‘Predatory’open access: a longitudinal study of article volumes and market characteristics. *Bmc Medicine, 13*(1), 230.

Shimshoni, D. (1970). The Mobile Scientist in the American Instrument Industry. *Minerva, 8*(1), 59-89.

Shove, E., Pantzar, M., & Watson, M. (2012). The dynamics of social practice : everyday life and how it changes. Los Angeles: SAGE. Retrieved from http://site.ebrary.com/id/10568302.

Sieber, J. E. (1993). The Ethics and Politics of Sensitive Research. In C. M. Renzetti & R. M. Lee (Eds.), *Researching Sensitive Topics* (pp. 14-26). Newbury Park, California: Sage.

Sieber, J. E., & Stanley, B. (1988). Ethical and professional dimensions of socially sensitive research. *American Psychologist, 43*(1), 49.

Sigl, L. (2016). On the Tacit Governance of Research by Uncertainty:How Early Stage Researchers Contribute to the Governance of Life Science Research. *Science, Technology, & Human Values, 41*(3), 347-374. doi: 10.1177/0162243915599069

SIGMA. (2005). SIGMA is arXiv overlay. from http://www.emis.de/journals/SIGMA/about.html#overlay

Singh Chawla, D. (2015). Researchers argue for standard format to cite lab resources. *Nature, 29*, 2015. doi: doi:10.1038/nature.2015.17652.

Sismondo, S. (2010). *An Introduction to Science and Technology Studies* (Second Edition ed.): Wiley - Blackwell.

Sismondo, S. (2017). Post-truth? *Social Studies of Science, 47*(1), 3-6. doi: 10.1177/0306312717692076

Skogan, W. G. (1977). Dimensions of Dark Figure of Unreported Crime. *Crime & Delinquency, 23*(1), 41-50. doi: 10.1177/001112877702300104

Smith, A., & Raven, R. (2012). What is protective space? Reconsidering niches in transitions to sustainability. *Research Policy, 41*(6), 1025-1036. doi: https://doi.org/10.1016/j.respol.2011.12.012

Smith, R. (1999). Opening up BMJ peer review: A beginning that should lead to complete transparency. *British Medical Journal, 318*(7175), 4-5. doi: 10.1136/bmj.318.7175.4

Smith, R. (2006). Peer review: a flawed process at the heart of science and journals. *Journal of the Royal Society of Medicine, 99*(4), 178-182.

Smith, R. (2010). Classical peer review: an empty gun. *Breast Cancer Research, 12*, 4. doi: 10.1186/bcr2742

Solomon, D. (2002). Talking past each other: making sense of the debate over electronic publication. *First monday, 7*(8).

Stacey, G. N. (2000). Cell contamination leads to inaccurate data: we must take action now. *Nature, 403*(6768), 356-356. doi: 10.1038/35000394

Stapel, D. (2012). *Ontsporing*. Amsterdam: Prometheus.

Star, S. L. (1999). The Ethnography of Infrastructure. *American Behavioral Scientist, 43*(3), 377-391. doi: 10.1177/00027649921955326

Steen, R. G. (2011). Retractions in the medical literature: how many patients are put at risk by flawed research? *Journal of Medical Ethics, 37*(11), 688-692. doi: 10.1136/jme.2011.043133

Steneck, N. H. (1994). Research universities and scientific misconduct: History, policies, and the future. *The Journal of Higher Education, 65*(3), 310-330.

Steneck, N. H. (2006). Fostering integrity in research: Definitions, current knowledge, and future directions. *Science and Engineering Ethics, 12*(1), 53-74. doi: 10.1007/pl00022268

Stephan, P. E. (2012). *How economics shapes science* (Vol. 1): Harvard University Press Cambridge, MA.

Stern, B. M., & O'Shea, E. K. (2019). A proposal for the future of scientific publishing in the life sciences. *PLOS Biology, 17*(2), e3000116. doi: 10.1371/journal.pbio.3000116

Stern, V. (2017). Publisher issues first retractions for fake peer review, starts new checking policy. Retrieved from http://retractionwatch.com/2017/11/21/frontiers-issues-first-retractions-fake-peer-review-starts-new-checking-policy/

Stewart, W. W., & Feder, N. (1987). THE INTEGRITY OF THE SCIENTIFIC LITERATURE. *Nature, 325*(6101), 207-214. doi: 10.1038/325207a0

Stitzel, B., Hoover, G. A., & Clark, W. (2018). More on Plagiarism in the Social Sciences*. *Social Science Quarterly*, n/a-n/a. doi: 10.1111/ssqu.12481

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research*: Sage publications.

Stroebe, W., Postmes, T., & Spears, R. (2012). Scientific Misconduct and the Myth of Self-Correction in Science. *Perspectives on Psychological Science, 7*(6), 670-688. doi: 10.1177/1745691612460687

Struiksma, J., van Mil, J., & Widdershoven, J. (2015). Adviesrapport CWI, Casus 2014-04 FEWEB, NN3 versus Nijkamp, Kourtit, Baycan. In C. W. I. VU/VUmc (Ed.). Amsterdam: Vrije Universiteit.

Sudbø, J., Kildal, W., Risberg, B., Koppang, H. S., Danielsen, H. E., & Reith, A. (2001). DNA content as a prognostic marker in patients with oral leukoplakia. *New England Journal of Medicine, 344*(17), 1270-1278.

Sudbø, J., Lee, J., Lippman, S., Mork, J., Sagen, S., Flatner, N., . . . Zhou, X. (2005). RETRACTED: Non-steroidal anti-inflammatory drugs and the risk of oral cancer: a nested case-control study. *The Lancet, 366*(9494), 1359–1366.

Sudbø, J., Lippman, S. M., Lee, J. J., Mao, L., Kildal, W., Sudbø, A., . . . Risberg, B. (2004). The influence of resection and aneuploidy on mortality in oral leukoplakia. *New England Journal of Medicine, 350*(14), 1405-1413.

Sudbø, J., Samuelsson, R., Risberg, B. r., Heistein, S., Nyhus, C., Samuelsson, M., . . . Reith, A. (2005). Risk markers of oral cancer in clinically normal mucosa as an aid in smoking cessation counseling. *Journal of clinical oncology, 23*(9), 1927-1933.

Südhof, T. C. (2016). Truth in Science Publishing: A Personal Perspective. *PLOS Biology, 14*(8), e1002547. doi: 10.1371/journal.pbio.1002547

Sun, Y. C. (2013). Do journal authors plagiarize? Using plagiarism detection software to uncover matching text across disciplines. *Journal of English for Academic Purposes, 12*(4), 264-272. doi: 10.1016/j.jeap.2013.07.002

Sykes, G. M., & Matza, D. (1957). Techniques of neutralization: A theory of delinquency. *American Sociological Review, 22*(6), 664-670.

Tavare, A. (2011). Managing research misconduct: is anyone getting it right? *BMJ : British Medical Journal, 343*. doi: 10.1136/bmj.d8212

Taylor & Francis. (2015). "Peer Review in 2015: A global view", A white paper by Taylor & Francis. from https://authorservices.taylorandfrancis.com/peer-review-global-view

Tennant, J. P., Dugan, J. M., Graziotin, D., Jacques, D. C., Waldner, F., Mietchen, D., . . . Crick, T. (2017). A multi-disciplinary perspective on emergent and future innovations in peer review. *F1000Research, 6*(1151).

Tennant, J. P., Ross-Hellauer, T., Matthias, L., Byers, A., & Dugan, J. M. (2018). *Thinking Outside the Black Box of Peer Review*. Paper presented at the International Conference on Peer Review, Rome. http://www.peere.org/wp-content/uploads/2018/03/Tennant.pdf

Thelwall, M., Haustein, S., Lariviere, V., & Sugimoto, C. R. (2013). Do Altmetrics Work? Twitter and Ten Other Social Web Services. *Plos One, 8*(5). doi: 10.1371/journal.pone.0064841

Thomas, J. J. (1993). Measuring the Underground Economy: A Suitable Case for Interdisciplinary Treatment? In C. M. Renzetti & R. M. Lee (Eds.), *Researching Sensitive Topics* (pp. 52-70). Newbury Park, California: Sage.

Thomson Reuters. (2016). Web of Knowledge. from https://apps.webofknowledge.com/

Thurman, A. (2001). Institutional responses to medical mistakes: ethical and legal perspectives. *Kennedy Institute of Ethics Journal, 11*(2), 147-156.

Thurman, R. H., Chervenak, F. A., McCullough, L. B., Halwani, S., & Farine, D. (2016). Self-plagiarism: a misnomer. *American Journal of Obstetrics and Gynecology, 214*(1), 91-93. doi: http://dx.doi.org/10.1016/j.ajog.2015.09.004

Tijdink, J. K., Bouter, L. M., Veldkamp, C. L. S., van de Ven, P. M., Wicherts, J. M., & Smulders, Y. M. (2016). Personality Traits Are Associated with Research Misbehavior in Dutch Scientists: A Cross-Sectional Study. *Plos One, 11*(9), e0163251. doi: 10.1371/journal.pone.0163251

Tramer, M. R., Reynolds, D. J. M., Moore, R. A., & McQuay, H. J. (1997). Impact of covert duplicate publication on meta-analysis: a case study. *British Medical Journal, 315*(7109), 635-640.

Treviño, L. J., Balkin, D. B., & Gomez-Mejia, L. R. (2017). How “doing gender” leads to gender imbalances in the higher ranks in colleges of business [and how to “undo gender”]. *Academy of Management Learning & Education, 16*(3), 439-453.

Trinkle, B. S., Phillips, T., Hall, A., & Moffatt, B. (2017). Neutralising fair credit: factors that influence unethical authorship practices. *Journal of Medical Ethics, 43*(6), 368-373. doi: 10.1136/medethics-2015-103365

Turnitin. (2006). from http://www.turnitin.com/en_us/

van Assen, M. A. L. M., van Aert, R. C. M., Nuijten, M. B., & Wicherts, J. M. (2014). Why Publishing Everything Is More Effective than Selective Publishing of Statistically Significant Results. *Plos One, 9*(1), e84896. doi: 10.1371/journal.pone.0084896

van Bokhoven, A., Varella-Garcia, M., Korch, C., Hessels, D., & Miller, G. J. (2001). Widely used prostate carcinoma cell lines share common origins. *Prostate, 47*(1), 36-51. doi: 10.1002/pros.1045

van Bokhoven, A., Varella-Garcia, M., Korch, C., & Miller, G. J. (2001). TSU-Pr1 and JCA-1 cells are derivatives of T24 bladder carcinoma cells and are not of prostatic origin. *Cancer Research, 61*(17), 6340-6344.

Van Buggenhout, M., & Christiaens, J. (2016). Deviance in science: a criminological analysis - PRINTEGER deliverable D II.5. from http://printeger.eu/wp-content/uploads/2016/12/D2.5.pdf van der Heyden, M. A. G., van de Ven, T. D., & Opthof, T. (2009). Fraud and misconduct in science: the stem cell seduction. *Netherlands Heart Journal, 17*(1), 25-29.

Van Leeuwen, T., Reyes Elizondo, A., & De Rijcke, S. (2017). Analysis of papers retracted in the Web of Science. In PRINTEGER (Ed.). Leiden.

Van Lente, H. (1993). *Promising technology. The dynamics of expectations in technological developments.* (PhD Thesis), Twente University, Delft.

Van Noorden, R. (2011). The trouble with retractions. *Nature, 478*(7367), 26-28.

van Vuuren, M. (2012). Merger, She Wrote: Improvising on the Script of Identification Processes During Mergers *Advances in Mergers and Acquisitions* (pp. 39-53): Emerald Group Publishing Limited.

Vandekerckhove, W. (2016). *Whistleblowing and organizational social responsibility: A global assessment*: Routledge.

Varella-Garcia, M., Boomer, T., & Miller, G. J. (2001). Karyotypic similarity identified by multiplex-FISH relates four prostate adenocarcinoma cell lines: PC-3, PPC-1, ALVA-31, and ALVA-41. *Genes Chromosomes & Cancer, 31*(4), 303-315. doi: 10.1002/gcc.1149

Vaughan, D. (1998). Rational choice, situated action, and the social control of organizations. *Law and Society Review*, 23-61.

Vaughan, D. (1999). The dark side of organizations: Mistake, misconduct, and disaster. *Annual review of sociology, 25*(1), 271-305.

Vaughan, D. (2002). Criminology and the sociology of organizations. *Crime, law and social change, 37*(2), 117-136.

Vaughan, L., Glanzel, W., Korch, C., & Capes-Davis, A. (2017). Widespread use of misidentified cell line KB (HeLa): Incorrect attribution and its impact revealed through mining the scientific literature. *Cancer Research*. doi: 10.1158/0008-5472.CAN-16-2258

Veldkamp, C. L. S., Hartgerink, C. H. J., van Assen, M. A. L. M., & Wicherts, J. M. (2017). Who Believes in the Storybook Image of the Scientist? *Accountability in Research, 24*(3), 127-151. doi: 10.1080/08989621.2016.1268922

Verbon, H. (2016). Peter Nijkamp (VU): verzon feitenvrij migratiebeleid. Retrieved from http://harrieverbon.blogspot.nl/2016/01/peter-nijkamp-vu-verzon-feitenvrij.html

Verbong, G., Geels, F. W., & Raven, R. (2008). Multi-niche analysis of dynamics and policies in Dutch renewable energy innovation journeys (1970–2006): hype-cycles, closed networks and technology-focused learning. *Technology Analysis & Strategic Management, 20*(5), 555-573. doi: 10.1080/09537320802292719

von Hippel, E. (1976). The dominant role of users in the scientific instrument innovation process. *Research Policy, 5*(3), 212-239. doi: https://doi.org/10.1016/0048-7333(76)90028-7

VSNU. (2012). De Nederlandse Gedragscode Wetenschapsbeoefening: Principes van goed wetenschappelijk onderwijs en onderzoek.

VSNU. (2013). Beschuldiging van plagiaat. Amsterdam: VU.

VSNU. (2014). The Netherlands Code of Conduct for Academic Practice, revised version. The Hague.

VSNU. (2015). VU Plagiaat - ongegrond 2: VU.

VU. (2017). Department of Spatial Economics. from http://www.feweb.vu.nl/nl/afdelingen-en-instituten/spatial-economics/index.aspx

Wagner, P. D., & Bates, J. H. T. (2016). Maintaining the integrity of peer review. *Journal of Applied Physiology, 120*(5), 479-480. doi: 10.1152/japplphysiol.00067.2016

Walker, R., & Rocha da Silva, P. (2015). Emerging trends in peer review—a survey. *Frontiers in Neuroscience, 9*(169). doi: 10.3389/fnins.2015.00169

Waltman, L., & van Eck, N. J. (2012). A new methodology for constructing a publication-level classification system of science. *Journal of the American Society for Information Science and Technology, 63*(12), 2378-2392. doi: 10.1002/asi.22748

Ware, M. (2008). Peer review: benefits, perceptions and alternatives. *Publishing Research Consortium, 4*, 1-20.

Weick, K. E. (1995). *Sensemaking in organizations* (Vol. 3): Sage.

Weick, K. E., Sutcliffe, K. M., & Obstfeld, D. (2005). Organizing and the process of sensemaking. *Organization science, 16*(4), 409-421.

Wejnert, B. (2002). Integrating models of diffusion of innovations: A conceptual framework. *Annual review of sociology, 28*, 297-326. doi: 10.1146/annurev.soc.28.110601.141051

Weller, A. C. (2001). *Editorial peer review: Its strengths and weaknesses*: Information Today, Inc.

Westlund, H., Martin, J. C., & Stough, R. R. (2014). Letter of support for Professor Peter Nijkamp.

Whelehan, I. (1995). *Modern Feminist Thought: From the Second Wave to\Post-Feminism*: NYU Press.

Wicherts, J. M. (2017). The Weak Spots in Contemporary Science (and How to Fix Them). *Animals (Basel), 7*(12). doi: 10.3390/ani7120090

Williams, R. H., & Benford, R. D. (2000). Two faces of collective action frames: A theoretical consideration. *Current perspectives in social theory, 20*, 127-152.

Wilsdon, J. (2018, 14-12-2018). *The role of researchers in valueing science.* Paper presented at the On Quality - Symposium on the occasion of Paul Wouters' farewell as director of CWTS, Leiden.

Wisdom, J. P., Chor, K. H. B., Hoagwood, K. E., & Horwitz, S. M. (2014). Innovation Adoption: A Review of Theories and Constructs. *Administration and policy in mental health, 41*(4), 480-502. doi: 10.1007/s10488-013-0486-4

Wouters, P. (1999). *The Citation Culture.* (Doctoral Doctoral Thesis), University of Amsterdam, Amsterdam. Retrieved from https://pure.uva.nl/ws/files/3164289/8218_thesis.pdf

Wouters, P., Beaulieu, A., Scharnhorst, A., & Wyatt, S. (2013). *Virtual knowledge: experimenting in the humanities and the social sciences*: MIT Press.

Yang, W. (2013). Research Integrity in China. *Science, 342*(6162), 1019-1019. doi: 10.1126/science.1247700

Yankauer, A. (1990). Editor's report: scientific misconduct and the responsibility of journal editors. *American journal of public health, 80*(4), 399-400.

Ye, F., Chen, C. G., Qin, J., Liu, J., & Zheng, C. Y. (2015). Genetic profiling reveals an alarming rate of cross-contamination among human cell lines used in China. *Faseb Journal, 29*(10), 4268-4272. doi: 10.1096/fj.14-266718

Yin, R. K. (2013). *Case study research: Design and methods*: Sage publications.

Yu, M., Selvaraj, S. K., Liang-Chu, M. M. Y., Aghajani, S., Busse, M., Yuan, J., . . . Neve, R. M. (2015). A resource for cell line authentication, annotation and quality control. *Nature, 520*(7547), 307-+. doi: 10.1038/nature14397

Zhang, H. Y. (2010). CrossCheck: an effective tool for detecting plagiarism. *Learned publishing, 23*(1), 9-14.

Zhang, M. H., & Grieneisen, M. L. (2013). The impact of misconduct on the published medical and non-medical literature, and the news media. *Scientometrics, 96*(2), 573-587. doi: 10.1007/s11192-012-0920-5

ZonMw. (2016). Bevorderen Verantwoorde Onderzoekspraktijken - Over dit programma. from https://www.zonmw.nl/nl/onderzoek-resultaten/fundamenteel-onderzoek/programmas/programma-detail/bevorderen-van-verantwoorde-onderzoekspraktijken/

Zuckerman, H. (1977). Deviant behavior and social control in science. *Deviance and social change, 1*, 87-138.

Zuckerman, H. (1984). Norms and deviant behavior in science. *Science, Technology, & Human Values, 9*(1), 7-13.

Zuckerman, H., & Merton, R. K. (1971). Patterns of evaluation in science: Institutionalisation, structure and functions of the referee system. *Minerva, 9*(1), 66-100. doi: 10.1007/bf01553188

Zwart, H., & ter Meulen, R. (2019). Addressing research integrity challenges: from penalising individual perpetrators to fostering research ecosystem quality care. *Life Sciences, Society and Policy, 15*(1), 5. doi: 10.1186/s40504-019-0093-6

Zwemmer, J., Gunning, J. W., & Grobbee, R. (2015). Rapport over verwijzingen in het werk van prof. dr. P. Nijkamp. Amsterdam: Vrije Universiteit.

Summary

Research integrity is one of the main pillars constituting trustworthy science. However, fuelled by large, often spectacular cases of research misconduct, research integrity has come under the attention of scholars, science policymakers and the larger public alike. Combined with increasing concerns about the (ir)reproducibility of scientific experiments and an apparent decline in public trust in science, some consider the global research enterprise to be in a status of crisis. This dissertation aims to study research integrity, by describing how it may be threatened and how it may be protected.

In three empirical parts, I consider three questions regarding research integrity and the scientific publication system. With these, I aim to contribute to an understanding about: How problematic research gets *spilled* into the research literature, how self-regulating mechanisms may *filter* such articles from the literature, and how organisations may *clean* up cases that slipped through.

Part I – *To Spill* directs its attention to instances of problematic research. Using the cases of academic text recycling and the usage of misidentified cell lines in biomedical research, I discuss why such research might, or might not, be considered problematic, how conventions differ between different research disciplines and contexts, and how instances of problematic research spread or spill through the academic literature. This part finds considerable levels of both types of problematic research practices in the academic literature. More specifically, it demonstrates that the occurrence of these practices varies between research disciplines and geographical locations. Moreover, it finds indications of other factors influencing the amount of problematic research ending up in the academic literature, such as the number of co-authors of a paper and the productivity of these authors.

Understanding how problematic research spreads through the literature, the important question about how to prevent this naturally emerges. Part II – *To Filter* considers one specific way of avoiding that questionable research practices harm the scientific enterprise by studying journal's editorial process as filtering mechanism. In four chapters it describes the establishment of peer review's current models. New procedures have currently been implemented as a response to changing expectations of the system. Chapter 4 describes the rationales for developing new review formats, resulting in a systematisation of peer review procedures. Chapter 5 uses this systematisation to assess the ability of different peer review formats to filter problematic research, by correlating review format with retraction rates. It demonstrates that several review procedures are associated with significantly lower retraction rates, although some uncertainty remains because of the complexity of retractions.

Chapters 6 and 7 focus on innovation and implementation of editorial procedures. In these chapters we study the innovation process by using data from a survey among editors and ethnographic data of field visits to the editorial offices of commercial publishers. Thereby we address open questions regarding the implementation of new review procedures, the occurrence rate of various peer review procedures and their distribution over scientific disciplines and academic publishers. We demonstrate that in spite of enthusiastic innovation, the adoption of new peer review procedures is in fact slow, with the exception of text similarity scanners. We therefore continue to analyse editors' and publishers' motivations for (not) implementing new review procedures in chapter 7. In this implementation process, the publisher's strong focus on commercial interest and their internal hierarchical structure turn out to be decisive factors. Hence, to understand why journal editors pick up on innovations, we need to take into account the business considerations of publishers. Although improved integrity can be a sales argument, journals are also forced to carefully weigh the costs of innovations against revenues, with efficiency and the support of authors and reviewers proving to be crucial factors.

Acknowledging that no format of editorial peer review will be impermeable to problematic research, further systems of academic self-regulation are required. We aim to contribute to the understanding of these in part III – *To Clean*, by studying the reporting of cases allegedly involving research misconduct as well as the handling of such cases by universities. In chapter 8 we study how power relations in academia influence researchers' motivation to report research misconduct, as well as the consequences of reporting. It identifies research seniority, temporary vs. permanent work appointments, and gender as factors affecting researchers' ability and willingness to report research misconduct. Analysing four cases in which such allegations were made, chapter 9 completes this part by studying organisational responses to cases of alleged misconduct. It demonstrates that universities' responses commonly show a lack of formal guidelines and procedures to deal with misconduct cases, that they are characterised by a fear of reputational damage, are influenced by (the possibility of) media coverage, and may include corrective action that seems to aim for restoration of symbolic order. Chapter 9 concludes by arguing that this may hinder effective learning from cases of alleged misconduct.

This dissertation finds some evidence for the need to include novel perspectives in discussions on research integrity. First, it shows that integrity challenges emerge differently in diverse academic contexts. Research discipline, epistemic culture, geographical location and organisational research climate all seem to influence the way in which integrity challenges are perceived and how they can be effectively addressed. Hence, one-size-fits-all solutions are not likely to be effective. Second, while several stakeholders may highly value research integrity, other, competing considerations might equally well influence their policy and strategy. These include efforts to obtain business value in commercial publishers, as well as reputational concerns in research performing organisations and individual researchers. Hence, this wider spectrum of values and interests should be taken into account when aiming to foster responsible research. Third, hierarchies and power structures may complicate effective handling of integrity issues, as well as hinder transparency and effective learning of such issues. This consequently asks for more attention for research cultures and research climate in addressing integrity challenges, rather than to focus solely on individual researchers.

Research integrity turned out to be a highly sensitive object of research. As I discuss in chapter 10, the results of our studies commonly evoked substantial responses from multiple stakeholders in the research process. These responses highlight both the importance of this topic and the wide range of stakeholders involved in it. This dissertation hence calls for more coordinated action from all these stakeholders to effectively foster research integrity. By explicating how different actors may contribute to achieving this goal and by initiating or continuing debate about integrity challenges, I hope this dissertation may ultimately support more trustworthy science.

Samenvatting

Wetenschappelijke integriteit is een van de belangrijkste pijlers van betrouwbare wetenschap. Door grote, vaak spectaculaire gevallen van wangedrag is wetenschappelijke integriteit echter onder de aandacht gekomen van zowel wetenschappers, beleidsmakers als het grote publiek. In combinatie met de toenemende bezorgdheid over de (niet) reproduceerbaarheid van wetenschappelijke experimenten en een schijnbare afname van het publieke vertrouwen in de wetenschap, wordt het wereldwijde onderzoeksysteem door sommigen beschouwd als een systeem in crisis. Dit proefschrift heeft als doel om wetenschappelijke integriteit vanuit verschillende perspectieven te bestuderen door te kijken naar hoe integriteit wordt bedreigd en hoe we haar kunnen beschermen.

In drie empirische delen behandel ik drie vragen over de integriteit van onderzoek en het wetenschappelijke publicatiesysteem. Hiermee wil ik bijdragen aan een beter begrip over: Hoe problematisch onderzoek in de onderzoeksliteratuur terechtkomt, hoe zelfregulerende mechanismen dergelijke artikelen uit de literatuur kunnen filteren, en hoe universiteiten omgaan met zaken die niet weggefilterd zijn.

Deel I - *To Spill* richt de aandacht op gevallen van problematisch onderzoek. Aan de hand van de casussen academische tekstrecycling en het gebruik van verkeerd geïdentificeerde cellijnen in biomedisch onderzoek, bespreek ik waarom dergelijk onderzoek al dan niet als problematisch kan worden beschouwd, hoe conventies hierover verschillen tussen verschillende onderzoeksdisciplines en -contexten, en hoe problematisch onderzoek zich verspreidt in de academische literatuur. Dit eerste deel van dit proefschrift vindt aanzienlijke niveaus van beide soorten problematische onderzoekspraktijken in de academische literatuur. Meer in het bijzonder toont het aan dat het aandeel van deze praktijken varieert tussen verschillende onderzoeksdisciplines en geografische locaties. Bovendien zijn er aanwijzingen dat andere factoren van invloed zijn op de hoeveelheid problematisch onderzoek dat in de academische literatuur terechtkomt, zoals het aantal co-auteurs van een artikel en de productiviteit van deze auteurs.

Wanneer we begrijpen hoe problematisch onderzoek zich door de literatuur verspreidt, rijst vanzelf de belangrijke vraag hoe dit kan worden voorkomen. Deel II - *To Filter* beschouwt één specifieke manier om te voorkomen dat twijfelachtige onderzoekspraktijken schade toebrengen aan de wetenschap, door het redactionele proces van wetenschappelijke tijdschriften als filtermechanisme te bestuderen. Dit redactionele proces wordt tegenwoordig op veel verschillende manieren ingericht. In vier hoofdstukken beschrijft deel II de totstandkoming van huidige modellen van peer review. Veel nieuwe procedures blijken ingevoerd als antwoord op de veranderende verwachtingen van het peer review systeem. Hoofdstuk 4 beschrijft de redenen voor het ontwikkelen van nieuwe beoordelingsmodellen, wat resulteert in een systematisering van peer review vormen. Hoofdstuk 5 gebruikt deze systematisering om te bestuderen in hoeverre verschillende peer review-formaten in staat zijn om problematisch onderzoek weg te filteren, door verschillende wijzen van review te correleren aan het aantal retractions van artikelen. Het toont aan dat verschillende beoordelingsprocedures gepaard gaan met aanzienlijk lagere retractionpercentages, hoewel er enige onzekerheid blijft bestaat vanwege de complexiteit van retractions.

De hoofdstukken 6 en 7 zijn gericht op innovatie en implementatie van redactionele procedures. In deze hoofdstukken bestuderen we het innovatieproces door gebruik te maken van een enquête onder tijdschriftredacteuren en van etnografische werk tijdens veldbezoeken aan de redacties van commerciële uitgevers. Daarbij buigen we ons over open vragen over de implementatie van nieuwe reviewprocedures, het voorkomen van verschillende peer review procedures en de verdeling ervan over wetenschappelijke disciplines en academische uitgevers. We tonen aan dat, ondanks enthousiaste innovatie, de invoering van nieuwe peer review procedures in feite vrij traag verloopt, met uitzondering van plagiaatscanners. Daarom analyseren we vervolgens de motivatie van redacteuren en uitgevers om nieuwe beoordelingsprocedures al dan niet toe te passen. De sterke focus van uitgevers op commerciële belangen en de interne hiërarchische structuur van uitgevers blijken in dit implementatieproces doorslaggevende factoren te zijn. Om te begrijpen waarom tijdschriftredacteuren innovaties oppikken, moeten we dus rekening houden met de zakelijke overwegingen van uitgevers. Hoewel meer aandacht voor integriteit een verkoopargument kan zijn, zijn tijdschriften ook genoodzaakt om de kosten van innovaties zorgvuldig af te wegen tegen de baten, waarbij efficiëntie en de steun van auteurs en reviewers cruciale factoren blijken te zijn.

Omdat geen enkele vorm van redactionele peer review ondoordringbaar zal zijn voor problematisch onderzoek, zijn verdere systemen van academische zelfregulering nodig. In deel III - *To Clean* willen we een bijdrage leveren aan het begrip hiervan door meldingen van vermeende onderzoeksfouten en de behandeling van dergelijke zaken door universiteiten te bestuderen. In hoofdstuk 8 onderzoeken we hoe machtsverhoudingen in de academische wereld van invloed zijn op de motivatie van onderzoekers om mogelijk wangedrag van collega's te melden. Ook bestuderen we de gevolgen van zo'n melding. Dit hoofdstuk identificeert wetenschappelijke senioriteit, tijdelijke versus permanente contracten en gender als factoren die het vermogen en de bereidheid van onderzoekers beïnvloeden om dubieuze onderzoekspraktijken te melden. Hoofdstuk 9 zet deze analyse voort met een studie van vier gevallen waarin dergelijke beschuldigingen zijn geuit. In dit hoofdstuk bestuderen we de manier waarop universiteiten met dergelijke aanklachten omgaan. Uit de reacties van de universiteiten blijkt dat er vaak een gebrek is aan formele richtlijnen en procedures om gevallen van wangedrag te behandelen, dat behandelingen worden gekenmerkt door angst voor reputatieschade, dat zij worden beïnvloed door (de mogelijkheid van) berichtgeving in de media en dat zij corrigerende maatregelen omvatten die gericht lijken op het herstel van symbolische orde. Hoofdstuk 9 concludeert daarmee dat deze processen een doeltreffende lering uit gevallen van vermeend wangedrag in de weg kan staan.

Samenvattend vindt dit proefschrift enig bewijs voor de noodzaak om nieuwe perspectieven op te nemen in discussies over de integriteit van wetenschappelijk onderzoek. Ten eerste toont het aan dat integriteitsuitdagingen in uiteenlopende academische contexten anders naar voren komen. Onderzoeksdiscipline, epistemische cultuur, geografische en culturele achtergrond en organisatorisch onderzoeksklimaat lijken allemaal van invloed te zijn op de manier waarop integriteitsuitdagingen worden ervaren en hoe ze effectief kunnen worden aangepakt. Daarom zullen one-size-fits-all initiatieven waarschijnlijk niet bijdragen aan een algehele oplossing. Ten tweede, hoewel verschillende belanghebbenden veel waarde hechten aan wetenschappelijke integriteit, kunnen andere, concurrerende overwegingen hun beleid en strategie evenzeer beïnvloeden. Dit geldt onder meer voor commerciële belangen bij uitgevers, maar ook voor de waarde van een goede reputatie van onderzoeksinstellingen en individuele onderzoekers. Bij de bevordering van verantwoord onderzoek moet dan ook rekening worden gehouden met dit bredere spectrum van waarden en belangen. Ten derde kunnen hiërarchieën en machtsstructuren een effectieve aanpak van integriteitskwesties bemoeilijken en de transparantie en het effectief leren van dergelijke kwesties belemmeren. Dit vraagt dan ook om meer aandacht voor onderzoeksculturen en het bredere onderzoeksklimaat bij het aanpakken van integriteitsproblemen, in plaats van ons uitsluitend te richten op individuele onderzoekers.

Wetenschappelijke integriteit is een zeer gevoelig onderzoeksonderwerp gebleken. Zoals ik in hoofdstuk 10 bespreek, hebben de resultaten van onze studies vaak uitgebreide en sterke reacties opgeroepen van betrokkenen bij het onderzoeksproces. Deze reacties benadrukken zowel het belang van dit onderwerp als het brede scala aan partijen dat erbij betrokken is. Dit proefschrift roept dan ook op tot meer gezamenlijke en gecoördineerde actie tussen al deze belanghebbenden om de integriteit van onderzoek effectief te bevorderen. Door uit te leggen hoe verschillende actoren kunnen bijdragen aan het bereiken van dit doel en door het initiëren en voortzetten van het debat over integriteitsuitdagingen, hoop ik dat dit proefschrift uiteindelijk een betrouwbaardere wetenschap kan ondersteunen.

Dankwoord

Zoals met veel goede of belangrijke dingen in het leven, was ook het afronden van dit proefschrift niet mogelijk geweest zonder de hulp, steun en toewijding van veel mensen om mij heen. Ik wil dan ook allen bedanken die op welke wijze dan ook hebben geholpen bij het tot stand komen van dit werk. Op deze plek wil ik graag even de ruimte nemen om de belangrijke bijdragen van sommige mensen in het bijzonder uit te lichten.

Ten eerste, en zonder twijfel belangrijkste, wil ik Willem bedanken voor zijn onmisbare hulp en steun. De keuze om na mijn afstudeertijd nog een tijdje samen door te werken was een hele bewuste en een waar ik nog steeds ontzettend blij mee ben. Ik ben je erg dankbaar voor de manier waarop je mij de afgelopen jaren hebt geïntroduceerd in de wetenschap, hoe je me door dit (soms ruige) wereldje begeleid hebt en hoe je altijd oog had voor hoe ik me het beste kon ontwikkelen. Je was een begeleider om tegen op te kijken, iemand die me voortdurend scherp gehouden heeft, maar je bent ook de gemoedelijkheid en zachtaardigheid zelve, iemand waarbij ik me altijd op mijn gemak heb gevoeld. Al met al een waar voorbeeld. Ik gun iedere promovendus een begeleider zoals jij, al ben ik bang dat ik toch echt een van de weinigen ben die van dit voorrecht heeft mogen genieten. Ik hoop dat we onze zeer prettige en productieve samenwerking nog lang door mogen zetten.

Daarnaast wil ik Hub bedanken voor zijn begeleiding van mijn promotietraject. Jij was degene die me in aanraking heeft gebracht met het thema wetenschappelijke integriteit. Van dat eerste gesprek tussen ons, heb ik nog jarenlang plezier gehad. Daarnaast wil ik je bedanken voor je feedback in verschillende fasen, en met name de laatste, van mijn promotietraject.

Mijn dank gaat uit naar Sarah, Barend en Henk voor het lezen en beoordelen van dit proefschrift. Dank jullie wel voor de moeite die jullie hierin hebben gestoken en voor de waardevolle feedback die ik nu en op andere momenten van jullie heb mogen ontvangen.

I am very grateful to the European Horizon2020 research program and ZonMw’s BVO program for supporting the studies in this dissertation. I hope similar programs will facilitate additional studies in the future.

Ook wil ik Martin en Thed hier in het bijzonder noemen. Ik heb mij jarenlang gelukkig geprezen met jullie adviezen, hulp en steun. Zonder het ooit zo te benoemen, zijn jullie al die tijd als twee mentoren voor me geweest. Vaak belangrijk op cruciale of juist hele alledaagse momenten, met kleine duwtjes in de rug, of door juist even aan de rem te trekken. En natuurlijk hebben de gezelligheid en wandelingetjes door het park of naar de koffieautomaat het promovendusbestaan een stuk aangenamer gemaakt. Dank jullie wel hiervoor.

De collega's bij zowel ISiS als het CWTS wil ik graag bedanken voor de fijne tijd die ik op beide instituten heb gehad. Ik heb de afgelopen jaren in twee verschillende, maar warme nesten mogen doorbrengen. Dank jullie wel voor de inspiratie en gezelligheid die jullie me gebracht hebben. Een apart woordje van dank gaat uit naar het Research Quality Team (RQT, spreek uit: 'our cutie') voor alle leerzame en uitdagende discussies die we het afgelopen jaar hebben gevoerd.

I want to thank Eric and Svenn-Erik for our collaboration that led to several of this dissertation's chapters. Thank you for the inspiring conversations, for your sharp feedback and for your stimulating enthusiasm. It has been a great pleasure to work with you and I look forward to collaborating again in the future. I also want to express my gratitude to the rest of the PRINTEGER-team for the very pleasant way in which they made me familiar with international research and the genuine delight of team science.

Ook wil ik alle andere mensen waarmee ik de afgelopen jaren heb samengewerkt bedanken, in het bijzonder Wytske, Michèle, Gareth, Joeri en Nees Jan. Het is altijd heel verfrissend geweest om met jullie te werken. Zonder alle ideeën, inzichten en motivatie die dit opgeleverd heeft, zou mijn proefschrift er heel anders uit hebben gezien.

Dit dankwoord zou absoluut niet compleet zijn zonder stil te staan bij de bijdrage van WTMC en alle WTMC'ers. De activiteiten binnen de graduate school hebben me wegwijs gemaakt in STS en gevormd als wetenschapper. De discussies, lezingen en creatieve opdrachten met jullie hebben me altijd enorm geïnspireerd en onwijs veel plezier en energie gegeven. Onze verblijven in Ravenstein hoorden telkens tot de hoogtepunten van het jaar. In het bijzonder wil ik Bernike, Govert en Anne bedanken voor hun inzet om dit alles mogelijk te maken. I also want to thank Elizabeth Shove for the amazing WTMC Summer School in 2018 and for hosting me in Lancaster the year after. Both have been a great time and I greatly look forward to joining you for floorball in the future.

Daarnaast wil ik WTMC bedanken voor haar bijdrage aan het drukken van dit proefschrift.

Samen met de BHV-piketploeg heb ik tijdens mijn promotietraject mogen bijdragen aan de veiligheid op de campus. Naast dat dit leerzaam, (hopelijk) nuttig en zonder twijfel erg leuk was, was het vooral heel fijn om in zo'n hecht team te werken. Dus, Christel, Danny, Geert-Jan, Gerben, Ivo, Lars, Lianne, Marty, Michel, Mirjam, René, Ricardo, Stefan en Victor, dank voor deze fantastische ervaring.

Zowel voor als tijdens mijn periode als promovendus heb ik betrokken mogen zijn bij het FNWI honoursprogramma. Dank aan alle mensen die dit mogelijk hebben gemaakt. Mijn bijzondere dank gaat uit naar Jan van de Broek, van wiens wijze lessen en adviezen ik nog dagelijks profijt heb.

Zij die mij kennen weten dat dit proefschrift nooit was geworden wat het nu is, zonder de nodige uurtjes sporten. Net als het werken aan een proefschrift gebeurt dit meestal in stilte en eenzaamheid, maar is de bijdrage van trainers, teamgenoten of sparringpartners ronduit essentieel. Mijn oprechte dank dus voor alle uitdaging, voldoening en plezier aan de karateka's van Kuroshiro, de lopers van 't Haasje en de squashers van ISiS.

Dit proefschrift was zeker ook niet hetzelfde geworden zonder de steun van mijn vrienden. Anouk, Carmen, Cindy, Dorien, Hannah, Inge, Iris, Jop, Laura, Myrte, Valijn, Vincent, Vincent, Wouter en alle andere, dank jullie wel voor alle wandelingen, spelletjes, etentjes, gekkigheid en goede gesprekken. Het heeft me altijd enorm veel vreugde en energie gegeven.

In het bijzonder noem ik natuurlijk Marco. Marco, ja, waar moet ik beginnen? Laat ik het naar goede traditie bij drie dingen houden. Drie maal Dank. Voor dat je er altijd was; bij vreugde en verdriet, je was altijd de eerste. Voor onze dagenlange wandelingen, met mooie gesprekken, een muziekje of samen in onze eigen stilte. En voor onze urenlange Skype gesprekken, met drie dingen, een boek en gezamenlijk op afstand gedronken thee. Normaal mogen we altijd meer dan drie dingen noemen, als die er zijn. Ik had hier dan ook nog een hele poos door kunnen gaan, maar laten we het daar bij onze volgende kop thee nog eens over hebben.

Heleen, ook aan jou wil ik mijn oprechte dankbaarheid hier tonen. Je bent al die jaren een enorme steun geweest. Je hebt me zoveel geleerd over wat echt belangrijk is en zoveel geleerd over mezelf. Je motiveerde me, remde me op de juiste momenten af en was er altijd om vreugde en tegenslag mee te delen. Dank je wel voor je liefde, vriendschap, luisterend oor en goede raad. Dank je wel ook Ria, voor dat je deur altijd voor me open stond.

Een speciaal woordje wil ik richten aan Axel, mijn grote kleine broertje. De laatste jaren was je fysiek iets verder weg dan voorheen, maar toch altijd dichtbij. Je bent al die tijd een voorbeeld voor me geweest. Ik bewonder je rust, je vermogen om alles tot je te laten komen of juist alles van je af te laten vallen, je kunst om met een gevatte opmerking alles vlot te trekken, en je onvoorstelbare en onophoudelijke werklust. Je hebt me enorm geïnspireerd. En wat kunnen we samen fijn lachen.

Als laatste: Paps, Mams, speciale en grote dank gaat uit naar jullie. Dank jullie wel voor alles wat jullie me meegegeven hebben. Dank voor alle liefde en onophoudelijke steun. Dank voor de interesse in mijn werk en voor de trots die jullie uitstraalden als er wel eens iets lukte. Zonder dit alles was ik nooit tot hier gekomen.

Als allerlaatste toevoeging: Dank jullie allemaal voor het lezen van (op zijn minst een klein deel van) mijn proefschrift.

Een paar regels in het dankwoord van dit proefschrift zijn wel een heel magere manier om mijn dank naar al deze mensen uit te spreken. En dan zijn het ook nog eens bijna de laatste regels. Met deze woorden houdt mijn dank voor alles wat jullie voor me betekend hebben echter geenszins op. Hopelijk mogen wij nog lang samen door gaan.

About the Author

Serge Horbach was born on April 22, 1993 in Roermond, the Netherlands. He obtained his Bachelor's (Summa Cum Laude) and Master's (Cum Laude) degree in Mathematics from the Radboud University Nijmegen. During his master's he followed the Science in Society track, getting acquainted with science-society interactions, science policy debates, and the field of science studies. As part of this master's track he wrote his master's thesis within the European PRINTEGER project about the shifting discourse on research integrity in academic, policy and societal debates. For this thesis, he was awarded Radboud University's University Study Award in 2017.

After obtaining his master's degree in 2016, he started his doctoral research at the Institute for Science in Society at Radboud University. He initially did this in the context of the PRINTEGER project, studying organisational dimensions of research integrity. After obtaining a ZonMw grant in the *Fostering Responsible Research Practices*-programme, he continued his doctoral research within the IMPER and PREP projects, focussing on editorial procedures and their ability to contribute to research integrity.

During his doctoral research he obtained training in Science and Technology Studies at the WTMC graduate school. Furthermore he stayed as a part-time guest researcher at the Centre for Science and Technology Studies, Leiden University. He also paid research visits of several weeks to the Department of Sociology at Lancaster University and the Danish Centre for Studies in Research and Research Policy at Aarhus University.

After his time as a doctoral candidate he aims to pursue his academic career as a postdoctoral research at Aarhus University's Danish Centre for Studies in Research and Research Policy, working together with Prof. dr. Jesper Schneider and Dr. Kaare Aagaard.

List of publications

Articles in this dissertation

Horbach, S. P. J. M., & Halffman, W. (2019). The extent and causes of academic text recycling or 'self-plagiarism'. *Research Policy, 48*(2), 492-502. doi: 10.1016/j.respol.2017.09.004

Horbach, S. P. J. M., & Halffman, W. (2017). The ghosts of HeLa: How cell line misidentification contaminates the scientific literature. *Plos One, 12*(10), e0186281. doi: 10.1371/journal.pone.0186281

Horbach, S. P. J. M., & Halffman, W. (2018). The changing forms and expectations of peer review. *Research Integrity and Peer Review, 3*(1), 8. doi: 10.1186/s41073-018-0051-5

Horbach, S. P. J. M., & Halffman, W. (2019). The ability of different peer review procedures to flag problematic publications. *Scientometrics, 118*(1), 339-373. doi: 10.1007/s11192-018-2969-2

Horbach, S. P. J. M., & Halffman, W. (2019). Journal Peer Review and Editorial Evaluation: Cautious Innovator or Sleepy Giant? *Minerva*. doi: 10.1007/s11024-019-09388-z

Horbach, S. P. J. M., & Halffman, W. (Under review). Innovating the editorial process: An ethnography of publishing. *Science, Technology & Human Values*.

Horbach, S. P. J. M., Breit, E., Halffman, W., & Mamelund, S.-E. (Under review). On the willingness to and consequences of reporting research misconduct: the role of power relations *Science and Engineering Ethics*.

Horbach, S. P. J. M., Breit, E., & Mamelund, S.-E. (2018). Organisational responses to alleged scientific misconduct: Sensemaking, sensegiving, and sensehiding. *Science and Public Policy*, 46(3): 415-429.. doi: 10.1093/scipol/scy068

Other articles

Horbach, S. P. J. M., & Halffman, W. (2017). Promoting Virtue or Punishing Fraud: Mapping Contrasts in the Language of 'Scientific Integrity'. *Science and Engineering Ethics, 23*(6), 1461-1485. doi: 10.1007/s11948-016-9858-y

Horbach, S. P. J. M., & Halffman, W. (2017). Text recycling reveals flaws in science's reward system. *Research Europe, 463*, 7.

Horbach, S. P. J. M., Van der Voort, I. E. H., & Halffman, W. (2018). Hoeveel cellen met het verkeerde etiket dolen er nog rond? *Analyse - Vakblad voor biomedisch laboratoriummedewerkers, februari 2018*, 2.

Forsberg, E.-M., Anthun, F. O., Bailey, S., Birchley, G., Bout, H., Casonato, C., . . **Horbach, S. P. J. M.,**., …, Zöller, M. (2018). Working with Research Integrity—Guidance for Research Performing Organisations: The Bonn PRINTEGER Statement. *Science and Engineering Ethics, 24*(4), 1023-1034. doi: 10.1007/s11948-018-0034-4

Horbach, S. P. J. M. (2018). Universiteiten, geef het goede voorbeeld en stimuleer integer onderzoek. *Trouw*, 22-10-2018, opinie. Amsterdam, De Persgroep Nederland.

Horbach, S. P. J. M., & Halffman, W. (2019). Evidence-based peer review can make research more robust. *Research Europe, 489*, 8.

Horbach, S. P. J. M., & Halffman, W. (2019). Rejoinder to 'Is there a Peter Nijkamp effect?' by Jasper Lukkezen. *Research Policy, 48*(2), 505. doi: 10.1016/j.respol.2018.10.024

Horbach, S. P. J. M., & Halffman, W. (2019) Wangedrag maskeert problemen in de wetenschap *De Volkskrant,* 19-09-2019, Opinie & Debat, Amsterdam, De Persgroep Nederland.

Horbach, S. P. J. M., Nuijten, M., O'Neil, G., &Tijdink, J. (2019). *Position Paper: Fostering Responsible Research Practices - Start document for ZonMw's BVO program*. The Hague: ZonMw.

van der Kamp, J., **Horbach, S. P. J. M.**, & Halffman, W. (Under Review). Journal editorial procedures should be more transparent. *British Medical Journal*.

Datasets

Horbach, S. P. J. M. & Halffman, W. (2017). The Ghosts of HeLa: How cell line misidentification contaminates the scientific literature. Nijmegen, DANS. https://doi.org/10.17026/dans-2ap-7bnu

Horbach, S. P. J. M. & Halffman, W. (2018). Peer review procedures - results from a survey among journal editors. Nijmegen, DANS. https://doi.org/10.17026/dans-z82-29dd