Towards a Responsible Research Climate

Findings from academic researchers in Amsterdam

Tamarinde Haven

Towards a responsible research climate

Findings from academic researchers in Amsterdam

Tamarinde Laura Haven

Design & Lay-out: Vera van Beek
Printing: ProefschriftMaken || www.proefschiftmaken.nl

This PhD project was part of the "The Epistemic Responsibilities of the University" project, sponsored by the Templeton World Charity Foundation. (Grant #TWCF0163/AB106.). The views expressed in the dissertation are those of the author and her co-authors, and don't necessarily coincide with those of the Foundation. This PhD thesis was also sponsored by the Vrije Universiteit Amsterdam, Universiteit van Amsterdam and the Amsterdam UMC.

VRIJE UNIVERSITEIT

Towards a responsible research climate

Findings from academic researchers in Amsterdam

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor of Philosophy
aan de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus prof.dr. V. Subramaniam,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Geesteswetenschappen
op donderdag 14 januari 2021 om 9.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Tamarinde Laura Haven

geboren te Amersfoort

promotoren:	prof.dr. L.M. Bouter
	prof.dr. R. van Woudenberg
copromotor:	dr. J.K. Tijdink

Thesis assessment committee:

Prof. Ana Marusic, MD, PhD
Prof. dr. Yvonne Erkens
Prof. dr. Maurice Zeegers
Prof. dr. Jelte Wicherts
Dr. Natalie Evans
Prof. dr. Ir. Jeroen de Ridder

For Dotty, great bun(dle) of love.

Table of contents

I

Chapter I

Introduction

I started my bachelor's degree in psychology in September of 2011, the year Diederik Stapel's transgressions became public knowledge (1-3). His case and others worldwide (for example, Hendrik Schön (4)) led to intense debates on research integrity (5-9). However, other than a vague remark about Stapel in a statistics tutorial, I remember no discussion about this subject among fellow students. And I certainly would not have predicted that I would devote a PhD thesis to it.

In this introduction, I will briefly review some thoughts on integrity in research that paved the way for this PhD project and led to the main research question. Next, I will clarify the terminology used throughout this dissertation, such as 'climate' and 'research integrity', and lay out the theoretical framework that describes how we believe the research climate influences research integrity. Finally, I will flesh out the applied methodology and close with an outline of the dissertation.

Some introductory remarks

The initial response focused on finding someone or something to blame. Stapel was generally taken to be an incidentally bad apple (10). In other words: Yes, his acts were bad and should not be condoned, but they don't reflect academic research in general.

All humans are prone to go astray, academic researchers included. What can we learn from areas of research that study misconduct? What explanations can and have been meaningfully applied to researchers that fabricate or falsify data? **Chapter two** looks into these different explanations and applies them to the Stapel case: What do we need to know about the various theories of misconduct to meaningfully apply them to research misconduct?

During the time that I completed my bachelor's and master's degrees, two ideas about research integrity gained prominence that are not unrelated[1] (11). Firstly, falsification, fabrication and plagiarism (FFP), the three classic deadly sins of research, are bad, but they may not be the most pressing problem (6,12–16). Secondly, the "just a bad apple" defense needs to be supplemented by a focus on the barrel as a whole (17–20). Let us review these in turn.

So what, then, is the most pressing problem, if not FFP? It became evident that there are a variety of smaller transgressions that may collectively be more harmful to the validity and reproducibility of academic research than the more extreme FFP cases like Stapel's, which were deemed to be infrequent occurrences. I refer to these throughout this dissertation as Questionable Research Practices (QRPs) (21). The reasoning is that when such practices are so much more frequent, their total harm to academic research may be larger and hence they deserve our attention. Furthermore, some argue that there

1 For an elaborate overview of the field of research integrity, the reader is referred to the report *Fostering Integrity in Research* by the National Academies of Sciences, Engineering, and Medicine (34).

is only so much you can do about 'bad apples' (22).

What exactly do I mean by 'harmful'? Let us say that the harm would be to the validity or the trustworthiness of the research (23), so the harm is some sort of epistemic harm — results are disseminated which are not fully correct (24,25) and/or not to be trusted.

Now, plagiarism's harm to validity and trustworthiness may be low, presuming the validity and the trustworthiness of research don't change when it's reported again without reference to its source[2]. Compared to this, the harm done by F&F is much greater. Still, there is solid reason to think that compared to FFP jointly, the harm done by QRPs is even greater.

The reason is this: take the 'harm' of falsification and fabrication to be 10 and their occurrence 1 in 50 over 3 years (so for every 50 researchers, 1 may commit falsification or fabrication, or FF, over a timespan of 3 years). Also, take the 'harm' of QRPs on average to be 1 and their occurrence to be 1 in 3, over the same timespan (so for every 3 researchers, one may commit QRP over a timespan of 3 years).

Imagine that a university has 1000 researchers and we follow this university for 3 years, then the harm of FF in our scenario would be 200, but the harm of QRP would be 333. In actuality, the prevalence of FF has been estimated to be 2% and the prevalence of QRPs is 34%[3] (12). Then one may argue that it might be more meaningful to focus on preventing QRPs where possible (as opposed to solely focusing on FFP), as their total harm seems to be greater.

The second recognition that took place is that the "just a bad apple" view needed to be supplemented by a focus on the barrel (14,22,26–30). I say supplemented, because there is a plurality of factors that have been related to research misconduct (that go beyond the 'bad apple'). The full proverb reads: "one bad apple ruins the barrel", so what was the state of the 'barrel' (31), or here: academic research? And what do I take the barrel to refer to, as there are a variety of terms for the 'barrel', like "research culture", "research climate", "research environment", and many more (32–34).

Research question and clarification of terminology

Throughout this dissertation, I will use the term 'research climate' to mean "the shared meaning organizational members attach to the events, policies, practices and procedures

2 However, it is not 0, since plagiarizing a finding and presenting it as new could be taken as incremental evidence for, say, the effectiveness of a particular treatment, which in turn could lead to inflated confidence in the treatments' results, or harm the validity of meta-analyses when the same data sets are, unintentionally, used repeatedly.

3 Of course, the harm of some QRPs, such as selective reporting of clinical trial results (54), is probably greater than 1, but this regards an imaginary average. The point of the explanation was to show that even on a charitable reading of the harm of FFP, QRPs are a problem that deserve attention.

they experience and the behaviors they see rewarded, supported, and expected." (35,36) (p. 115). Most readers will be more familiar with the term 'culture', so it may help to briefly put these terms side by side. I take the organizational culture to mean: "the shared basic assumptions, values, and beliefs that characterize a setting…" (37) (p. 362)). These two concepts are by no means unrelated, as studying the research climate involves a focus on the more tangible (e.g. behaviors) with a view to gaining insight into the more intangible (e.g. values) research culture (35). And importantly, it seems easier to intervene on policies — procedures of behaviors — than basic assumptions or values.

The research question guiding this dissertation is a mouthful, namely*: What do scientists of the four academic institutes in Amsterdam consider to be the most salient aspects of the research climate of their institution that promote or hinder research integrity, and which do they believe to be the most important barriers to responsible conduct of research (RCR) and the most promising interventions to prevent research misconduct (FFP) and questionable research practices (QRP)?*

Some conceptual clarification is useful here, such as what I take 'research integrity' and 'responsible conduct of research' to mean and how they relate to a concept more familiar to some readers, namely 'research ethics'. Let us briefly review some key sources.

Steneck (38) defines RCR as "conducting research in ways that fulfill the professional responsibilities of researchers, as defined by their professional organizations, the institutions for which they work and, when relevant, the government and public." (p. 55). According to him, RCR comprises both research ethics (that he takes as concerning moral principles) and research integrity (that he takes as having to do with the professional standards as defined by professional organizations).

The World Conferences on Research Integrity Foundation's description of research integrity is broader: "'Research integrity' refers to the principles and standards that have the purpose to ensure validity and trustworthiness of research." (23).

The European Code of Conduct for Research Integrity (ALLEA, 39) mentions in its preamble that it is: "A basic responsibility of the research community is to formulate the principles of research, to define the criteria for proper research behaviour, to maximise the quality and robustness of research, and to respond adequately to threats to, or violations of, research integrity."

The Netherlands Code of Conduct for Research Integrity (40) writes: "…research integrity is essential. This holds true for all disciplines. Research in the sciences and the humanities derives its status from the fact that it is a process governed by standards. That normativity is partly methodological and partly ethical in nature, and can be expressed in terms of a number of guiding principles: honesty, scrupulousness, transparency, independence and responsibility. Researchers who are not guided by these principles risk harming both the quality and the trustworthiness of research." (p. 7).

What these descriptions have in common is a focus on researchers' behavior, not on the applications or consequences of research findings. Research integrity thus has to do with the extent to which researchers' behavior is in line with some ideal as formulated by the research community (this may for example regard methodological ideals or moral ideals[4]), no matter whether we refer to these ideals as "professional standards" (Steneck), "criteria for proper research" (ALLEA), or "guiding principles" (Netherlands Code of Conduct).

These behaviors are often grouped into three classes[5], or shades (16): research misconduct (FFP – black), QRPs (grey) and RCR (white) (23,38). RCR thus refers to research that is conducted in line with the normative ideals and lives up to their meaning. See figure 1.

Figure 1. Types of behaviors as displayed in shades.

Theoretical framework

Our research question above could be approached from a variety of fields. Yes, I am situated in the department of philosophy, but I will be the first to note that this is not a traditional philosophical thesis. To name just a few options available: it could have been situated in the department of Science and Technology studies, and connected to Merton's (41) work on scientific norms. It could also have been approached from behavioral economics and built on work from Ariely (42) on why people lie. The research question could also have been tackled from the perspective of organizational psychology and the evidence we have from Simha and colleagues (43) on ethical climates.

The theory that most heavily influenced our research is organizational justice theory (44). Organizational justice theory, in a nutshell, supposes that the fairer people feel treated, the more likely they are to trust their organization, accept its decisions and not engage in questionable behavior or worse (29,44). But the reverse is also true, and applied to academic research: in an organizational research climate where perceived injustice is high, researchers should be more likely to engage in research misbehavior

4 Besides, there is a strong moral component to intentionally complying with these (methodological) ideals.
5 This class distinction is a little oversimplified, as these behaviours presumably exist along a continuum. Other taxonomies exist too, e.g. Hall & Martin (55) discuss four 'shades', distinguishing between inappropriate conduct and blatant misconduct.

or QRPs.

The diversity of approaches available to us is both an advantage and a disadvantage. It is an advantage because it allows us to compare and contrast different approaches. It is a disadvantage because none of the preceding approaches are entirely focused on research integrity, so the generalizability to academic research (in Amsterdam) is not always clear (said differently: sometimes the comparison is not valid, as for instance students may cheat for different reasons than academics who misbehave).

Methodology

The research question also left open what sort of methodologies to use in pursuit of an answer. Still, it limited us in the sense that we had to, in some way or another, connect directly with academic researchers in Amsterdam. We chose to do so by means of a survey questionnaire and focus group interviews that each came with their inherent strengths and weaknesses that I will reflect on throughout this dissertation and more elaborately in the discussion.

The first method, survey questionnaires, is quantitative. The survey questionnaire consisted of different questionnaires put together. Potential participants received an invitation via email to complete the questionnaires online in their own time. This allowed us to reach out to a large group of academic researchers. Participants provided their answers using a rating system, meaning that they express their individual perceptions through indicating a number on an answering scale and that the raw data consists of numbers (e.g. if the answering scale ranges from 1 'totally disagree' to 5 'totally agree', then a participant who fully agreed with the statement "I cannot find sufficient time to work on my publications" would put down 5). We then used all these *individual* perceptions to look for patterns on a *group* level. For example: To what extent do PhD students (as a group) feel treated fairly by their PhD supervisors?

The second method, focus group interviews, is qualitative. Focus group interviews are group debates, led by a moderator. Participants interact with the moderator, as well as with each other. They can follow-up on someone's remark and pose questions, so in some way the focus group tries to mimic a natural group conversation. The raw data are participants' language, recorded and transcribed. Here, we are interested in *these* researchers' viewpoints and less in to what extent those viewpoints reflect what *the average member* in their group thinks or believes. We then used all these different perceptions to write a detailed and context-sensitive story of what we learned through talking to these academic researchers.

Outline of the dissertation

Back to the case of Diederik Stapel. Some may argue that Diederik Stapel was a psychologist and Hendrik Schön a physicist and that the practices, procedures and behaviors in psychology and physics are so different, from those in their own discipline that these cases are beyond comparison: bad apples defy comparison with bad pears. Similarly, some may think the barrel looks rather different from the perspective of an established professor compared to a novel PhD student. In other words: it may not be useful to think of *one* barrel, as the research climate may be perceived differently depending on the academic rank or disciplinary field. In this dissertation, I distinguish three academic ranks (namely: PhD students, postdoc or assistant professors, and associate or full professors) and four broad disciplinary fields (namely: biomedical sciences, natural sciences, social and behavioral sciences and the humanities). **Chapter three** studies how, within each of these academic ranks and disciplinary fields, the academic research climate in Amsterdam is perceived. As its title reveals, the perceptions of the research integrity climate (the 'barrels') do indeed differ — at least among our sample of academic researchers in Amsterdam.

Yet there are some pressures that all researchers are subject to, because they are rooted in how academic research and contemporary knowledge dissemination are organized. One of these is publication pressure, as it is hard to climb the academic ladder without publishing (45). But measuring this can be difficult. Many readers will tend to agree with the statement: "Publication pressure harms science" (46). But does that tell us anything about whether they in fact have experienced publication pressure? I wholeheartedly agree with the statement "Domestic abuse harms marriage", but rest assured I am not in an abusive relationship (nor yet married). What this comparison is meant to convey, is that I can agree with a proposition about a phenomenon even if the phenomenon in question has never happened to me. If I want to say someone has experienced publication pressure, I need to know whether they experienced the phenomenon themselves and whether it was stressful to them. In **chapter four**, we describe how we revised the Publication Pressure Questionnaire. We used models on work stress (that look at stress as an interplay between demands and resources (47)) and added two subscales to tease apart whether someone 'just' disliked publication pressure or whether they themselves felt subjected to publication pressure.

With this revised instrument (48), we surveyed academic researchers about their degree of perceived publication pressure. We reasoned that if academic researchers are under extreme publication pressure, they may be more likely to cut corners (49–51). Now, a reader may wonder whether publication pressure is even part of the research climate (the 'barrel'), and whether we are not all of a sudden wandering the fruit market (more commonly referred to as the 'system of science'). To some extent we are. Yet, publication criteria for academic advancement are often formulated by research institutes

themselves. Similarly, publication resources are also partly determined by the research climate: Is there a tendency to help a colleague with a critical reviewer (resource) or is the departmental climate such that colleagues are mainly judged on the basis of their publication quantities (stressor)? **Chapter five** details our findings and, here again, we look into perceptions across different academic ranks and disciplinary fields because it seems natural to think that these perceptions differ depending on where one is situated on the academic ladder, or where the ladder stands.

Sometimes it is clear that an apple has gone 'bad' -- it may look mouldy or smell sour. But an apple can be less than perfect in different respects. Unbeknownst to you, it may have been produced under terrible labor conditions where the people that actually did the work are never acknowledged. Somewhat analogously, there are a range of research misbehaviors, like omitting an author that contributed significantly to a paper (52), that are not salient when you read some scientific work. Bouter and others (53) compiled a list of research misbehaviors that included FFP, but also many others. How pervasive might their effect be in Amsterdam? In what different ways might our apples have been produced unethically, unbeknownst to the naïve consumer? Again, we surveyed academic researchers' perceptions of their research climate, this time asking them to report misbehaviors they perceived and to rate how impactful they found them to be. Yet, there is only so much any list can cover and we wanted to look beyond that. In our focus groups, we asked researchers to deliberate on the different misbehaviors they perceived. When they had all the misbehaviors on the table, they had to arrive at a consensus on which were more harmful than others. We attempt to integrate our survey and focus group findings in **chapter six**.

Now it was time to connect some dots. We had studied misbehavior (the ways in which apples go bad), publication pressure (the 'fruit market') and the research climate ('barrel') separately, but we wanted to know more about the relationship between high publication pressure and a poor-quality research climate to research misbehavior (put in fruit-terms: understand the extent to which apples going bad can be ascribed to their barrel or to the fruit market conditions). In **chapter seven** we look at how much variance in perceived research misbehavior it is that publication pressure and the research climate explain.

So, the research climate matters, but what *is* a responsible research climate? What sort of characteristics would researchers ascribe to it? We look into this question in **chapter eight** and also asked researchers in the focus group interviews to discuss possible barriers to a responsible research climate. That is, given those characteristics they listed, what is standing in the way? Here, we wanted to look beyond the grand pressures such as the need to obtain funding. What policies or practices that are part of the *current* research climate may be hampering the creation of a responsible research climate? And what sort of creative approaches can a group of researchers come up with, collectively, to alleviate those barriers?

References

1. Levelt Committee, Noort Committee, Drenth Committee. Flawed science: The fraudulent research practices of social psychologist Diederik Stapel. 2012.
2. Vogel G. Psychologist accused of fraud on "astonishing scale." Science (80). 2011;334(6056):579.
3. Wise J. Extent of Dutch psychologist's research fraud was "unprecedented." BMJ. 2011;343(7831):969.
4. Beasley MR, Datta S, Kogelnik H, Kroemer H, Monroe D. Report of the investigation committee on the possibility of scientific misconduct in the work of Hendrik Schön and coauthors. 2002.
5. Kornfeld DS. Perspective: Research misconduct: The search for a remedy. Acad Med. 2012;87(7):877–82.
6. Bouter LM. Commentary: Perverse incentives or rotten apples? Account Res. 2015;22(3):148–61.
7. Simonsohn U. Just post it: The lesson from two cases of fabricated data detected by statistics alone. Psychol Sci. 2013;24(10):1875–88.
8. Wicherts JM. Psychology must learn a lesson from fraud case. Vol. 480, Nature. 2011. p. 7.
9. Lacetera N, Zirulia L. The economics of scientific misconduct. J Law, Econ Organ. 2011;27(3):568–603.
10. Kraut A. Despite occasional scandals, science can police itself. Chron High Educ. 2011;58(16):A72.
11. Aubert Bonn N, Pinxten W. A decade of empirical research on research integrity: What have we (not) looked at? J Empir Res Hum Res Ethics. 2019;14(4):338–52.
12. Fanelli D. How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. PLoS One. 2009;4(5):e5738.
13. De Vries, Raymond; Anderson, Melissa; Martinson B. Normal misbehavior: Scientists talk bbout the ethics of research. J Empir Res Hum Res Ethics. 2006;1(1):43–50.
14. Martinson BC, Anderson MS, de Vries R. Scientists behaving badly. Nature. 2005;435(7043):737–8.
15. John LK, Loewenstein G, Prelec D. Measuring the prevalence of questionable research practices with incentives for truth telling. Psychol Sci. 2012;23(5):524–32.
16. Johnson DR, Ecklund EH. Ethical ambiguity in science. Sci Eng Ethics. 2016;22(4):989–1005.
17. Lafollette MC. The evolution of the "scientific misconduct" issue: An historical overview. Proc Soc Exp Biol Med. 2000;224(4):211–5.
18. Geller G, Boyce A, Ford DE, Sugarman J. Beyond "compliance": The role of institutional culture in promoting research integrity. Acad Med. 2010;85(8):1296–302.
19. Steneck NH, Bulger RE. The history and future of instruction in the responsible conduct of research. Acad Med. 2007;82(7):829–34.
20. Sacco DF, Bruton S V., Brown M. In defense of the questionable: Defining the basis of research scientists' engagement in questionable research practices. J Empir Res Hum Res Ethics. 2018;13(1):101–10.
21. Medicine I of. Responsible Science: Ensuring the Integrity of the Research Process. Volume I. Washington, DC; 1992.

22. Sovacool BK. Exploring scientific misconduct: Isolated individuals, impure institutions, or an inevitable idiom of modern science? J Bioeth Inq. 2008;5(4):271–82.
23. World-Conferences-on-Research-Integrity. Mission of the WCRIF [Internet]. 2020 [cited 2020 May 11]. Available from: https://wcrif.org/foundation/mission
24. Ioannidis JPA. Why most published research findings are false. PLoS Med. 2005;2(8):0696–701.
25. Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, Julious S, et al. Reducing waste from incomplete or unusable reports of biomedical research. Lancet. 2014;383(9913):267–76.
26. Alberts B, Kirschner MW, Tilghman S, Varmus H. Rescuing US biomedical research from its systemic flaws. Proc Natl Acad Sci. 2014;111(16):5773–7.
27. Casadevall A, Fang FC. Reforming science: Methodological and cultural reforms. Infect Immun. 2012;80(3):891–6.
28. Steneck NH. Institutional and individual responsibilities for integrity in research. Am J Bioeth. 2002;2(4):51–3.
29. Martinson BC, Crain LA, De Vries R, Anderson MS. The importance of organizational justice in ensuring research integrity. J Empir Res Hum Res Ethics. 2010;5(3):67–83.
30. De Vries R, Anderson MS, Martinson BC. Normal misbehavior: scientists talk about the ethics of research. J Empir Res Hum Res Ethics. 2006;1(1):43–50.
31. Gino F, Ayal S, Ariely D. Contagion and differentiation in unethical behavior: The effect of one bad apple on the barrel. Psychol Sci. 2009;20(3):393–8.
32. Mumford M, Murphy S, Connelly S, Hill J, Antes A, Brown R, et al. Environmental influences on ethical decision making: climate and environmental predictors of research integrity. Ethics Behav. 2007;17(4):337–66.
33. IOM. Integrity in scientific research: Creating an environment that promotes responsible conduct. Washington, D.C.: National Academy of Sciences; 2002.
34. Medicine), NASEM (National Academies of Sciences, Engineering A. Fostering Integrity in Research. Washington, D.C.; 2017.
35. Schneider B, Ehrhart MG, Macey WH. Organizational climate and culture. Annu Rev Psychol. 2013;64(1):361–88.
36. Wells JA, Thrush CR, Martinson BC, May TA, Stickler M, Callahan EC, et al. Survey of organizational research climates in three research intensive, doctoral granting universities. J Empir Res Hum Res Ethics. 2014;9(5):72–88.
37. Schneider B, Ehrhart MG, Macey WH. Organizational Climate and Culture. Annu Rev Psychol. 2013;64(1):361–88.
38. Steneck N. Fostering integrity in research: Definition, current knowlege, and future directions. Sci Eng Ethics. 2006;12(1):53–74.
39. ALLEA (All European Academies). The European code of conduct for research integrity. 2017.
40. Netherlands Code of Conduct for Research Integrity. 2018.
41. Merton R. The sociology of science: Theoretical and empirical investigations. Chicago: University of Chicago Press; 1942. 1–605 p.

42. Gino F, Ayal S, Ariely D. Self-serving altruism? The lure of unethical actions that benefit others. J Econ Behav Organ. 2013;93:1–14.
43. Simha A, Cullen JB. Ethical climates and their effects on organizational outcomes: Implications from the past and prophecies for the future. Acad Manag Perspect. 2012;26(4):20–34.
44. Martinson BC, Anderson MS, Crain AL, De Vries R. Scientists' perceptions of organizational justice and self-reported misbehaviors. J Empir Res Hum Res Ethics. 2006;1(1):51–66.
45. Van Dalen HP, Henkens K. Intended and unintended consequences of a publish-or-perish culture: A worldwide survey. J Am Soc Inf Sci Technol. 2012;63(7):1282–1293.
46. Tijdink JK, Smulders YM, Vergouwen ACM, de Vet HCW, Knol DL. The assessment of publication pressure in medical science; validity and reliability of a Publication Pressure Questionnaire (PPQ). Qual Life Res. 2014;23(7):2055–62.
47. Karasek R, Amick B. The Job Content Questionnaire (JCQ): An instrument for internationally comparative assessments of psychosocial job characteristics. J Occup Health Psychol. 1998;3(4):322–55.
48. Haven TL, Tijdink JK, De Goede MEE, Oort F. Personally perceived publication pressure - Revising the Publication Pressure Questionnaire (PPQ) by using work stress models. Res Integr Peer Rev. 2019;4(7):1–9.
49. Tijdink JK, Verbeke R, Smulders YM. Publication pressure and scientific misconduct in medical scientists. J Empir Res Hum Res Ethics. 2014;9(5):64–71.
50. Martin BR. Ethics and integrity in publishing. In: Clark T, Wright M, Ketchen DJ, editors. How to get published in the best management journals. 1st ed. Cheltenham, UK: Edward Elgar Publishing; 2016. p. 29–48.
51. Necker S. Scientific misbehavior in economics. Res Policy. 2014;43(10):1747–59.
52. Anderson MS, Kot CF, Shaw MA, Lepkowski CC, De Vries RG. Cross-national differences complicate allocation of credit and responsibility. Am Sci. 2011;99(3).
53. Bouter LM, Tijdink J, Axelsen N, Martinson BC, ter Riet G. Ranking major and minor research misbehaviors: results from a survey among participants of four World Conferences on Research Integrity. Res Integr Peer Rev. 2016;1(17):1–8.
54. Dwan K, Williamson PR, Gamble C, Higgins J, Sterne J, Altman DG, et al. Guidance to detect, evaluate and prevent the problem of selective reporting in trial publications. Trials. 2013;14(S1):O91.
55. Hall J, Martin BR. Towards a taxonomy of research misconduct: The case of business school research. Res Policy. 2018;48(2):414–27.

2

Chapter 2

Explanations of Research Misconduct, and How They Hang Together

Accepted for publication in Journal for General Philosophy of Science

Abstract

In this paper, we explore different possible explanations for research misconduct (especially falsification and fabrication), and investigate whether they are compatible. We suggest that to explain research misconduct, we should pay attention to three factors: (i) the beliefs and desires of the misconductor, (ii) contextual affordances, (iii) and unconscious biases or influences. We draw on the three different narratives (individual, institutional, system of science) of research misconduct as proposed by Sovacool to review six different explanations. Four theories start from the individual: rational choice theory, bad apple theory, general strain theory and prospect theory. Organizational Justice Theory focuses on institutional factors, while New Public Management targets the system of science. For each theory, we illustrate the kinds of facts that must be known in order for explanations based on them to have minimal plausibility. We suggest that none can constitute a full explanation. Finally, we explore how the different possible explanations interrelate. We find that they are compatible, with the exception of explanations based on Rational Choice Theory and Prospect Theory respectively, which are incompatible with one another. For illustrative purposes we examine the case of Diederik Stapel.

Introduction

Over the past few years, interest in research misconduct has substantially increased (1). While not everyone agrees about what should be labeled a research mis*behavior*, there is general consensus on what has been called research mis*conduct*: falsification, fabrication and plagiarism (FFP) (2,3). This consensus is reflected in codes of conduct, both national and international (4,5).

This paper has a twofold aim. First, to explore and discuss a number of possible *explanations* of research misconduct, and second to use this as a case study for the more philosophical question: how do these different explanations relate to one another: are they compatible, or are they not?

This paper potentially has practical relevance in that explanations of research misconduct can be expected to give a handle on what can be done to prevent research misconduct. This being said, this paper focuses on explanation, not prevention.

This paper is organized as follows. In section 1 we describe various types of research misconduct, and describe one actual case for concreteness' sake, as well as for the sake of future reference. Section 2 discusses what to expect from an explanation. The next section presents and discusses a number of explanations of research misconduct and explores what needs to be known if those explanations are to have some minimum level of credibility. In section 4 we discuss the more philosophical question of how these explanations hang together. We conclude with some overall remarks.

1. Research Misconduct

The most extreme kinds of research misbehaviors –fabrication, falsification, and plagiarism (FFP)— are at the same time not the most frequent ones (6,7). Much more frequent are the numerous 'minor offences', the many cases of 'sloppy science', the 'questionable research practices' (QRPs)(3). According to recent surveys, examples of frequent QRPs are: failing to report all dependent measures that are relevant for a finding (8), insufficient supervision of junior co-workers (9); selective citing to enhance one's own findings or conviction; and not publishing a 'negative' study (10,11). Despite their presumed frequency, assessment of the wrongness of the QRPs can be less than straightforward. Here, context, extent and frequency matter. The wrongness of FFP is more evident and codes of conduct are typically developed in order to prevent these (For an excellent overview of different reasons for using a wide or narrow concept of research misconduct, see (12)).

The reason why research misconduct needs to be prevented is somewhat different for falsification and fabrication compared to plagiarism. Whereas falsification and fabrication distort the *creation* of scientific knowledge, plagiarism need not distort the

field nor hamper its progress. Plagiarism fails to connect the knowledge to its proper origin, but it need not distort scientific knowledge per se (3,7). Also, explanations for plagiarism can be expected to differ from explanations for falsification and fabrication. Some plagiarism, for example, is committed by non-fluent English authors who borrow well-written sentences or even entire paragraphs for their own work, which is an explanation that is not available for cases of falsification and fabrication. We therefore focus on the latter two.

For illustrative purposes, we will examine a case of actual research misconduct in order to review the applicability of explanatory theories of research misconduct. We chose the case of Diederik Stapel for two main reasons. First, because his fraud has been established beyond reasonable doubt. Second, because there is sufficient publicly available information about the case: information about the committees' way of assessing the case, as well as about Stapel's own responses and reflections on his case. The more details of a case that are available, the better we can discuss the explanatory power of the theories we shall review. With the disclaimer that it is not our aim to provide an explanation of Stapel's fraudulent behavior, and that others have produced interesting accounts of it (for example, see (13,14)), we now offer a very brief description of the Stapel case.

Diederik Stapel was a professor of cognitive and social psychology. His research included topics such as the influence of power on morality, the influence of stereotyping and advertisements on self-perception and performance, and other eye-catching topics (15). He was an established figure whose findings often appeared in (inter)national newspapers. Stapel was accused of data falsification by three whistle blowers from within Tilburg University, where Stapel was employed in 2011, the year the case became public. In total, three committees investigated whether Stapel's work at the University of Amsterdam, University of Groningen and finally University of Tilburg, was indeed fraudulent (16). The committees established that, whilst the studies were carefully designed in consultation with collaborators, Stapel fabricated the data sets from scratch. In another variant, the data were gathered but altered by Stapel after a student-assistant had forwarded them to him. Finally, Stapel had at times reached out to colleagues inviting them to use some data he claimed to have 'lying around'.

Stapel has admitted that he engaged in these practices. The committees concluded that Stapel intentionally falsified and fabricated data. None of Stapel's co-authors were found to have collaborated with him in this regard. We will provide more information about the case as we proceed.

2. What to Expect from an Explanation

It is fair to say that currently, we have no single unifying theory of explanation (Woodward starts his book with a similar remark, see (17)). What we have is a wide assortment of ideas that are all claimed to be at least sometimes relevant for understanding explanation. One idea is that explanation is closely linked with *causation*: an explanation of X can be achieved by pinpointing the causal factors relevant to X. Another that it is closely linked to *laws*: an explanation of X is achieved by referring to laws under which X can be subsumed. Yet another idea is that explanation is linked with *unification*: an explanation of the phenomena X, Y and Z is achieved by showing that X, Y and Z are special cases of a more general phenomenon GP. A further idea is that explanation sometimes has to do with *reasons* (as opposed to causes): an explanation of a person's action A is achieved by citing her reasons, i.e. her beliefs and desires, for doing A.[1] In the social and behavioural sciences, this idea is sometimes coupled with the idea mentioned above that explanation is linked with laws. This approach to explaining human behaviour aims to formulate empirical generalizations of the form: If person P desires D, and believes that action A is the most efficient means of attaining D, then P does A. The hope is that such generalizations can be improved so as to state genuine laws, laws that enable prediction. Whether this hope is a realistic one need not detain us here. The important point to note is that reference to a person's reasons often has explanatory force.

However, it is often not *just* a person's reasons that have explanatory force; they often have it in conjunction with what we shall call "affordances": the specific situations in which a person acted and in which certain possibilities are open to him. The explanation of the fact that A shot B cannot consist of merely citing A's desire that B be dead and his belief that pulling the trigger was a way to attain that goal. A factor in the explanation should surely be the availability of a gun to A. The availability of the gun is a contextual affordance for A.

We should add that some behaviors can be explained independently of the actor's reasons, and independently even of the actor's being aware of displaying those behaviors. There are unconscious influences on human behavior, like the biases and heuristics that psychologists have been researching, and reference to them can also do explanatory work (see (18,19)).

To conclude: if we want to explain cases of research misconduct, we should pay attention, among possible others, to the following factors:

1 We note that reasons can serve different roles: they can be *motivating* and they can, even at the same time, be *normative*. P's motivating reasons are the reasons *for which* P did A—*the considerations in light of which* P did what she did, and that motivated her for doing A. Normative reasons are the reasons that P would cite in favor of her action A, reasons that would show that A was the sensible, or right thing to do. This way of making this distinction is borrowed from (40). Anscombe (41) offers a subtle analysis of the notion "explaining behavior".

- I: the desires and beliefs of the misconductor, meaning his or her (motivating) reasons;
- II: the contextual affordances available to the misconductor;
- III: unconscious influences.[2]

In an actual case of misconduct, all these factors may be at work. We should therefore heed the distinction between *partial* and *full* explanations. A full explanation of an event specifies *all* the factors that jointly guarantee the occurrence of the event. A partial explanation, by contrast, specifies a factor, or several factors, that facilitate the occurrence of the event, but do not guarantee it. It remains an open question (for us at least) whether full explanations of human behavior are even possible.

Explanations in the social sciences can take various forms. One that will figure quite prominently in our discussion are inferences to the best explanation (IBEs). A key feature of IBEs is that the factor doing explanatory work is not directly observed, but concluded to.[3]

3. Explanations of Research Misconduct

In a helpful article, Benjamin Sovacool (20) distinguishes three 'narratives' about research misconduct: one in terms of (i) impure individuals, another in terms of the (ii) failures of this-or-that particular university or research institute, and yet another in terms of (iii) the corrupting structure of the practice of modern science as such—three narratives that he suggests are incommensurable. Even if these narratives don't explain in any straightforward way individual cases of research misconduct, they are helpful for two reasons.

First, narratives can deliver cognitive goods that are distinct from explanations - they can provide understanding. And, as Peter Lipton (21) has argued, there can be understanding without explanation. Even if we have no explanation of Stapel's fraudulent behavior, it does give insight into the whole affair if the evidence indicates that Stapel was only one bad apple, or if it indicates that the institute at which he worked was failing in important respects, or if the whole structure of science turns out to be corruptive. Second, Sovacool's narratives are helpful as they *do* point to places we could look for explanations. For example, the narrative that a case of research misconduct is due to an

2 Note that the factors we describe seem to match up with what has been termed levels of explanation, e.g. an explanation using desires and beliefs would be an explanation on the personal level, etc. (see Owens (42). Yet, we will not focus on the question of whether an explanation on one level is more fundamental than an explanation on another; our aim is merely to assess the plausibility of the explanations and whether they are compatible with each other.

3 See Lipton (43), ch. 4. Standard examples of IBEs are the doctor's inference that his patient has measles, since this is the best explanation of the symptoms; and the astronomer's inference to the existence and motion of Neptune, since that is the best explanation of the observed perturbations of Uranus.

impure individual (and not a failing research institute, nor something like the corruptive structure of science as such) doesn't explain in any detail why Stapel engaged in the misbehavior he did, but the narrative (if true) *does* point to what is needed for such an explanation: the nature of his impure character needs to be understood, so that we can see how Stapel's specific impurity led to the misbehaviors that made him notorious. Likewise, the narrative that the misconduct is due to a failing research institute doesn't explain Stapel's behavior, but it *does* point (if true) to where to look for an explanation: to the operative rules and procedures of the institute, perhaps, to its 'culture' or 'climate' ('there was an atmosphere of terror'), etc.

Of course, things get complicated here. For if Stapel's misbehaviors are due exclusively to factors covered in the narratives about the institutions he was part of (or about the structure of science as such), then we should expect other members of those institutions to have displayed similar misbehaviors—which, as far as we know, they haven't. And this is a reason for thinking that Stapel's misbehaviors are due *not* exclusively to institutional failings, but also, say, to personal impurities like character flaws. The distinction between partial and full explanations is a recognition of this complication.

We draw attention to the fact that whereas explanations under Sovacool's first narrative will typically refer to type I and III factors (beliefs and desires; unconscious influences), explanations under Sovacool's second and third narrative will refer to type II factors (contextual affordances). Since all these factors, possibly and likely, can play a role in cases of research misconduct, we need not assume that the explanations under Sovacool's three narratives are *per se* incommensurable if that entails they are incompatible. In fact, as we will argue in section 4, most of these explanations are compatible with each other, as they are partial at best.

To conclude: Sovacool's narratives don't offer explanations of cases of research misconduct, but they point to where to look for explanations. We discuss six[4] different (types of) theories that might help explain research misconduct.[5] Our aim here is to specify what we need to know about a specific case in order for such explanations to get a good start. Whether they are credible, is a further issue. We begin with four theories that fall under Sovacool's first kind of narrative.

4 Our search for theories was guided by a similar endeavor of Gjalt de Graaf's (44), in which he discusses a number of theories that purport to explain corrupt or fraudulent behavior in public administration, such as taking bribes. We supplemented his list with additional theories where relevant. Interestingly, de Graaf also included correlation 'theories', but as these are not theories in that they do not contain an idea about the explanatory mechanism, we do not review them in-depth but will elaborate briefly on their relevance in the concluding section.

5 Note that the (types of) theories De Graaf (44) reviewed are theories of human behavior that come from different fields such as economics, sociology or criminology and seem to work on different levels. Hence, these theories often apply to terrains beyond fraud in public administration and may not even have been designed to explain fraud in public administration, but *have* been *appealed to* to explain fraud in public administration. In a similar vein, we explore whether these theories that *have* been coined to explain research misconduct are actually applicable and compatible.

First: Rational Choice Theory.

Sometimes labeled 'rational choice theory', this theory has its origins in economics. It starts from an individual that is portrayed as rationally considering different options to tackle a particular problem. Rational Choice says that an individual actor faced with a risky outcome selects the specific behavioral action that yields the maximized anticipated payoffs, where the utility of his behavior is weighted by the probabilities of its occurrence. The domain of the utility function is absolute benefits and costs. The individual weighs the costs and benefits attached to each option, and next makes the calculation, on the basis of which she makes a decision[6]. This theory, that refers to type I factors only (beliefs and desires), is appealed to in the research integrity literature by Wible (22) as well as by Lacetera and Zirulia (23).

Suppose we apply this theory to Stapel's case. We will first describe what we think needs to be the case if this theory is going to provide an adequate explanation of his misconduct. Next we discuss whether (we know) these things are indeed the case.

If this theory is to explain Stapel's misconduct, we should envisage Stapel as a rational agent[7] who is calculating the costs and benefits, i.e. the utility, of cheating compared to playing it fair (i.e. observing the rules and principles that we now find in the numerous Codes of Responsible Conduct of Research). The benefits of (undetected) cheating probably include: more publications (or: more publications with outcomes that would be considered remarkable), which would contribute to greater prestige, which would increase the chances of obtaining more research funds, which would mean gaining more visibility, power and influence. The costs of cheating probably include: the fear of being found out (and fear of whatever else is set in motion by it: retraction of publications, loss of research funds, loss of prestige, loss of job, etc.), which means that one must always be on one's guard; loss of self-respect; not contributing to the (great!) cause of science. The costs of playing it fair include, probably: often having research results that

6 We are aware that rational choice theory is sometimes used as a general paradigm that is not to be applied to individual cases because the notion of a "rational choice" is deemed to be no more than a useful theoretical fiction. We side with those authors that have used rational choice theory to shed light on individual cases of human behavior.

7 What does it mean to be a 'rational' agent in this case? In jurisprudence, an important consideration for holding someone accountable is whether that person had the right mentality, or *mens rea*. The four generally distinguished levels of mentality are purpose, knowledge, recklessness and negligence,. Each of these correspond to a different extent to which the researcher, in our case Stapel, could be held accountable for his deeds, with the first being the highest level of accountability, purpose. Stapel's case maps most closely onto this level — in his own writings, he is explicit about his intention to deceive others. The mental state of knowledge would look something like this: a colleague of Stapel had reasonably strong doubts about Stapel's conduct, but decided to work with him regardless. Recklessness could perhaps be applied to cases of falsification, where a researcher runs a data analysis she does not fully understand and finds a significant result that she reports regardless. The level of negligence does not seem to work in our case, as it is unlikely to engage in misconduct out of negligence. So if one applies rational choice theory to cases of misconduct, one should be clear about whether that conduct was intentional, with knowledge, reckless or negligent, as rational choice theory seems more apt to explain cases where the trespasser had a *mens rea* of purpose or knowledge, compared to negligence or recklessness.

are not significant and/or interesting, which decreases the likelihood of one's research being published, which decreases the chances of getting research funds, and of making an impact. The benefits of playing it fair include: doing what one, from a moral point of view, ought to do, behaving in a responsible way (and virtue is its own reward, as the proverbial wisdom has it); increasing the chance that you will have research results that are genuine contributions to the cause of science; increasing the chance of receiving recognition that is based on substance.

If rational choice theory is going to give an adequate explanation of the falsifications and fabrications committed by Stapel, he must have engaged in a cost/benefit analysis of the cheating option as compared to the playing fairly option—and on that basis have decided that falsification and fabrication 'pay'.

Is there any evidence that Stapel did engage in a cost/benefit analysis of this sort? There are two main types of possible evidence here: the misconduct investigation reports and Stapel's own accounts. From the report (16) on Stapel's misconduct, we could deduce that the costs – at least, the fear of being found out – seemed low: "It was easy for researchers to go their own way. Nobody looked over their shoulder..." (p. 39). Stapel's own account[8] also points in this direction: "So when I started to fake my studies, it was very, very easy. I was always alone. Nobody checked my work; everyone trusted me. I did everything myself. Everything. I thought up the idea, I did the experiment, and I evaluated the results. I was judge, jury, and executioner. I did everything myself because I wanted to be one of the elite, to create great things and score lots of major publications" (p. 118-119).

Yet, it remains somewhat questionable whether Stapel actually engaged in a cost/benefit analysis. But this doesn't mean that the rational choice theory explanation is false or wrong. Stapel's engaging in such an analysis is at least a *possible* outcome of a rational choice IBE, for his fabrications and falsifications *may* be best explained by his having made a cost/benefit analysis. Whether it indeed *is* the *best* explanation, depends, of course, on the strength of alternative explanations. Moreover, as we noted, explanations can be *partial*. Rational choice theory, then, may offer only a *part* of a full (or fuller) explanation. As a matter of fact, this IBE, even if it is correct, can at best be a partial explanation only. For, as we suggested in the previous section, there must be contextual affordances (so type II factors), in this case: structures and systems that allow for the possibility of falsification and fabrication. And these affordances fall outside the scope of rational choice theory, as do type III factors.

8 Note that this regards the translation of Stapel's 2012 autobiographical book by Brown (26). Caution is needed when interpreting these statements, as it is arguably *oratio pro domo*. Stapel seems to acknowledge this as his foreword reads: "This is my own, personal, selective, biased story about my downfall." (p. iii). Similar to Zwart (14), it is not our primary concern whether "the autobiographical account actually corresponds with the facts… but rather what can be learned and gained from this ego-document" (p. 211-212).

Second: Bad apple theories.

Like rational choice theory, this theory too has its roots in economics. Here, the individual is depicted as someone with a flawed (moral) character. This flawed character is subsequently causally linked to corrupt acts. Greed is sometimes deemed to be an element in a flawed character. An example of a full-scale faulty character has been labelled in the literature as the Machiavellian personality type, that deems that the prestige associated with a particular goal justifies any means to attain it, even if those would be seen as unethical. Hren et al. (24) studied Machiavellism in relation to moral reasoning and Tijdink et al. linked personality types such as a Machiavellian character to research misbehaviour (25). Bad apple theories refer to type I factors only—to reasons that motivate certain characters to behave in certain ways.

If we apply this theory to Stapel and ask what should be the case if bad apple theories are to provide an adequate explanation of his misconduct, it is clear that he needs to have, or at the time have had, a flawed moral character—he needs to have a Machiavellian personality type for example, or some other flawed moral character[9].

Is there evidence that Stapel had a flawed moral character at the time—evidence coming from psychologists and psychiatrists, for example, who have done something like a personality-analysis on him? The only evidence that would point in that direction appears in Stapel's own book (26): "*It takes strong legs to carry the weight of all that success.* My legs were too weak. I slipped to the floor, while others—maybe wobbling, maybe with a stick to lean on—managed to stay upright. I wanted to do everything, to be good at everything. I wasn't content with my averageness; I shut myself away, suppressed my emotions, pushed my morality to one side, and got drunk on success and the desire for answers and solutions." (p. 148, emphasis original). Yet, this one passage seems insufficient as a basis for a solid psychological verdict on his character, and as far as we know we have nothing else to go on that is publicly available and would reliably demonstrate a flawed character.

Note that when we refer to a flawed character, we do not mean to insinuate that Stapel had no moral awareness whatsoever. The report (16) on his misconduct explicitly mentions that he taught the research ethics course. Stapel's account (26) confirms this: "I'm the teacher for the research ethics course, in which I get to discuss all the dilemmas with which I'm confronted every day, and for which I always make the wrong choice." (p. 129).

Even if we have no solid basis to draw a conclusion about Stapel's moral character, this doesn't mean a bad apple explanation can be ruled out. For it is possible to make an IBE, based on a bad apple theory, to the effect that Stapel's fraudulent conduct is best explained by the fact that he had, at the time, a flawed (moral) character. Whether this is really the best explanation, depends, again, on the strength of the available alternatives.

9 Although we mostly discuss moral character flaws, it seems plausible that intellectual character flaws, such as insouciance, play a similar role – insouciance example taken from Cassam (45).

It seems clear, however, that bad apple theories, even insofar as they are correct, cannot give us a full explanation of Stapel's misconduct. For there must be contextual affordances that allow flawed moral characters to commit acts of fabrication and falsification—and these are part of a full(er) explanation of the misconducts at hand.

Three: General Strain Theory.

Another theory that could be headed under the individual narrative is General Strain Theory (henceforth: GST) as originally developed by Agnew (27) who worked in the sociology of crime. GST sees misconduct as originating in stress or strain. These states of stress and strain bring about a negative emotional state in the researcher, like anger, sadness or depression—which are, broadly speaking, type I factors. As a third step, GST posits that the behavioral strategies researchers adhere to in order to cope with these negative states differ, and, importantly, strategies may include deviant behavior (in our case: research misconduct). This theory, which has been coined as playing a role in explaining research misbehavior by Martinson and colleagues (28), is put forth in the Institute of Medicine's report *Fostering Integrity in Research* (29), and recently came forward in research by Holtfreter and colleagues (30) wherein they asked US scientists what factors they believed to play a role in research misconduct.

If this theory is to do explanatory work, we need to know whether Stapel faced prolonged stressful situations, so prolonged that they put him in a persistent negative state. Strain is thought to be more difficult to mitigate if the strain affects one's central identity, and indeed, to many being a scientist is considered a vocation. The report on Stapel's misconduct is silent on this issue. In his book, Stapel himself, though, talks of a persistent state of stress he experienced: "Nothing relaxes me any more… but I feel stressed and restless. I want everything, and everything has to happen now. I want out. I don't want to have to write papers any more. I want to start over, get away from this fantasy world I've created, get out of this system of lies and half-truths, to another city, another job.". However, he experienced this after he got into the habit of altering his data.

GST theory presupposes that behavioral strategies to cope with the negative emotional states differ. Some individuals cope with the strain through downplaying it or distracting themselves (e.g., 'H-index really isn't that important'). Coping strategies differ between individuals depending on traits such as self-esteem, intelligence, and self-efficacy. The more an individual possesses these traits, the more effective they will be in coping with strain in legitimate ways (e.g., 'I don't need confirmation of my academic skills, I know that I am competent'). Thus, whereas Stapel's colleagues facing similar strains found other ways to cope, he turned to deviant behavior. But this is also a caveat: What exactly made Stapel turn to deviant strategies? Perhaps his environment was crucially different in some way, which fuelled his urge to create spectacular results? Here GST holds that if an individual works in an environment that functions as a constant 'reality-check', then this environment can hamper the effectiveness of the coping

strategies. In any case, GST can thus, at best, be a partial explanation. That is not to say that GST can be ruled out entirely, as it is possible, via an IBE, that his misconduct could be explained by GST – whether that is also the best explanation depends on the explanatory force of the alternative theories.

Four: Prospect theory.

The final theory that we shall consider under Sovacool's first narrative is prospect theory. The roots of prospect theory lie in the psychology of risk, but the theory has also been used in behavioral economics. In their study of risky choice, Kahneman & Tversky (31,32) found that individuals are more strongly motivated by fear of loss than potential gain, and are inclined to avoid risk when faced with potential gains, yet seek risk when faced with potential losses. Bearing in mind that the reference point of the individual researcher matters (their context -- whether that is one in which the researcher is faced with potential losses or gains), prospect theory would predict that researchers faced with potentially losing their job, tenure or other meaningful resources would be more prone to take risks, or in our case, to engage in research misconduct, than colleagues who face no such threats. This theory refers to type I and II factors, as the behavioral tendencies involved may, but need not, go unnoticed by the subject. The National Academies' report *Fostering Integrity in Research* offers this as a possible explanation in its chapter on the causes of deviance (29).

For this theory to explain Stapel's deeds, we need to know whether, at the point in time when he falsified or fabricated datasets, he was faced with the threat of losing his job, or tenure, or other meaningful resources. In addition, it would be useful to know if the opposite situation occurred, where Stapel was faced with a potential gain, perhaps greater chance of having his research accepted in a high-impact journal through the risky behavior of falsifying his data, and decided against it.

Stapel's book contains a passage of his reflection that reads: "There was no pressure, no power politics, no need to produce patents or pills, to compete in the marketplace or make a pile of money. It was always purely academic, scientific research, which makes any form of cheating even harder to understand." (p. 188). Another passage seems to point more at the potential for gain as a driving force: "I couldn't resist the temptation to go a step further. I wanted it so badly. I wanted to belong, to be part of the action, to score. I really, really wanted to be really, really good. I wanted to be published in the best journals and speak in the largest room at conferences." (p. 102-103).

The report (16) does not provide direct information on these issues, but it does detail that 55 of Stapel's publications rested on falsified or fabricated data. Even if we put aside the idea that different papers can be based on the same dataset, how often can one be faced with potentially losing their job, tenure or another meaningful resource? It seems likely that there were other factors at play, too. Again, that is not to say that prospect theory cannot be an explanation for research misconduct, but that it can at best

be a partial explanation. And even if, in Stapel's case, there was no direct evidence that he feared losing his job, this potential threat could be inferred via an IBE. This in turn sparks the question whether it is also the *best* explanation, given its competitors.

We now move on to consider a theory that aims to explain misconduct by referring to the institutions and organizations in which the perpetrator works, and thereafter to a theory that aims to explain it by referring to the structure of the practice of modern science in general. Explanations based on these theories refer to type II factors, contextual affordances.

Five: Organizational culture theories.

These theories find their roots in organizational psychology. They have in common that they consider people as working in an organization with a specific culture and a particular structure, and argue that these have an effect on individuals and their behavior. An assumption underlying these theories is that there is a causal path from a certain organizational culture, to a particular mental state, to an individual's behavior.

One particular organizational culture theory, called organizational justice theory, is based on the idea that people who perceive themselves to be treated fairly by their organization[10], behave more fairly themselves. Conversely, when the organizational procedures are perceived as unfair, people are more likely to engage in acts that make up for the perceived unfairness, e.g. falsifying or fabricating their data. Martinson and colleagues (33,34) have investigated this theory and they report that researchers who perceived their treatment as unfair were more likely to engage in research misconduct.

There are various ways in which the organization can influence the behavior of researchers, and the organization itself is not immune to external influences[11]. The Institute of Medicine's report *Integrity in Scientific Research: Creating an Environment that Promotes Responsible Conduct* (35) conceptualized the research organization as an open systems model. Within the organizational structure itself, there are policies and procedures in place that influence researchers, and within the organizational processes the IOM report emphasizes the role of leadership and supervision. These last two are

10 The organization is often studied through the organizational culture (the values, beliefs, and norms that help shape members' behavior) and the organizational climate, defined as "the shared meaning organizational members attach to the events, policies, practices, and procedures they experience and the behaviors they see being rewarded, supported, and expected." (46) (p. 115). We will look at both in our consideration of organizational justice theory, but as policies and procedures are more observable than values and beliefs, we will focus more heavily on the former when reviewing the empirical materials available.

11 It can be hard, in the case of academic research, to pinpoint the boundary at which the culture ends and the outside begins, which can be seen as a caveat of applying organizational justice theory to research misconduct. Sometimes we speak of the research culture in, say, psychology, referring to the scientific field at large. Related, internal means of promotion or tenure are influenced by review committees of papers and grants, which would traditionally be placed outside of the organizational culture ((see also (28)). Nevertheless, it seems reasonable to suppose that an individual researcher is most profoundly influenced by their local climate— by the policies that directly apply to them and by the practices they see their colleagues engage in and be rewarded for.

especially important, as studies on the organizational climate in academic and other settings found that organizational leadership, ethics debates and ethical supervision were associated with an ethical climate. The system is open in that it produces various outputs in the form of papers and other research related activities that in turn influence organisational inputs through funding and human resources, which in turn influence the organization again.

Another idea is that the organizational dynamics themselves can take such a form that everyone in the organization begins to engage in questionable practices. This type of unethical conduct may then become so frequent that it slowly becomes the normal way of conducting research.

If we apply this theory to the misconduct of Stapel and ask what should be the case if his misconduct is to be adequately explained by it, we must say that the culture and structure of the organizations he was employed by, somehow induced his conduct. Either there should be indications that he was mistreated by his organizations or there should be evidence that his work environment was perverted altogether. Delving deeper: Is there information available on their policies, the degree to which leadership emphasized integrity, or whether open debates about integrity issues were a regular occurrence? There must, perhaps, have been reward systems in place that triggered misconduct, or some element of an organization's culture that did the trick.

So, if such an explanation is to work for the Stapel case, what we need is insight into the culture and structure of the organizations that he worked with. Stapel seems to believe that culture played a role (26): "I'm not the only bad guy, there's a lot more going on, and I've been just a small part of a culture that makes bad things possible." (p. 171) Even if there was no direct evidence available about Stapel's research culture, it might be possible to make an IBE here too: from his misconduct we can draw conclusions suggesting a bad organizational culture and bad organisational structures—the latter explaining the occurrence of the former.

Interestingly, the report (16) about Stapel's misconduct devotes an entire chapter to the culture in which his fraud took place. It is described as "a culture in which scientific integrity is not held in high esteem" (p. 33) and "even in the absence of fraud in the strict sense, there was a general culture of careless, selective and uncritical handling of research and data." (p. 47). This may prompt one to believe that the culture indeed played a role in fostering Stapel's fraudulent behavior. However, the report (16) presents culture as an explanation for why the fraud could *sustain* for so long—"The Committees are of the opinion that this culture partly explains why the fraud was not detected earlier." (p.47)—not as one that brought about the fraud. Of course, this does not preclude the organizational culture from being a potential explanatory factor in the origination of the misconduct as well.

Are there indications that Stapel was structurally undervalued by his respective organizations, and treated unfairly? The report's (16) information points in the

opposite direction: "These more detailed local descriptions also reveal Mr Stapel's considerably powerful position, at any rate within the University of Groningen and even more so within Tilburg University. At the University of Amsterdam he already enjoyed a reputation as a 'golden boy'." (p.38). To our knowledge, there is no public evidence of a culture that treated researchers unfairly or that suggests Stapel's deeds could be interpreted as a means to make up for perceived unfairness done unto him. Can we know enough about the organization's culture and the structures of the units Stapel belonged to? Perhaps we can. But even if we do, the organizational culture explanation can at best be a *partial* one. For many other individuals who worked in the same organization, have not (we assume this to be so) committed acts of fabrication and falsification. For this reason we may think of an organization's climate and structure as contextual affordances that don't forestall misconduct, and don't cause it either, but do enable it.

Until a certain stage of investigation, it is possible to propose an organizational culture explanation of Stapel's behavior, namely as long as we have no evidence that any of the other explanations even partly explain it. At a later stage of the investigation, however, it should be possible to have more direct access to the organizational culture, as it should in principle be observable.

Six[12]: Ethos of public administration.

Ethos of public administration theories, at times labelled Taylorism or New Public Management (NPM) theories, have their roots in economics, and, applied to research misconduct, fall under Sovacool's third kind of narrative. These theories center around a complex set of ideas and concepts: specialization, command, unity, efficiency and atomism. The ideas that connect these concepts are, firstly, that individuals are naturally isolated from one another and that only an organization, through a chain of command and a sense of mission, can unify individuals into a single, efficient and rational working unit. The second is that individuals tend to laziness, selfishness and are not interested in any social good beyond their own individual good, and that therefore organizational unity and discipline must always be maintained.

12 De Graaf's (44) fourth type of theory is the theory of clashing moral values. The idea builds on Sellin (47) and is that particular values that are held in high regard in the private atmosphere may lead to behaviors that are undesirable in the public or work atmosphere. Davis (48) applied this to research misconduct cases: Take a researcher that comes from a culture where scientific productivity is the holy grail. After working for some time in a culture where adherence to ethical practice is regarded pivotal, some of the researcher's behaviors may be regarded fraudulent. Davis' argues that this can be fixed by subjecting the researcher to ethical training or developing codes of conduct that are endorsed widely. If this theory is to stand ground, Stapel should have been (successfully) socialized in a culture that held values which clash with ethical practice in extremely high regard. We think the analogy to research misconduct does not work here. Which value in the private atmosphere is supposed to do the explanatory work? If we look at the example from Davis, it is far from obvious that adherence to ethical practice clashes with productivity. We regard clashing moral values a nonstarter in the case of research misconduct.

The perverting influences of NPM or Taylorism on the academic system can be expressed through different phenomena that Halffman & De Radder (36) eloquently captured in their *Academic Manifesto*. They describe, among other phenomena, the "measurability for accountability" (p. 167), meaning the obsession with output quantifiers, be it publication indices, metrics, or impact factors. They also elaborate on the "permanent competition under the pretense of 'quality'" (p. 168), referring to the 'hypercompetition' where researchers compete against each other for funding in a 'winner takes it all' system, where it is the junior staff that do the bulk of the work, faced with temporary contracts and poor career opportunities (36).

Now, this extreme emphasis on effectiveness and performance can come at the cost of neglecting ethical issues and crowding out the values that motivate professional behavior and institute the organization's mission. When this happens, it can lead to corrupt individuals. Overman, Akkerman and Torevlied (37) seem to subscribe to this proposition when they write: "Academic misconduct is considered to be the logical behavioral consequence of output-oriented management practices, based on performance incentives." (p. 1140).

If this theory is going to explain Stapel's misconduct, what should be the case is that he worked in an organization with a strong focus on performance and output in a way that crowds out values and the acknowledgement thereof. Perhaps he started out with an intrinsic desire to do good research. However, the more his work's merit was determined by performance indicators and the more the focus was put on effectiveness, the more this intrinsic motivation was replaced by a desire to do good *according to* these performance indicators – to be effective and publish lots of papers. In addition, the emphasis on these performance incentives shifted attention away from responsible conduct of research.

Is there evidence that Stapel worked in such a system? Overman and colleagues describe that performance indicators indeed have become more evident among academic institutions in The Netherlands (they draw on research by Teelken (38)). Do we have evidence that increased emphasis on performance accounts for Stapel's actions? His own account (26) acknowledges the pressures in contemporary science: "Science is an ongoing conflict of interests. Scientists are … all in competition with each other to try and produce as much knowledge as possible in as short a time, and with as little money, as possible, and they try to achieve this goal by all means possible. They form partnerships with business, enter the commercial research market, and collect patents, publications, theses, subsidies, and prizes." (p. 189-190)

Perhaps we should consider the role of these performance indicators plus the reality of hypercompetition as biasing Stapel's view on research. Under their influence, he unconsciously focused more and more on effectiveness at the expense of ethical conduct. At some point, effectiveness itself became his main desire. One is reminded of Goodheart's law: "When a measure becomes a target, it ceases to be a good measure".

However, we are again left with the question why these indicators biased Stapel towards extreme efficiency and not his peers. Maybe his affordances were different from those of his peers, but these fall outside the scope of this theory. Hence the ethos of public administration or NPM, even if it is an acceptable explanation of misconduct, can best be thought of as a *partial* explanation.

As with the other theories, even if (so far) there is no direct evidence that a case of scientific fraud was caused (at least in part) by excessive emphasis on effectiveness and performance indicators, excessive emphasis could, indirectly, be inferred via an IBE. In which case the question arises whether it is also the *best* explanation, given its competitors.

4. Are the Different Explanations of Research Misconduct Compatible?

Having discussed six explanations of research misconduct, and having explicated what, for each of them, needs to be the case if they are to be accurate, if only partial, explanations, we now address the second question that we have for this paper: how do these explanations relate to each other? Two different explanations of the same phenomenon, E1 and E2, can be compatible, or they can be incompatible. And if they are compatible, further qualifications can be added—for example that E1 and E2 "add up", or that they reinforce each other, or that one weakens the other. We will focus our discussion on the compatibility question, and abstract from further qualifications. The reason for this is that which of the qualifications applies, will depend on the details of the case whose explanation one seeks, whereas the mere compatibility of explanations can be discussed in the abstract.

Given that we have six explanations on our hands, this means there are 15 pairs of explanations to consider. We can reduce this number to six pairs, because each of the four explanations under the first narrative are individually, and in their very nature, compatible with the explanations under the second (institutional) and third (system of science) narratives. This is in the nature of the case, as the first focus on qualities of the misconductor, and the latter two on contextual affordances—none of which, we suggested, constitute full explanations. We don't want to make this point only at this abstract level, but want to offer one illustration. Consider bad apple explanations and organizational culture explanations. It would seem that such explanations (of the same behavior) are *at least* compatible. If cheating can be adequately, if only partly, explained by reference to the ill treatment that the cheater has suffered in an earlier stage, then this explanation can be augmented by the additional explanation that the cheater has a failed moral character. And if cheating can adequately, if only partly, be explained by reference to the culture within the organization that the cheater worked with, then

this explanation can be augmented by the additional explanation that the cheater has a failed moral character. So these explanations are *at least* compatible. *At least*, for it is possible (and plausible) that these explanations reinforce each other in this way: failed moral characters will tend to make organizational cultures bad, and bad organizational culture will tend to make moral characters fail. Failed moral characters in organizations with a bad culture, will tend to feel at home like fish in water. Applied to Stapel: the explanation of his misconduct can be explained by reference to his failed moral character (to akrasia perhaps), but also by reference to the culture of the organizations with which he worked as bad characters will tend to breed bad cultures, and bad cultures tend to breed bad characters.

As is in the nature of the case, the explanations under the second and third narratives, being Organizational Culture explanations and NPM explanations, are compatible as well. This point can also be made in a more concrete way. Since NPM will foster a particular kind of culture within an organization, and since a particular kind of culture will be especially sensitive to the down-sides of NPM, explanations of misconduct that refer to culture and to NPM are compatible, and they even reinforce each other. Applied to Stapel's case, his misconduct can be explained, partly, by reference to organizational culture, and this can be augmented (and so make for a more complete explanation) by reference to the down-sides of NPM—and these two reinforce each other.

Since the first narrative covers four explanations, there are six pairs to check for compatibility. The *first* pair we consider is Rational Choice explanations and Bad Apple explanations. We may feel pulled in two directions here. Suppose someone is a bad apple, i.e. displays a defective moral character (perhaps the person suffers from akrasia), then we may think that his choice can never be rational, because his defective moral character prevents him from making such a choice. On the other hand, if making a rational choice consists of weighing the costs and benefits of an action as compared to alternative actions, then it would seem that someone with a defective moral character can engage in rational choice making as well—even if the outcome of the calculation is not what we would like it to be. Since it is *formal* ("means-end") rationality that rational choice theory works with, it seems that a rational choice explanation is compatible with a bad apple explanation of the same behaviour. Applied to the Stapel case: an explanation of his misconduct in terms of character flaws (like akrasia) is compatible with the claim that his choice to cheat was the outcome of a rational cost-benefit analysis.

The second pair of explanations we consider is Rational Choice and General Strain. This pair puts before us the question whether strain and stress prevent a person from making a rational choice. On the face of it, stress and strain may lead a person to select a goal that he wouldn't have selected in the absence of it; and given the goal, he may have calculated the means to attain it. Alternatively, a person may have set himself a goal, while stress and strain influence the calculation of the means to attain it. The influence may be that certain means become live options that were dead, or that options that were

alive, die. But given the options, a stressed person may still make what he thinks is a fair calculation—fair not in a moral but in a formal sense. Either way, the explanations based on Rational Choice and General Strain are compatible. Applied to the Stapel case: stress and strain may have led him to set the goal of achieving high-profile publications, and rational choice deliberation suggested to him that fabrication and falsification were the ways to attain that goal. Alternatively, Stapel had set himself the goal of achieving high-profile publications, and strain and stress led him to calculate that fabrication and falsification were the best ways to attain the goal.

Third, Rational Choice and Prospect Theory, by contrast, do not deliver compatible explanations. For the former assumes that an actor will always seek maximal gains based on the probability of occurrence, while the latter says that fear of loss tends to be a much stronger motivator of behavior than the potential for gain, and also that individuals tend toward risk aversion when confronted with potential gains but bias toward risk seeking when confronted with avoiding potential losses.

Applied to Stapel: Rational Choice explains his fraudulent behavior by reference to a rational calculation he has made so as to have maximal gains, while Prospect Theory predicts that, given Stapel's stable job's situation (he had a tenured position with no fear of losing it), he would be less likely to make the risky choices that he did make.

The incompatibility should come as no surprise, as Prospect Theory was expressly developed as an alternative to Rational Choice and to overcome the limitations of the latter (39).

General Strain and Bad Apple approaches are compatible. If stress and strain induce deviant behavior, then it does so in virtuous persons and bad apples alike. Strain explanations and bad apple explanations are compatible, and they may even reinforce each other, in that it is plausible to think that bad apples make even worse choices if they also experience stress and strain—and that strained persons make worse choices if they have flawed moral characters. Applied to Stapel: if he had a flawed moral character, he may already have been open to cheating, but if he was also under stress and strain, then the cheating option may have become even more salient.

Prospect Theory and Bad Apple theory are also compatible. As indicated, Prospect Theory predicts that people faced with the prospect of losing their job or other meaningful resources, will be more inclined to take risks—and if this holds, it holds for bad and good apples alike. The two explanations of behavior it suggests, can both be correct, if only partially. Applied to our case: if we counterfactually assume that Stapel's position was at stake, and also that he had a flawed moral character, then both these factors can be referred to for explanatory purposes—and both explanations can be correct.

The sixth and final pair to consider is that of Prospect and General Strain. Strain and stress may be real in a person who, when faced with serious loss of meaningful resources, is more prone to take risks than when not so faced. Hence, two explanations of a person's behavior based on their own respective theories, can both be true and

hence be compatible. However, if a person is experiencing stress and strain, while at the same time there is no threat of loss of meaningful resources, then the two theories yield incompatible explanations. After all, Prospect Theory tells us that people are risk-aversive. The impulse to deviant behavior generated by strain and stress would be mitigated by the impulse to risk-aversion. In that case we might say that the two theories are compatible—but that the explanations don't reinforce each other, nor do they add up, but rather the one weakens the other in the sense that the effect that one theory predicts doesn't occur to the degree it would have in the absence of the other effect. If we again assume that Stapel was experiencing stress and strain (which already motivated him towards deviant behavior) and he was also facing the threat of losing meaningful resources (which inclined him to take more risks than he would otherwise have taken), then the explanations reinforce each other. But if he was experiencing stress and strain, yet there was no fear of losing meaningful resources, then the theories lead us to expect deviant behavior to a lesser degree than if there was also a threat of loss.

We have discussed, then, six pairs of explanations under the first narrative, and argued that the only incompatible pair is the Rational Choice/Prospect Theory pair. But what about triplets? Which triplets of explanations under the first narrative can form a coherent whole? The answer is: only those triplets that don't contain the incompatible pair. This leaves us with the two remaining triplets, viz. Rational Choice/ Bad Apple/ General Strain, and also Bad Apple/ General Strain/ Prospect Theory.

5. Concluding remarks

We have discussed six explanations of research misconduct, and how they relate to each other. We argued that most theories are compatible with each other, with the exception of Rational Choice and Prospect. Suppose now we concentrate on explanations that are compatible. Can we conclude that those pairs offer *full* explanations? For a number of reasons we cannot. First, we have only looked at pairs among the six theories we have discussed. But triplets of them may offer fuller explanations, and quartets of them even fuller. Second, there are explanations of research misconduct that we haven't discussed, but that can be added to the fold.[13] Third, a large body of research that investigates research misconduct takes the form of correlation 'theories' that map significant correlations between (some measure of) research misconduct and some other factor of interest. Of course, correlation does not equal causation. Take this one step further: on

13 See for example Rajah-Kanagasabai & Roberts (49) that use the theory of planned behavior to explain research misconduct in students. Because our review focused on misconduct among academic researchers, and it had not been coined outside the realm of students, we chose to not review the theory of planned behavior in-depth here. Alternatively, Hackett (50) reviewed anomie as a possible explanation for researchers engaging in research misconduct, but he disregarded anomie so persuasively that we chose not review it here.

a narrow reading of theory – "an idea that is used to account for a situation" – it seems incorrect to speak of correlation *theories*. Correlations map temporal co-occurrence beyond some degree of doubt. The idea or link that is to explain this co-occurrence is often thought up post-hoc as a rationalization, but it is not (yet) a fully-fledged theory.

However, that does not render correlational research meaningless for explaining research misconduct. Similar to narratives, correlational research results deliver cognitive goods— they give knowledge about factors that in some way play into the misconduct. Along that same line of reasoning, they serve as a pointer for further theorizing that may at some point be formalized into a theory.

Still, we are left with the question whether it is sensible to suppose that, drawing on all correlational research and supplemented with the types of theories we reviewed here, one can *fully* explain research misconduct. There seem two avenues to take, both reconcilable with what we argued above, and these avenues are connected to one's stance on free will. Either one believes that humans are free and this will render some part of their behavior – especially complex behaviors, like research misconduct – inexplicable. Or one believes that humans are not free and that scholars have not yet found the (final) key to the explanatory puzzle. It seems natural to think that this key, if it exists, is to be found somewhere along the lines of unconscious factors that influence human behavior, such as biases or heuristics. We tend to the first view.

A further point we would like to make is that although this paper is focused on theories coined to explain falsification and fabrication, these theories also seem relevant when explaining lesser trespasses, such as QRPs. In fact, for those QRPs that teeter on the edge of falsification – take p-hacking or HARKing (hypothesizing after results are known) – it seems natural to suspect that when we apply the theories reviewed here to explain the occurrence of those QRPs, we likely run into similar problems that we encountered when trying to explain research misconduct. And since explanations of research misbehavior – here encompassing both FFP and QRPs – feed into our ideas about prevention of research misbehavior, extending our theories and models on how to explain may help us to prevent.

A further point we would like to make is that although various theories have been used to explain research misbehavior of individual scientists, our discussion brought to light that in order for such explanations to have some minimal level of plausibility, we need to know quite a bit about the personal situation of the researcher, as well as her contextual affordances at an institutional level. The suggestion of our paper is that such knowledge is not easily obtained.

Our final point concerns the role of the Stapel case in our discussion. It should be clear that we have not tried to offer the fullest possible explanation of his fraudulent behavior. We have used Stapel merely to illustrate the kinds and amounts of facts that should be known if an explanation of research misconduct, based on any of the six theories discussed in this paper, is to have minimal plausibility.

Acknowledgments

TH and RvW would like to thank the reviewers and the members of the Theoretical Philosophy reading group for their critical yet constructive comments on earlier versions of this chapter.

References

1. Gunsalus CK. Make reports of research misconduct public. Nature. 2019;570(7).
2. Lafollette MC. The evolution of the "scientific misconduct" issue: An historical overview. Proc Soc Exp Biol Med. 2000;224(4):211–5.
3. Steneck N. Fostering integrity in research: Definition, current knowlege, and future directions. Sci Eng Ethics. 2006;12(1):53–74.
4. ALLEA (All European Academies). The European code of conduct for research integrity. 2017.
5. Netherlands Code of Conduct for Research Integrity. 2018.
6. Martinson BC, Anderson MS, de Vries R. Scientists behaving badly. Nature. 2005;435(7043):737–8.
7. Fanelli D. How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. PLoS One. 2009;4(5):e5738.
8. Fiedler K, Schwarz N. Questionable research practices revisited. Soc Psychol Personal Sci. 2016;7(1):45–52.
9. Haven TL, Tijdink JK, Pasman HR, Widdershoven G, Riet G, Bouter LM. Researchers' perceptions of research misbehaviours : a mixed methods study among academic researchers in Amsterdam. Res Integr Peer Rev. 2019;4(25):1–12.
10. Bouter LM, Tijdink J, Axelsen N, Martinson BC, ter Riet G. Ranking major and minor research misbehaviors: results from a survey among participants of four World Conferences on Research Integrity. Res Integr Peer Rev. 2016;1(17):1–8.
11. Maggio L, Dong T, Driessen E, Artino A. Factors associated with scientific misconduct and questionable research practices in health professions education. Perspect Med Educ. 2019;8(2):74–82.
12. Faria R. What Is Research Misconduct? In: Research Misconduct as White-Collar Crime: A Criminological Approach. Cham: Springer International Publishing; 2018. p. 39–70.
13. Abma R. De publicatiefabriek. Over de betekenis van de affaire-Stapel. Nijmegen: Van Tilt Uitgeverij; 2013. 183 p.
14. Zwart H. Tales of research misconduct : a Lancanian diagnostics of integrity challenges in science novels. Cham, Switzerland: Springer; 2017. p. 1–254. (Library of ethics and applied philosophy ; v. 36).
15. Vogel G. Psychologist accused of fraud on "astonishing scale." Science (80-). 2011;334(6056):579.
16. Levelt Committee, Noort Committee, Drenth Committee. Flawed science: The fraudulent research practices of social psychologist Diederik Stapel. 2012.
17. Woodward J. Making Things Happen: A Theory of Causal Explanation. Oxford University Press; 2003.
18. Gilovich T. How We Know What Isn't so: The Fallibility of Human Reason in Everyday Life. Free Press; 1991.
19. Kahneman D. Thinking, Fast and Slow. London: Penguin Books; 2011.
20. Sovacool BK. Exploring scientific misconduct: Isolated individuals, impure institutions, or an inevitable idiom of modern science? J Bioeth Inq. 2008;5(4):271–82.
21. Lipton P. Understanding without explanation. In: de Regt HW, Leonelli S, Eigner K, editors. Scientific Understanding: Philosophical Perspectives. University of Pittsburgh Press; 2009.

22. Wible JR. Fraud in Science: An economic approach. Philos Soc Sci. 1992;22(1):5–27.
23. Lacetera N, Zirulia L. The economics of scientific misconduct. J Law, Econ Organ. 2011;27(3):568–603.
24. Hren D, Vujaklija A, Ivanišević R, Knežević J, Marušić M, Marušić A. Students' moral reasoning, Machiavellianism and socially desirable responding: Implications for teaching ethics and research integrity. Med Educ. 2006;40(3):269–77.
25. Tijdink JK, Bouter LM, Veldkamp CLS, Van De Ven PM, Wicherts JM, Smulders YM. Personality traits are associated with research misbehavior in Dutch scientists: A cross-sectional study. PLoS One. 2016;11(9):1–12.
26. Stapel D. Faking Science: A True Story of Academic Fraud. Brown NJL, editor. 2014.
27. Agnew R. Foundation for a general strain theory of crime and delinquency. Criminology. 1992;30:47–87.
28. Martinson BC, Crain LA, De Vries R, Anderson MS. The importance of organizational justice in ensuring research integrity. J Empir Res Hum Res Ethics. 2010;5(3):67–83.
29. Medicine), NASEM (National Academies of Sciences, Engineering A. Fostering Integrity in Research. Washington, D.C.; 2017.
30. Holtfreter K, Reisig MD, Pratt TC, Mays RD. The perceived causes of research misconduct among faculty members in the natural, social, and applied sciences. Stud High Educ. 2019;1–13.
31. Kahneman D, Tversky A. Prospect Theory: An Analysis of Decisions Under Risk. Econometrica. 1979;47(2):263–91.
32. Kahneman D. Maps of bounded rationality: Psychology for behavioral economics. Am Econ Rev. 2003;93(5):1449–75.
33. De Vries R, Anderson MS, Martinson BC. Normal misbehavior: scientists talk about the ethics of research. J Empir Res Hum Res Ethics. 2006;1(1):43–50.
34. Crain LA, Martinson BC, Thrush CR. Relationships between the Survey of Organizational Research Climate (SORC) and self-reported research practices. Sci Eng Ethics. 2013;19(3):835–50.
35. IOM. Integrity in scientific research: Creating an environment that promotes responsible conduct. Washington, D.C.: National Academy of Sciences; 2002.
36. Halffman W, Radder H. The Academic Manifesto: From an Occupied to a Public University. Minerva. 2015;53(2):165–87.
37. Overman S, Akkerman A, Torenvlied R. Targets for honesty: How performance indicators shape integrity in Dutch higher education. Public Adm. 2016;94(4):1140–54.
38. Teelken C. Hybridity, coping mechanisms, and academic performance management: Comparing three countries. Public Adm. 2015;93(2):307–23.
39. Thomas KJ, T.A L. Rational Choice and Prospect Theory. In: Bruinsma G, Weisburd D, editors. Encyclopedia of Criminology and Criminal Justice. New York: Springer; 2014.
40. Dancy J. Practical Reality. Oxford University Press; 2000.
41. Anscombe GEM. The Causation of Action. In: Geach M, Gormally L, editors. Human life, action and ethics: essays by GEM Anscombe. Andrews UK; 2011.
42. Owens D. Levels of explanation. Mind. 1989;98:57–79.

43. Lipton P. Inference to the best explanation. 2nd ed. London and New York: Routledge; 2008.
44. De Graaf G. Causes of corruption: towards a contextual theory of corruption. Public Adm Q. 2007;31:39–86.
45. Cassam Q. Vices of the mind. Oxford: Oxford University Press; 1992.
46. Schneider B, Ehrhart MG, Macey WH. Organizational climate and culture. Annu Rev Psychol. 2013;64(1):361–88.
47. Sellin T. Culture and Conflict. New York: Social Science Research Council; 1983.
48. Davis MS. The role of culture in research misconduct. Account Res. 2003;10(3):189–201.
49. Rajah-Kanagasabai CJ, Roberts LD. Predicting self-reported research misconduct and questionable research practices in university students using an augmented Theory of Planned Behavior. Front Psychol. 2015;6:535.
50. Hackett EJ. A Social Control Perspective on Scientific Misconduct Author. J Higher Educ. 1994;65(3):242–60.


3

Chapter 3

Perceptions of research integrity climate differ between academic ranks and disciplinary fields: Results from a survey among academic researchers in Amsterdam

This chapter has been published as:
Haven TL, Tijdink JK, Martinson BC, Bouter LM. Perceptions of research integrity climate differ between academic ranks and disciplinary fields: Results from a survey among academic researchers in Amsterdam. *PLoS One. 2019;14(1):e0210599. https://doi.org/10.1371/journal.pone.0210599*

Abstract

Breaches of research integrity have shocked the academic community. Initially explanations were sought at the level of individual researchers but over time increased recognition emerged of the important role that the research integrity climate may play in influencing researchers' (mis)behavior. In this study we aim to assess whether researchers from different academic ranks and disciplinary fields experience the research integrity climate differently. We sent an online questionnaire to academic researchers in Amsterdam using the Survey of Organizational Research Climate. Bonferroni corrected mean differences showed that junior researchers (PhD students, postdocs and assistant professors) perceive the research integrity climate more negatively than senior researchers (associate and full professors). Junior researchers note that their supervisors are less committed to talk about key research integrity principles compared to senior researchers (*MD* = -.39, *CI* = -.55, -.24). PhD students perceive more competition and suspicion among colleagues (*MD* = -.19, *CI* = -.35, -.05) than associate and full professors. We found that researchers from the natural sciences overall express a more positive perception of the research integrity climate. Researchers from social sciences as well as from the humanities perceive less fairness of their departments' expectations in terms of publishing and acquiring funding compared to natural sciences and biomedical sciences (*MD* = -.44, *CI* = -.74, -.15; *MD* = -.36, *CI* = -.61, -.11). Results suggest that department leaders in the humanities and social sciences should do more to set fairer expectations for their researchers and that senior scientists should ensure junior researchers are socialized into research integrity practices and foster a climate in their group where suspicion among colleagues has no place.

Keywords:

Research integrity, research climate, academic ranks, disciplinary fields, responsible conduct of research

Introduction

Recent breaches of research integrity in The Netherlands and worldwide have shocked the academic community (1–4). Such events led to a new field of inquiry that aimed to better understand how common the problems are and what drives researchers to misbehave (5–7). Initially, studies in this area mainly focused on research misconduct, in which there is generally an intent to deceive (fabrication, falsification, plagiarism). However, over time the focus broadened to the more frequent questionable research practices (QRPs). Accumulating empirical evidence has indicated QRPs are much more prevalent than formal research misconduct (8–10). Consequently, QRPs probably have on the aggregated level more impact. Initially, explanations for research misconduct were sought at the level of individual researchers (11) but over time increased recognition emerged of the important role that structural and institutional factors such as research climate may play in influencing researchers' behavior (12–16). This has shifted the focus to the organizational climate in research settings as a potential target for intervention (17,18).

Studying organizational climates implies investigating the environment researchers work in and how this climate can strengthen or erode research integrity (19,20). The organizational climate here is defined as "the shared meaning organizational members attach to the events, policies, practices, and procedures they experience and the behaviors they see being rewarded, supported, and expected." (p. 115) (21,22). Crain et al. (23) have documented that a favorable organizational research climate is positively associated with lower levels of self-reported questionable research practices. The Survey of Organizational Research Climate (henceforth: SOuRCe©) is designed to measure the organizational research integrity climate in academic research settings (18,20,22,24)

The SOuRCe© is embedded in two conceptual frameworks, the first being organizational justice theory (25). In a nutshell: the fairer people regard decisions and decision-making processes in their organization, the more likely they trust their organization, abide by decisions made and do not engage in questionable behavior (26,27). When people perceive procedural or distributional injustice in their organization, they are more likely to behave in ways that, in their mind, compensates for the perceived unfairness (27). Applied to research integrity, in a research climate where perceived injustice is high, researchers would be expected to be more likely to engage in intentional research misconduct (falsification, fabrication and plagiarism) or questionable research practices (27).

The second conceptual framework underpinning the SOuRCe© stems from the Institute of Medicine report *Integrity in Scientific Research: Creating an Environment That Promotes Responsible Conduct* (28). This report describes the research environment as an open systems model where different factors influence research integrity. The report specifies that the research integrity climate can both stimulate or diminish responsible research (18,28,29). Some key factors herein that are reflected in the SOuRCe are ethical

leadership, integrity policy familiarization and communication, and the degree to which these are known by people in the organization (18,28).

Previous research with the SOuRCe© found that researchers in different phases of their career perceive the research integrity climate differently (22). PhD students perceived the climate to be fairer compared to senior scientists in that scholarly integrity was valued (e.g. acknowledging work of others). Senior scientists perceived there to be more resources for conducting research responsibly (e.g. policies to deal with integrity breaches were well known) (22).

Wells et al. (22) also found large differences in SOuRCe© scores for different organizational subunits. Some had scores twice as negative compared to others or compared to overall mean scores. This indicates that overall high mean scores on an institutional level offer departmental leaders little comfort (20) and research climate may vary significantly within institutions. One factor that accounts for these stark differences between subunits was disciplinary field (22).

Our study aims to determine how scientists experience the research integrity climate, stratified for academic rank and disciplinary field, in two university medical centers and two universities in Amsterdam. This is the first study that investigates research integrity climate in The Netherlands. Assessing research integrity climate will provide insight what factors may hinder responsible research practices (26).

We hypothesized that we would observe significant variability in SOuRCe© scale-scores based on (1) the disciplinary field in which academic researchers work and (2) the academic ranks of respondents. As our aim is descriptive in nature, we did not specify the direction of these differences.

Materials and methods

Ethical considerations

The Scientific and Ethical Review board of the Faculty of Behavior & Movement Sciences (Vrije Universiteit Amsterdam) approved our study (Approval Number: VCWE-2017-017R1).

Participant selection and procedure

The institutions that participated in our study included two universities (Vrije Universiteit Amsterdam and University of Amsterdam) and two academic medical centers (Amsterdam Medical Centers). Upon securing endorsement from the deans and rectors of the participating institutions and finalizing a data sharing agreement, each institution provided a list of e-mail addresses of all researchers and PhD students. We distributed the electronic survey in May 2017 via email among all academic researchers. Researchers were eligible to participate if they were doing research at least one day per week (>0.2fte)

on average. Our cross-sectional online survey contained three instruments (SOuRCe©, the Publication Pressure Questionnaire (30) and a list of 60 major and minor research misbehaviors (9)). This article presents the SOuRCe© results. The survey concluded with three demographic items about gender, academic rank and disciplinary field.

We used the online survey program Qualtrics (Qualtrics, Provo, UT, USA) to create and distribute the survey. Researchers first received an information e-mail explaining the purpose, goal and procedure of the study. After one week, we sent the official invitation with a unique link to the survey and a link to the non-response survey (see S1 appendix). The invitation also included a link to our privacy policy and the protocol (see S2 appendix and S3 protocol), both available on the project's website (www.amsterdamresearchclimate.com). The survey started with an online informed consent form. After consenting, participants were asked to indicate whether they were doing research for at least one day per week (inclusion criterion). We sent three reminders to those who had not responded yet. All correspondence explicitly stated that the data would remain confidential and that participation was voluntary.

Instruments

We used the Survey of Organizational Research Climate (22–24,31). The SOuRCe© evaluates what factors play a role in the perceived research climate on a scale that ranges from 1 ("not at all") to 5 ("completely") (18,24). It consists of 28 items forming 7 subscales that detail the organizational climate of integrity on a departmental and institutional level (29). For an overview of the SOuRCe© subscales, see Table 1.

Table 1. Overview of SOuRCe subscales.

Subscale	Level	# items	Description constructs measured
RCR Resources	Institutional	6	degree to which respondents perceive the existence of effective educational opportunities about RCR, available policies and professionals to whom concerns can be addressed, and leaders who actively support RCR
Regulatory Quality	Institutional	3	factors such as the degree to which regulatory committees such as the Medical Ethical Testing Committee treat researchers fairly.
Integrity Norms	Departmental	4	degree to which norms about research integrity exist in one's department.
Integrity Socialization	Departmental	4	degree to which organizational departments engage in activities that effectively socialize junior researchers in research integrity.
Supervisor/ Supervisee Relations	Departmental	3	relations between supervisors and their supervisees in terms of fairness, availability and respect
(Lack of) Integrity inhibitors*	Departmental	6	degree to which conditions like lack of adequate resources or suspicion and competition between colleagues produce difficulties for conducting research responsibly.
Expectations	Departmental	2	degree to which the department's expectations for publishing and obtaining external funding are fair

Columns stipulate level of measurement, number of items per subscale and a description of the constructs that subscale measures. *This scale was reversely scored so that all subscales can be interpreted using the same logic (i.e. a higher score means a greater lack of inhibiting factors, which indicates a better research integrity climate).

SOuRCe© subscale scores are calculated by taking the average of all valid non-missing items in that subscale. The respondent needs to validly answer at least half of the items in the subscale for the subscale score to be valid. Valid scores are all response options except for "No basis for judging OR not relevant to my field of work". All subscale scores can be interpreted using the same logic: the higher the score, the stronger the presence of that factor. Higher scores thus express a more favorable perception of the research integrity climate. While most SOuRCe© items ask about the perceived presence of integrity supporting aspects of the local climate, the Integrity Inhibitors scale is comprised of items that ask about the perceived presence of factors that may inhibit research integrity. For analysis and reporting, the items contributing to this scale are reverse-coded so that the higher this subscale's score; the greater the ***lack*** of integrity inhibiting conditions (29).

The SOuRCe© was designed for a biomedical research setting. To make the items more applicable to all disciplinary fields from our study population, we slightly altered the wording of three items in consultation with the design team of the SOuRCe© (see S1 Table). We also extended the response option: "No basis for judging" to "No basis for judging OR not relevant to my field of work". Unfortunately, two of the original 28 SOuRCe© items were inadvertently omitted from the final distribution of the questionnaire due to a programming error.

Statistical analyses

The intended statistical analyses were preregistered under the title 'Academic Research Climate Amsterdam' at the Open Science Framework. Briefly, for the univariate analyses, we computed overall mean subscale scores and stratified scores per academic rank and disciplinary field. For those subscales where academic rank or disciplinary field was significantly associated, we tested whether stratified scores differed significantly using post hoc Bonferroni corrected *F* tests. We then created association models with academic rank or disciplinary field as independent variable and subscale score as dependent variable. For the multivariate analyses, we corrected for potential confounders (e.g. gender) or added effect modifiers when inspecting the relations between disciplinary field or academic rank and the SOuRCe© subscales.

Results

We collected 7548 e-mail addresses from academic researchers in Amsterdam. When we sent out the information letter, 83 bounced immediately as undeliverable. Also, 109 researchers decided not to participate and asked Qualtrics to be unsubscribed

from the data base. 2274 researchers opened the questionnaire (30%). Of those who opened the questionnaire, 1298 (17% of the total sample) researchers answered enough questions to complete at least one SOuRCe© subscale (57% of those who opened the questionnaire). See Fig 1. Only 2% filled in the ultra-brief non-response questionnaire.

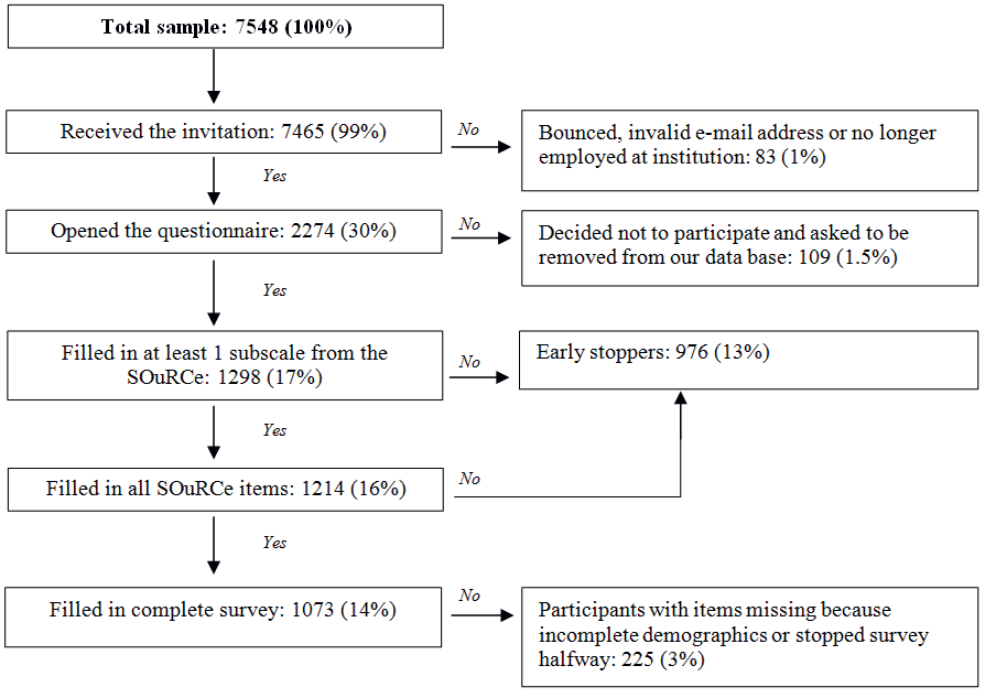

Fig 1. Flow diagram of response and completion rate.

Percentages are expressed in reference to the total population of academic researchers in Amsterdam (n = 7548).

Differences between academic ranks

Overall mean subscale scores of the total sample are given in Figs 2 and 3 as a general reference to our stratified results. Investigating our first hypothesis (differences between academic ranks), for those subscales that were significantly associated with rank (Integrity Norms, Integrity Socialization, Integrity Inhibitors, Supervisor-Supervisee relations, Expectations and RCR Resources, respectively), we ran post-hoc Bonferroni corrected F tests. The purpose was to see whether PhD students, postdocs & assistant professors or associate & full professors perceived the climate differently on these subscale (see Table 2). PhDs students as well as postdocs and assistant professors scored significantly lower than associate and full professors on 4 subscales (Expectations, Supervisor-Supervisee relations, Integrity Socialization and RCR Resources, respectively). PhDs students (M

= 3.73) also scored significantly lower than associate and full professors (*M* = 3.92) on Integrity Inhibitors. Postdocs and assistant professors (*M* = 3.67) scored significantly lower on Integrity Norms than did associate and full professors (*M* = 3.82) See Fig 2. Finally, we tested whether the relation between academic rank and the SOuRCe© subscale scores was confounded or modified by other independent variables (i.e. gender or disciplinary field). Expectations and Integrity Norms were confounded by gender, respectively. Adding gender to these models made the associations between academic rank and SOuRCE© subscale scores slightly weaker but the effect remained significant. We found effect modification by gender on RCR Resources only, these stratified results are given in Table 3. Therefore, Fig 2 and Table 2 display statistics corrected for confounding or reporting effect modification if applicable. We have calculated the effect sizes for the significant differences, see Table 4.

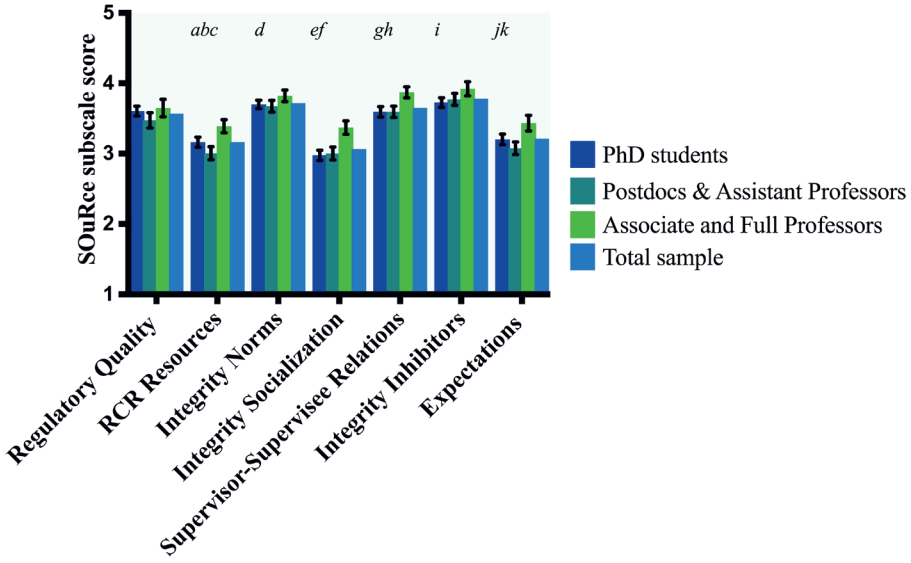

Fig 2. Differences between academic ranks.

Gender adjusted (if applicable) and Bonferroni corrected mean differences (*MD*) between pairs of academic ranks on SOuRCe subscale scores with 95% confidence intervals (*CI*). Letters indicate significant differences at the $\alpha = 0.05$ level.

a: PhD students scored lower on RCR Resources than associate & full professors (*MD* = -.23, *CI* = -.39, -.07)

b: Postdocs & assistant professors scored lower on RCR Resources than PhD students (*MD* = -.16, *CI* = -.3, -.02)

c: Postdocs & assistant professors scored lower on RCR Resources than associate & full professors (*MD* = -.39, *CI* = -.56, -.21)

d: Postdocs and assistant professors scored lower on Integrity Norms than associate & full professors (*MD* = -.15, *CI* = -.29, -.00)

e: PhD students scored lower on Integrity Socialization than associate & full professors *(MD = -.39, CI* = -.55, -.24*)*

f: Postdocs & assistant professors scored lower on Integrity Socialization than associate & full professors (*MD* = -.37, *CI* = -.54, .20)

g: PhD students scored lower on Supervisor-Supervisee Relations than associate & full professors (*MD* = -.29, *CI* = -.43, -.14)
h: Postdocs & assistant professors scored lower on Supervisor-Supervisee Relations than associate & full professors (*MD* = .28, *CI* = -.44, -.11)
i: PhD students scored lower on Integrity Inhibitors than associate & full professors (*MD* = -.19, *CI* = -.35, -.05)
j: PhD students scored lower on Expectations than associate & full professors (*MD* = -.23, *CI* = -.39, -.07)
k: Postdocs & assistant professors scored lower on Expectations than associate & full professors (*MD* = -.36, *CI* = -.53, -.18)

Table 2. Regression models of SOuRCe subscales by academic rank.

Academic rank*	**PhD students** (*N* = 481)		**Postdocs & assistant professors** (*N* = 294)		**Associate and full professors^** (*N* = 210)	
Scale *F* (*p*, *df*)	**Beta (SE)**	**(CI)**	**Beta (SE)**	**(CI)**	**Beta (SE)**	**(CI)**
Integrity Norms 3.21 (0.041, 2)	-.108 (.058)	(-.221, -.005)	-.138 (.062)	(-.259, -.016)	-	-
RCR Resources 14.043 (<.001, 2)	-.386 (.102)	(-.586, -.186)	-.476 (.110)	(-.691, -.260)	-	-
Integrity Inhibitors 4.908 (.008, 2)	-.195 (.063)	(-.317, -.073)	-.150 (.068)	(-.283, -.017)	-	-
Integrity Socialization 19.584 (<.001, 2)	-.394 (.065)	(-.522, -.266)	-.368 (.071)	(-.507, -.228)	-	-
Supervisor-Supervisee Relations 11.552 (<.001, 2)	-.284 (.062)	(-.405, -.162)	-.278 (.068)	(-.411, -.144)	-	-
Expectations 11.772 (<.001, 2)	-.202 (.070)	(-.340, -.064)	-.335 (.075)	(-.482, -.189)	-	-

Regression coefficients (Beta), standard errors (SE) and confidence intervals (CI). F-tests (*F*) between groups are given in the left column associated *p*-value and degrees of freedom (*df*).
* 313 respondents did not disclose their academic rank.
^ =Reference group.

Table 3. Stratified scores on RCR Resources for academic rank and gender.

Academic rank	**Male**	**Female**
PhD student	3.29	3.10
Postdoc and Assistant Professor	2.99	3.01
Associate and Full Professor	3.33	3.50

Table 4. Overview of effect sizes of significant differences (p<.05).

Subscale SOuRCe	Group vs. Group	Effect size*	Interpretation**
RCR Resources	PhD students < Associate & full professors	.29	Small
RCR Resources	Postdocs & assistant professors < PhD students	.19	Small
RCR Resources	Postdocs & assistant professors < Associate & full professors	.47	Small
Integrity Norms	Postdocs & assistant professors < Associate & full professors	.22	Small
Integrity Socialization	PhD students < Associate & full professors	.18	Small
Integrity Socialization	Postdocs & assistant professors < Associate & full professors	.87	Large
Supervisor-Supervisee Relations	PhD students < Associate & full professors	.36	Small
Supervisor-Supervisee Relations	Postdocs & assistant professors < Associate & full professors	.43	Small
Integrity Inhibitors	PhD students < Associate & full professors	.25	Small
Expectations	PhD students < Associate & full professors	.27	Small
Expectations	Postdocs & assistant professors < Associate & full professors	.43	Small
Regulatory Quality	Humanities < Biomedical sciences	.50	Medium
Regulatory Quality	Humanities < Social sciences	.42	Small
Expectations	Social sciences < Biomedical sciences	.29	Small
Expectations	Social sciences < Natural sciences	.42	Small
Expectations	Humanities < Biomedical sciences	.39	Small
Expectations	Humanities < Natural sciences	.55	Medium

* based on Hedges' G that is calculated as: $M1 - M2 / SDpooled$

** Interpreted based on Cohen (32), an effect size of .20 is small, .50 is medium, .80 is large and 1.30 is very large.

Differences between disciplinary fields

Regarding our second hypothesis (differences between disciplinary fields), disciplinary field was associated with Regulatory Quality and Expectations, see Table 5. Humanities scored significantly lower on Regulatory Quality than biomedicine. Social Sciences (M = 3.05) as well as humanities (M = 2.97) score significantly lower on Departmental Expectations than both natural sciences (M = 3.41) and biomedicine (M = 3.29). See Table 5 and Fig 3. The associations between disciplinary field and both

Regulatory Quality as well as Expectations were confounded by rank, yet again the main effect of discipline remained significant. Therefore, Fig 3 and Table 5 display statistics corrected for confounding. We have calculated the effect sizes of each difference, see Table 4.

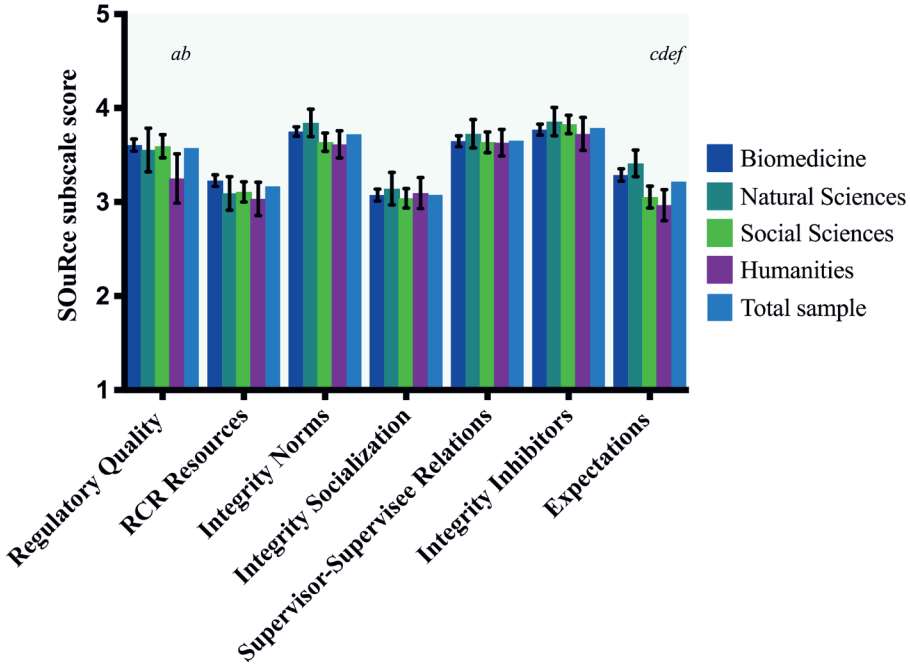

Figure 3. Differences between disciplinary fields.

Rank adjusted (if applicable) and Bonferroni corrected mean differences (*MD*) between pairs of disciplinary fields on SOuRCe subscale scores with 95% confidence intervals (*CI*). Letters indicate significant differences at the $\alpha = 0.05$ level.
a: Humanities scored lower on Regulatory Quality than social sciences (*MD* = -.34, *CI* = -.66, -.01)
b: Humanities scored lower on Regulatory Quality than biomedical sciences (*MD* = -.38, *CI* = -.68, -.08)
c: Social sciences scored lower on Expectations than biomedical sciences (*MD* = .26, *CI* = -.44, -.09)
d: Social sciences scored lower on Expectations than natural sciences (*MD* = -.38, *CI* = -.61, -.11)
e: Humanities scored lower on Expectations than biomedical sciences (*MD* = -.34, *CI* = .57, -.10)
f: Humanities scored lower on Expectations than natural sciences (*MD* = -.45, *CI* = -.75, -.15)

Table 5. Regression models of SOuRCe subscales by disciplinary field

Disciplinary field*	**Biomedical sciences**		**Natural sciences**		**Social sciences**		**Humanities^**	
	(*N* = 557)		(*N* = 103)		(*N* = 237)		(*N* = 100)	
Scale *F* (*p*, *df*)	**Beta (SE)**	**(CI)**	**Beta (SE)**	**(CI)**	**Beta (SE)**	**(CI)**	**Beta (SE)**	**(CI)**
Regulatory quality	.395	(.174, .617)	.391	(.070, .711)	.395	(.150, .640)	-	-
3.472 (.016, 3)	(.113)		(.164)		(.125)			
Expectations	.366	(.193, .540)	.483	(.261, .705)	.142	(-.052, .337)		
9.709 (<.001, 3)	(.089)		(.113)		(.099)			

Regression coefficients (Beta), standard errors (SE) and confidence intervals (CI). F-tests (*F*) between groups are given in the left column associated *p*-value and degrees of freedom (*df*).
* 281 respondents did not disclose their disciplinary field.
^ =Reference group.

Discussion

We assessed the research integrity climate in Amsterdam using the SOuRCe©. We hypothesized that we would observe significant variability in SOuRCe© scale-scores based on (1) the disciplinary field in which academic researchers work and (2) the academic ranks of respondents. For the sake of brevity, we therefore discuss only the significant differences between academic ranks and disciplinary fields below.

Differences between academic ranks

Departmental Expectations were perceived more negatively by PhD students, postdocs and assistant professors. This could be because their career prospects often directly depend on fulfilling these expectations whereas more senior scientists are less directly dependent on meeting publication and funding requirements for retaining their job (33,34). This result is similar to Martinson et al. (2006) who found mid- and early career scientists to perceive higher amounts of organizational injustice compared to senior scientists as measured by asking scientists about the efforts they put into scientific work and rewards they receive in return (35–37).

We found PhDs as well as postdocs and assistant professors to score lower on Supervisor-Supervisee relations than associate and full professors. Martinson et al. (20) found the opposite effect in their study of researchers within the U.S. Department of Veterans Affairs Healthcare System in which the senior staff perceived this scale to be lower than more junior staff. In contrast, in a study of more traditional academic researchers in the U.S., Wells et al. (22) did not find notable differences on this scale by academic rank. The fact that junior researchers in our sample perceive their supervision as suboptimal could be alarming as poor mentoring is associated with the risk of emotional stress (38,39) and poor mentoring is viewed by some as one of the most impactful research misbehaviors (9).

Contrary to Wells et al. (22) who found U.S. junior researchers to report the highest levels of Integrity socialization, we found PhDs and postdocs to report lower levels of Integrity socialization than professors. Junior researchers are the ones who would have to be 'socialized' into research integrity whereas senior researchers in charge of this socialization process report higher levels. This discrepancy could indicate that senior researchers acknowledge the importance of research integrity when it comes to effective socialization of junior researchers into the department, yet we may conclude that in practice this socialization into research integrity does not get sufficient attention.

Communication about research integrity policies, part of the RCR Resources subscale, from the various bodies in academia is often addressed to the deans, department heads or principal investigators. This could explain why Wells et al. (22) found the same result as we have here: senior researchers score higher on RCR Resources than junior researchers. Being a senior researcher (associate & full professor) in an academic organization inevitably means that research integrity policies created at the top are more likely to land on your desk.

Interestingly, this effect depended on gender: female researchers perceived more RCR Resources, except for PhD students where male researchers perceived more resources to conduct their research responsibly. Perhaps female PhD students are also more likely to express their concern about the availability of resources for responsible conduct of research than their male counterparts. There is some evidence that women value procedural justice, the way in which resources are distributed, more than men do (40) but as no gender interactions in SOuRCe© subscales have been reported, it seems premature to conclude that this applies here.

PhD students perceived the [lack of] Integrity Inhibitors to be lower than did associate and full professors. Mirroring the pattern for this subscale of Wells et al. (22), PhD students perceive a larger presence of such integrity inhibiting conditions (such as suspicion among colleagues or a hostile atmosphere) than more senior researchers. Associate and full professors may have gotten used to inhibitors such as publication pressure and regard these as less of a threat to integrity (22,41).

Finally, postdocs and assistant professors perceive Integrity Norms to be lower than associate and full professors, indicating a more negative attitude towards research integrity. Maybe postdocs and assistant professors witness less responsible research and more QRPs. This again parallels the three U.S. universities findings where postdocs scored lowest on more than half of the SOuRCe© subscales (22). Postdocs and assistant professors perceiving more questionable conduct of research also aligns with studies assessing the frequency of misbehavior, where mid-career scientists admitted to more research misbehaviors than did senior scientists (5).

Differences between disciplinary fields

Similar to Wells et al. (22), we found the humanities to score lowest on both Departmental Expectations and Regulatory Quality. The difference was to be expected as regulatory bodies play a more important role in fields where rules and regulation are pivotal (such as biomedicine). In areas like literature or philosophy, regulatory bodies are less important or non-existent. Hence, researchers from the humanities might score lower because they do not encounter these regulating bodies.

The subscale Departmental Expectation measures the degree to which researchers perceive their department's expectations regarding publishing or obtaining funding as fair. Alike Wells et al. (22) natural sciences score highest and the humanities score lowest. One explanation could be that in areas like philosophy or law the traditional way of disseminating academic work is via books, national or specialist journals. Nowadays when performance is measured the focus is predominantly on publishing in (high-impact) international journals. This can cause dissatisfaction from researchers from the humanities, as their books and national contributions are not valued the same way by their department as other academic products such as journal publications.

Strengths of our study

Ours is the first publicly available study to investigate the research integrity climate in a European country. It is too premature to compare our data to the U.S. studies available, as differences in research integrity climates found could be due to a range of factors (known and unknown) that neither of these studies has measured. Our data can provide a useful baseline measurement so that repeated administration of the SOuRCe© could provide information on developments over time. With this knowledge we can better inform universities about interventions tailored to specific disciplines and ranks. This can be used to create a better climate for research integrity.

Furthermore, the SOuRCe© subscales focus on observable characteristics in the local environment. This means that the SOuRCe© provides direct feedback for academic leaders on what can be improved in the organizational structure for fostering research integrity. For example, we found Integrity Socialization is perceived low by junior researchers. This result might target investigation at the institution to find out how socialization can be boosted, how and what means are necessary to foster embedding of research integrity socialization.

Study limitations

Although our completion rate of 18% is low, it is similar to other online surveys. This does not necessarily indicate response bias (42,43). Response bias could occur when non-responders are dissimilar to responders. We tried to estimate this by asking non-responders to fill in a brief non-response questionnaire, but only 2% of non-responders did which we regard too little to base solid conclusions on. We thus tried to

assess the representativeness of our respondents for the total population by comparing our demographics to publicly available data on researchers in Amsterdam. Because data on researchers in medical centers is not readily available, we decided to filter out all researchers who indicated working in biomedical sciences. When comparing the researchers in our sample from the two universities (excluding all researchers from the two medical centers) to the publicly available data on researchers at the two universities in Amsterdam, it appears that we had a reasonably representative sample taking part from the various ranks: 27% of researchers are full or associate professor (our sample: 21%), 40% are assistant professor or postdoc (our sample 38%) and 32% are PhD-student (our sample: 41%).

However, there may be a gender bias as more than half of the researchers in our sample were female (57% respectively). In the Netherlands as a whole, females only account for 39% of academics (https://www.vsnu.nl/f_c_ontwikkeling_aandeel_vrouwen.html). In Amsterdam, that is 42%. This is most likely accounted for by an overrepresentation of female PhD students in our sample (68% versus the national 45% of PhD students in academia). This could be due to women's greater willingness to participate in surveys (44,45). However, we accounted for this selectivity by correcting for gender where necessary. In the case of RCR Resources, gender modified the results. Hence, we report this effect separately for men and women (see Table 3). To conclude, this selectively of the sample is unlikely to bias our results.

Also, to protect respondents' and institutions' privacy, we decided to only collect personal information about gender, academic rank and disciplinary field. This restricted our ability to obtain institutional-level, department-level and specific field of study level classifications, making it likely that we have missed meaningful variability between institutes or departments within our broad disciplinary categories. This way of collecting our data on relatively large group level only (academic rank and disciplinary field) also makes a more advanced multilevel model infeasible, so results from our multivariate association models (see Table 2 and 5) should be interpreted with caution as the standard errors of observed associations may be under-estimated due to clustering (46). We tried to estimate the impact of clustering using unpublished ICCs from the data used by Wells et al. (22) for institute (they had three participating institutions, we have four). Applying the clustering correction affected the relation with rank and Integrity Norms and Integrity Inhibitors: rank was no longer significantly associated with these three subscales. Other associations with rank remained significant despite the VIF correction, see S2 Table. Disciplinary field remained significantly associated with both Expectations and Regulatory Quality, see S3 Table).

Implications

The core finding that the research integrity climate is perceived differently by juniors and seniors as well as by researchers from different disciplinary fields, stresses the need

for tailored interventions. A one-size-fits all approach to improve the academic research integrity climate will likely not yield the desired effect (23). Interestingly, nowadays more attention is paid to proper research integrity education via means of tutorials, seminars and other courses. This does not align with the low score on Integrity Socialization and RCR resources in our sample. However, integrity is not something someone learns from one course, responsible research has to become a habit, not an exception. There is terrain to win by integrating research integrity into daily practice by taking time to make every new researcher in the department familiar with research integrity. Furthermore, it can help to focus discussions about research integrity on the actual situation in the department: what standard procedures have been implemented to foster responsible research without having to compromise research integrity.

A rather alarming observation in our results is PhD students' perception of integrity inhibiting factors. The novices in academic research already have to cope with suspicion and competition among colleagues. Navigating in a research integrity climate with such challenges asks for thoughtful guidance from senior researchers that sadly seems no to have no priority (9).

In conclusion, the research integrity climate is perceived differently by researchers from different disciplinary fields. Small fields like the humanities perceive their department's expectations as more negative compared to other disciplinary fields. The natural sciences overall seem to perceive the climate more positively.

Associate and full professors perceive a more positive research integrity climate than assistant professors, postdocs and PhD-students. This might be a key for improving the research integrity climate. Senior scientists should ensure that new researchers are socialized into research integrity practices and foster a climate in their group where suspicion among colleagues has no place.

Acknowledgements

We would like to thank all members of the steering committee of the Academic Research Climate Amsterdam project (Frans Oort, Yvo Smulders, Gerben ter Riet, Hanneke de Haes, Guy Widdershoven and René van Woudenberg) for their critical comments during presentations on the study design and/or draft communications regarding the study. Peter van de Ven has been indispensable in discussing the statistical models. We also wish to acknowledge Herm Lamberink for his independent analysis of the survey data.

References

1. Bouter LM. Commentary: Perverse incentives or rotten apples? Account Res. 2015;22(3):148–61.
2. Levelt Committee, Noort Committee, Drenth Committee. Flawed science: The fraudulent research practices of social psychologist Diederik Stapel. 2012.
3. Stroebe W, Postmes T, Spears R. Scientific misconduct and the myth of self-correction in science. Perspect Psychol Sci. 2012;7(6):670–88.
4. Kornfeld DS. Perspective: Research misconduct: The search for a remedy. Acad Med. 2012;87(7):877–82.
5. Martinson BC, Anderson MS, de Vries R. Scientists behaving badly. Nature. 2005;435(7043):737–8.
6. Titus SL, Wells JA, Rhoades LJ. Repairing research integrity. Nature. 2008;453(7198):980–2.
7. Fanelli D. How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. PLoS One. 2009;4(5):e5738.
8. Steneck N. Fostering integrity in research: Definition, current knowlege, and future directions. Sci Eng Ethics. 2006;12(1):53–74.
9. Bouter LM, Tijdink J, Axelsen N, Martinson BC, ter Riet G. Ranking major and minor research misbehaviors: results from a survey among participants of four World Conferences on Research Integrity. Res Integr Peer Rev. 2016;1(17):1–8.
10. De Vries R, Anderson MS, Martinson BC. Normal misbehavior: scientists talk about the ethics of research. J Empir Res Hum Res Ethics. 2006;1(1):43–50.
11. Shaw D. The quest for clarity in research integrity: A conceptual schema. Sci Eng Ethics. 2018;
12. Casadevall A, Ellis LM, Colloquium AAM, Committee S, Davies EW, Manager PE, et al. A framework for improving the quality of research in the biological sciences. MBio. 2016;7(4):e01256-16.
13. Sovacool BK. Exploring scientific misconduct: Isolated individuals, impure institutions, or an inevitable idiom of modern science? J Bioeth Inq. 2008;5(4):271–82.
14. Steneck NH. Institutional and individual responsibilities for integrity in research. Am J Bioeth. 2002;2(4):51–3.
15. Casadevall A, Fang FC. Reforming science: Methodological and cultural reforms. Infect Immun. 2012;80(3):891–6.
16. Alberts B, Kirschner MW, Tilghman S, Varmus H. Rescuing US biomedical research from its systemic flaws. Proc Natl Acad Sci. 2014;111(16):5773–7.
17. Martinson BC, Mohr DC, Charns MP, Nelson D, Hagel- E, Bangerter A, et al. Main Outcomes of an RCT to Pilot Test Reporting and Feedback to Foster Research Integrity Climates in the VA. AJOB Empir Bioeth. 2017;8(3):211–9.
18. Thrush CR, Vander Putten J, Gene Rapp C, Pearson C, Simms Berry K, O'Sullivan P. Content validation of the Organizational Climate for Research Integrity survey (OCRI). J Empir Res Hum Res Ethics. 2007;2(4):35–52.
19. Mumford M, Murphy S, Connelly S, Hill J, Antes A, Brown R, et al. Environmental influences on ethical decision making: climate and environmental predictors of research integrity. Ethics Behav. 2007;17(4):337–66.

20. Martinson BC, Nelson D, Hagel-Campbell E, Mohr D, Charns MP, Bangerter A, et al. Initial results from the Survey of Organizational Research Climates (SOuRCe) in the U.S. department of veterans affairs healthcare system. PLoS One. 2016;11(3).
21. Schneider B, Ehrhart MG, Macey WH. Organizational Climate and Culture. Annu Rev Psychol. 2013;64(1):361–88.
22. Wells JA, Thrush CR, Martinson BC, May TA, Stickler M, Callahan EC, et al. Survey of organizational research climates in three research intensive, doctoral granting universities. J Empir Res Hum Res Ethics. 2014;9(5):72–88.
23. Crain LA, Martinson BC, Thrush CR, Crain AL, Martinson BC, Thrush CR. Relationships Between the Survey of Organizational Research Climate (SORC) and Self-Reported Research Practices. Sci Eng Ethics. 2013;19(3):835–50.
24. Martinson BC, Thrush CR, Crain AL. Development and validation of the Survey of Organizational Research Climate (SORC). Sci Eng Ethics. 2013;19(3):813–34.
25. Greenberg J. Allocator-recipient similarity and the equitable division of rewards. Soc Psychol (Gott). 1978;41(4):337–41.
26. Martinson BC, Anderson MS, Crain AL, De Vries R. Scientists' perceptions of organizational justice and self-reported misbehaviors. J Empir Res Hum Res Ethics. 2006;1(1):51–66.
27. Martinson BC, Crain LA, De Vries R, Anderson MS. The importance of organizational justice in ensuring research integrity. J Empir Res Hum Res Ethics. 2010;5(3):67–83.
28. Institute of Medicine. Integrity in scientific research: Creating an environment that promotes responsible conduct. Washington, D.C.: National Academy of Sciences; 2002.
29. Thrush CR, Martinson BC, Crain AL, Wells JA. User's Manual for the Survey of Organizational Research Climate (SOURCE). 2014.
30. Tijdink J, Smulders Y, Vergouwen A, de Vet H, Knol DL. The assessment of publication pressure in medical science; validity and reliability of a Publication Pressure Questionnaire (PPQ). Qual life Res An Int J Qual life Asp Treat care Rehabil. 2014;23(7):2055–62.
31. Martinson BC, Nelson D, Hagel-Campbell E, Mohr D, Charns MP, Bangerter A, et al. Initial results from the Survey of Organizational Research Climates (SOuRCe) in the U.S. department of veterans affairs healthcare system. PLoS One. 2016;11(3):1–18.
32. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale, N.J.: Erlbaum; 1995.
33. Van Dalen HP, Henkens K. Intended and unintended consequences of a publish-or-perish culture: A worldwide survey. J Am Soc Inf Sci Technol. 2012;63(7):1282–1293.
34. Mentzelopoulos SD, Zakynthinos SG. Research integrity, academic promotion, and attribution of authorship and nonauthor contributions. JAMA - J Am Med Assoc. 2017;318(13):1221–2.
35. Peter R, Alfredsson L, Hammar N, Siegrist J, Theorell T, Westerholm P, et al. High effort, low reward, and cardiovascular risk factors in employed Swedish men and women : baseline results from the WOLF Study. J Epidemiol Community Heal. 1998;52:540–7.
36. Siegrist J. Adverse health effects of high-effort/low-reward conditions. J Occup Health Psychol. 1996;1(1):27–41.

37. Siegrist. Effort-reward imbalance at work and health. Hist Curr Perspect Stress Heal. 2015;2:261–91.
38. Peluso DL, Carleton RN, Asmundson GJG. Depression symptoms in Canadian psychology graduate students: Do research productivity, funding, and the academic advisory relationship play a role? Can J Behav Sci. 2011;43(2):119–27.
39. Levecque K, Anseel F, De Beuckelaer A, Van der Heyden J, Gisle L. Work organization and mental health problems in PhD students. Res Policy. 2017;46(4):868–79.
40. Sweeney PD, Mcfarlin DB. Process and outcome: Gender differences in the assessment of justice. J Organ Behav. 1997;18(1):83–98.
41. Tijdink JK, Schipper K, Bouter LM, Pont PM, De Jonge J, Smulders YM. How do scientists perceive the current publication culture? A qualitative focus group interview study among Dutch biomedical researchers. BMJ Open. 2016;6(2).
42. Cook C, Heath F, Thompson R. A meta-analysis of response rates in web-or internet-based surveys. Educ Psychol Meas. 2000;60(6):821–36.
43. Groves R. Nonresponse rates and nonresponse bias in household surveys: What do we know about the linkage between nonresponse rates and nonresponse bias? Public Opin Q. 2006;70(5):646–75.
44. Cull WL, O'Connor KG, Sharp S, Tang SFS. Response rates and response bias for 50 surveys of pediatricians. Vol. 40, Health Services Research. 2005. p. 213–26.
45. Smith W. Does Gender Influence Online Survey Participation? ERIC Doc Reprod Serv No ED 501717. 2008;1–21.
46. Hox J. Multilevel analysis: techniques and applications. 2nd ed. Mahwah, N.J.: Lawrence Erlbaum Associates; 2002. (Quantitative methodology series).

Supporting information is available on the journal's website:
https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0210599

4

Chapter 4

Personally perceived publication pressure: Revising the Publication Pressure Questionnaire (PPQ) by using work stress models

This chapter has been published as:
Haven TL, Tijdink JK, De Goede MEE, Oort F. Personally perceived publication pressure - Revising the Publication Pressure Questionnaire (PPQ) by using work stress models. *Res Integr Peer Rev. 2019;4(7):1–9.*
https://doi.org/10.1186/s41073-019-0066-6

Abstract

Background: Emphasis on impact factors and the quantity of publications intensifies competition between researchers. This competition was traditionally considered an incentive to produce high quality work, but there are also unwanted side-effects, like publication pressure. To measure the effect of publication pressure on researchers, the Publication Pressure Questionnaire (PPQ) was developed. Upon using the PPQ, some issues came to light that motivated a revision.
Method: We constructed two new subscales based on work stress models using the facet method. We administered the revised PPQ (PPQr) to a convenience sample together with the Maslach Burnout Inventory (MBI) and the Work Design Questionnaire (WDQ). To assess which items best measured publication pressure, we carried out a Principal Component Analysis (PCA). Reliability was sufficient with a Cronbach's alpha of > 0.7. Finally, we administered the PPQr in a larger, independent sample of researchers to check the reliability of the revised version.
Results: Three components were identified as 'stress', 'attitude' and 'resources'. We selected 3 x 6 = 18 items with high loadings in the three-component solution. Based on the convenience sample, Cronbach's alphas were 0.83 for Stress, 0.80 for Attitude, and 0.76 for Resources. We checked the validity of the PPQr by inspecting the correlations with the MBI and the WDQ. Stress correlated .62 with MBI's emotional exhaustion. Resources correlated .50 with relevant WDQ subscales. To assess the internal structure of the PPQr in the independent reliability sample, we conducted a principal components analysis. The three-component solution explains 50% of the variance. Cronbach's alphas were 0.80, 0.78, and 0.75 for Stress, Attitude, and Resources, respectively.
Conclusion: We conclude that the PPQr is a valid and reliable instrument to measure publication pressure in academic researchers from all disciplinary fields. The PPQr strongly relates to burnout and could also be beneficial for policy makers and research institutions to assess the degree of publication pressure in their institute.

Key words: publication pressure, burnout, research integrity, validity, reliability

Background

Scientific output (publications) is the standard performance criterion for individual researchers and research institutions at large (1). The rising prestige of impact factors and emphasis on quantity (of publications), intensifies competition between researchers (2). This competition was traditionally considered an incentive to produce high quality work, but in practice there are also unwanted effects of this hyper-competitive and demanding publication climate in which you are mainly evaluated by the number of publications. This can result in (perceived) publication pressure (3). Publication pressure is studied for its effects on research integrity, as the pressure to publish may persuade researchers to cut corners (4,5). Publication pressure has also been linked to burn-out in senior researchers as well as academic drop-out among junior researchers (6,7).

To measure these effects on research and researchers, the Publication Pressure Questionnaire (henceforth: PPQ) was developed (8). The PPQ aimed to assess publication pressure as perceived by biomedical researchers and has been used to measure publication pressure in both The Netherlands and Belgium. Publication pressure was found to relate to burnout and associated with scientific misconduct (9,10).

The PPQ was the first instrument to measure publication pressure in biomedical researchers. Upon using the PPQ in various studies, a few methodological limitations came to light: 1) The relation between the PPQ and burnout is moderate; 2) the PPQ fails to cover the construct of personally experienced stress; 3) it is unknown how publication pressure relates to general work pressure in academics; and 4) the PPQ is particularly focused on (bio)medical research(ers).

Firstly, although intended to assess publication pressure, the PPQ items mostly ask about the researchers' attitude regarding current publication culture. The fact that a researcher perceives the current publication culture as negative, does not necessarily indicate severe personal publication pressure. Since the majority of the PPQ items do not reflect the core question (does this researcher experience publication pressure and if so, how much?), this ambiguity in interpretation of the PPQ sum score threatens the content validity of the PPQ.

Secondly, in the PPQ's validation study, the relationship between the PPQ and the Maschlach Burnout Inventory (MBI) (11) was investigated. The PPQ correlated only moderately with relevant MBI subscale scores (r = .34 with emotional exhaustion and r = .31 with depersonalisation). Yet, a meta-analysis by Lee & Ashfort (1996) found all relevant job stressors to correlate above .5 with emotional exhaustion (12). If burnout is a feasible outcome of publication pressure (8), one would expect correlations to be higher. Gillespie et al. (2001) found that two-thirds of the academics they consulted in their study described psychological problems resulting from stress, with burnout featuring prominently (13). If the PPQ scores are not consistent with initial ideas about its relationship with burnout, then this leaves doubts about the convergent construct

validity of the PPQ.

Thirdly, publication pressure could be viewed as one aspect of work pressure in academics, which raises the question of how (the measurement of) publication pressure relates to (more general measurements of) work pressure, which is one of the most important research areas in work psychology (14). There are interesting parallels between work pressure and publication pressure. Both work pressure and publication pressure may lead to burnout-like symptoms and both may encourage one to think about performing potential misbehaviours (15–18). Yet, there is little mention in research integrity literature of some of the prominent work stress models from psychology (19). As the PPQ was initially designed to study the effect of publication pressure on researchers' tendency to misbehave, work stress models may be a helpful extension for studying the effect of publication pressure on researchers and their integrity (8).

In addition, how does publication pressure relate to other well-known causes of stress and burnout, such as work-home interference and job insecurity (the fear of losing one's job)? Both work-home interference and job insecurity seem highly relevant to academic researchers (13,20,21). Since the PPQ does not mention any of these (arguably relevant) constructs, that leaves its divergent construct validity to be desired.

Lastly, the PPQ was constructed and tested with professors working in biomedicine in mind. Arguably, biomedical professors constitute only a small subset of the total population of academic researchers that may experience publication pressure. Some PPQ items explicitly mention the medical field (e.g. "My scientific publications contribute to better (future) medical care"). Since the current phrasing is tailored to biomedical researchers, it is hard to assess the generalisability of the results to other academic disciplines, lowering the external validity of the PPQ.

These validity issues formed the motivation for a revision of the PPQ. Below, we first present the methods used to revise the PPQr and to construct new items. Second, we assess the factorial structure and examine the reliability and (internal and external) validity of the new PPQr subscales, by calculating Cronbach's alphas and correlations between the PPQr subscales and relevant work pressure and burnout constructs. Finally, we administer the PPQr to a larger, independent sample of researchers to check its reliability.

Methods

Study's aim

This study's aim is threefold. First, to revise the PPQ and address the above-mentioned concerns that should lead to the design of the revised version of the PPQ (the PPQr, see *Instrument construction* section). Second, to study the PPQr in relation to work pressure and burnout (see *Pilot study* section). Lastly, we want to redistribute the

PPQr in an independent sample and test the presupposed structure and reliability in a more diverse group of academics (see *Reliability study* section).

Instrument construction

In work psychology literature, stress and the consequences of stress at work are some of the most frequently studied topics (14). One prominent conception of how stress is moderated stems from the Job Demands-Control model (JDC) (22), the Effort-Reward Imbalance Model (ERI) (23) and later the Job Demands-Resources model of burnout (JD-R) (24). These models propose that the balance between positive and negative work characteristics is important for various work outcomes. As a result, stress is seen as an interplay between (high) job demands and (low) job resources (24).

Within the JD-R model, demands refer to physical, social, or organizational aspects of the job that require sustained effort, such as work pressure, ambiguity about an employee's role, or stressful events in general. Resources, on the other hand, refer to aspects of the job that are helpful for achieving work goals, stimulate development and reduce the costs of job demands. Examples of resources are social support from family or colleagues, possibilities for career development, and autonomy. Job demands and resources interact; job resources can buffer the impact of job demands in predicting employee health and motivation. In a nutshell: when demands exceed resources, someone is likely to perceive stress or even burn-out symptoms (24).

There are warning signs that burnout is a growing problem in academia (11). A Flemish study found 50% of PhD students faced psychological distress which caused them to be more at risk for developing burnout compared to the general higher educated population (6). Moreover, a UK-study demonstrated that 15% of academics experience such profound levels of stress that they needed medical advice (25) and roughly one in five Dutch medical professors met official burnout criteria (7).

Based on the literature on stress and burnout, we tried to determine what content should be included in an instrument to measure publication pressure. The PPQ mostly enquired into attitudes towards the current publication culture, but items about perceived publication stress or publication resources were missing. As a consistent definition of work stress in academia is lacking (26), we identified possible job demands for academic researchers inspired by the Job Content Questionnaire (27), i.e. (lack of) social support, (lack of) autonomy, authority, psychological demands and skills. All these constructs have been extensively studied in relation to stress and burnout (28).

To ensure content validity, we used the facet method to formulate new items (29). The facet method strengthens content validity by structuring the analysis of the concept one wishes to study (30). We used the facet method to ensure that we did not miss any relevant aspects of publication pressure. Relevant work stress characteristics, applied to publishing, appear in the left column of Additional file 1. The top row of Additional file 1 specifies two types of experiences: first, whether respondents experience stress

and second, whether respondents experience lack of resources based on the Perceived Stress Scale (31). In total, we formulated 37 items, with the aim of ending up with a shorter and more user-friendly questionnaire. We kept the response options for the PPQr the same as the original PPQ: items are scored on a 5-point Likert scale (1 = "totally disagree", 5 = "totally agree").

To check whether our drafted items were understandable and generalisable, we asked PhD candidates and assistant professors from biomedical and behavioural sciences to test and inspect the items for comprehensiveness (n = 9). This resulted in minor modifications in wording to improve clarity and correct interpretation.

Pilot study

Materials

In addition to the PPQr items and questions about demographics, we included the complete Maslach Burnout Inventory-General Survey (MBI) (11), subscales of the Work Design Questionnaire (WDQ) (32), the Job Insecurity Scale (JIS) (33) and items about negative work-home interference taken from the Survey Work-home Interaction—NijmeGen, (SWING) (34).

The Maslach Burnout Inventory-General Survey (MBI) was included to measure burnout and stress and to examine the PPQr's convergent construct validity. Being the most used instrument to measure burnout, the MBI consists of 22 items spread over 3 subscales: emotional exhaustion (9 items, α = .90), depersonalisation (5 items, α =.80) and personal accomplishment (8 items, α =.73). Emotional exhaustion is the feeling of a depletion of energy during work and a negative attitude towards work related activities. Depersonalisation is alienation from work, where someone's interested in work or colleagues is completely lost. Personal accomplishment is a positive subscale: it regards feelings of contentment and a sense of being capable of doing the work. Responses are scored on a Likert scale from 1 ("never") to 5 ("every day") (11).

We chose the Work Design Questionnaire (WDQ) to measure the PPQr's divergent construct validity in relation to work pressure (32). Not all subscales of the WDQ are relevant to working in academia (Contextual Characteristics such as physical demands are arguably less relevant for academics), so we chose a selection of the WDQ items from categories Task Characteristics (12 items), Knowledge Characteristics (12 items) and Social Characteristics (9 items). From Task Characteristics, we took subscales Work Scheduling Autonomy (α = .85), Decision-Making Autonomy (α = .85) and Work Methods Autonomy (α = .88), and Feedback from Work (α = .86). From Knowledge Characteristics, we took Information Processing (α = .87), Problem Solving (α = .84) and Specialisation (α = .84). Finally, we took two subscales from Social Characteristics, namely Social Support (α = .82) and Feedback from Others (α = .88). All items are scored on a 5-point scale from "strongly disagree" to "strongly agree".

Two other constructs commonly referred to as causes of stress and burnout in

academics are job insecurity and negative work-home interference. Job insecurity is the subjective feeling that you may lose your job at any moment (33,35). The Job Insecurity Scale measures personally experienced job insecurity. The questionnaire consists of one subscale that measures the perceived threat of losing one's job and the worries that accompany this threat (4 items). The reliability of this scale is good (α = .82). Answers are scored on a 5-point Likert scale with 1 being "strongly disagree" and 5 "strongly agree".

Negative work-home interference regards the hindrance that people experience at home as a result of their work. Typical examples entail over-working, staying at work for long hours during the week, or having to always work on the weekends. We used the subscale work-home interference (9 items) of the Survey Work-home Interaction—NijmeGen (SWING) (34). Reliability of the negative work-home interference scale is good (α = .85). Answers indicate how often participants experience certain situations on a 4-point scale from "practically never" (0) to "practically always" (3).

Procedure

We distributed the survey through our own network and social media. The survey was available via an online link through Qualtrics. The questionnaire included (parts of) existing instruments (complete MBI, relevant parts of the WDQ, JIS and SWING) as well as the PPQr items and demographics. All items were in English. After reading about the purpose and procedure of the study, participants had to give informed consent before continuing to the actual questions. The questionnaire took approximately 15 minutes to complete.

Participants

All researchers (including PhD students) currently employed at an academic institution were eligible to participate. 205 researchers started the questionnaire, 129 respondents provided enough useful answers to include them for analyses, of which 66% were female. The majority of the respondents worked in biomedicine (52%), besides 43% that worked in social science and 5% that had a background in natural sciences or humanities. 38% of the participants were PhD students, 24% were currently employed as postdoctoral researchers, 20% as assistant professors and 19% associates or full professors. The average age was 37.

Results

In order to assess which items best measured publication pressure, we carried out a Principal Component Analysis (PCA). The first three components were identified as 'Stress', 'Attitude' and 'Resources'. We selected 3 x 6 = 18 items with high loadings on the three-component solution, but not necessarily the items with the highest loadings, because we tried to cover as many aspects relevant to experiencing lack of resources when working on publications or experiencing stress when working on publishing as

possible. See Additional file 3 for the pattern matrix with the three components. See Table 1 for an overview of selected items per subscale.

Item-rest correlations (aka corrected item-total correlations) for Stress were between .44 and .66. Attitude item-rest correlations ranged from .38 to .60. Finally, Resources' items correlated between .30 and .54 with their subscale.

Table 1. PPQr subscales' items with alphas, means, standard deviations and item-rest correlations.

(Cronbach's alpha)	Mean*	*SD**	*r**	Items
Stress (α = .80)	3.22	.80		
	2.98	1.22	.55	I experience stress at the thought of my colleagues' assessment of my publications output.
	3.88	1.09	.44	I feel forced to spend time on my publications outside office hours.
	3.52	1.10	.43	I cannot find sufficient time to work on my publications.
	2.79	1.12	.60	I have no peace of mind when working on my publications.
	2.87	1.01	.50	I can combine working on my publications with my other tasks.
	3.27	1.12	.57	At home, I do not feel stressed about my publications.
Attitude (α = .78)	3.59	.68		
	3.39	1.10	.46	The current publication climate puts pressure on relationships with fellow-researchers.
	3.84	.94	.47	I suspect that publication pressure leads some colleagues (whether intentionally or not) to cut corners.
	3.41	1.08	.46	In my opinion the pressure to publish scientific articles has become too high.
	3.93	.93	.50	My colleagues judge me mainly on the basis of my publications.
	2.99	.98	.40	Colleagues maintain their administrative and teaching skills well, despite publication pressure.
	4.01	.93	.47	Publication pressure harms science.
Resources (α = .75)	2.21	.63		
	2.09	.82	.45	When working on a publication, I feel supported by my co-authors.
	1.84	.78	.42	When I encounter difficulties when working on a publication, I can discuss these with my colleagues.
	2.26	1.04	.39	I have freedom to decide about the topics of my publications.
	2.30	1.00	.37	When working on a publication, many decisions about the content of the paper are outside my control.
	2.46	1.05	.50	I cannot cope with all aspects of publishing my papers.
	2.31	.90	.46	I feel confident in the interaction with co-authors, reviewers and editors.

* These means, standard deviations and item-rest correlations are taken from the reliability sample.

We calculated the reliability of the subscales. Cronbach's alphas were 0.83 for Stress, 0.80 for Attitude, and 0.76 for Resources, which are all considered acceptable (36).

Items were recoded in such a way that higher subscale scores indicate more publication pressure. A respondent that scores low on all subscales experiences little stress from publishing, has a positive attitude about publishing, and has sufficient resources.

We checked the validity of the PPQr by inspecting the correlations with the MBI, the WDQ, the JHI and WHI. Stress correlated .62 with MBI's emotional exhaustion scale and .46 with the total MBI. Work-home interference and stress were also highly correlated (r = .69). Resources correlated between -.41 and -.50 with relevant included WDQ subscales and moderately with job-insecurity (r = .33). For a full overview of subscale correlations, see Table 2. For PPQr item correlations with PPQr subscales, see Additional file 2.

Table 2. Correlations between included constructs.

Constructs and subscales	PPQr Stress	PPQr Attitude	PPQr Resources
MBI total	.46	.32	.19
MBI – Emotional Exhaustion	.62	.42	.34
MBI – Depersonalisation	.33	.37	.38
MBI – Personal Accomplishment	-.22	-.23	-.40
WDQ – Task	-.37	-.22	-.47
WDQ – Knowledge	.25	.06	-.05
WDQ – Feedback	-.30	-.48	-.41
WDQ – Social	-.30	-.39	-.50
Job Insecurity	.17	.17	.33
Work-home interference	.69	.41	.31
PPQ - Stress	1	.48	.43
PPQ - Attitude	.48	1	.36
PPQ - Resources	.43	.36	1

Note: sample size is 129

To assess the added value of publication pressure as an indicator of burnout, we conducted hierarchical regression analyses with emotional exhaustion (the most prototypical burnout indicator from the MBI) as the outcome variable. Various predictor selection procedures yielded the same result. We found emotional exhaustion to be best predicted by work-home interference, followed by social support and publication stress (r^2 =.59). This indicates that Publication Stress is a relevant indicator of burnout, even when considering the influence of other burnout predictors such as work-home interference and (lack of) social support. See Additional file 4 for the prediction model(s).

We conclude that the PPQr is sufficiently reliable (all Cronbach's alpha's > 0.7; (36), Construct validity is also good, as evidenced by its strong correlations with the relevant MBI and WDQ subscales. As Publication Stress is a significant predictor of burnout, this indicates good predictive validity.

Still, these are preliminary conclusions, as we used a single sample for both item selection and reliability and validity analysis. In order to check whether the proposed structure and reliability would hold, we administered the PPQr in a large and independent sample, as part of a study investigating the academic research climate (37), see www.amsterdamresearchclimate.nl.

Reliability study

Materials

Besides the PPQr (18 items) and demographics, the survey contained the Survey of Organisational Research Climate (SOuRCe©) (38) and a list of 60 major and minor misbehaviours (39). In this paper, we only report on the structure and reliability of the PPQr.

Procedure

We obtained ethical approval from the Scientific and Ethical Review Board of the Faculty of Behavioural and Movement Sciences from the VU University Amsterdam. A data sharing agreement with participating institutions University of Amsterdam, Amsterdam Medical Centre, and VU University Medical Centre, secured the e-mail addresses of all academic researchers. We designed and distributed the survey using Qualtrics.

First, we sent an informational letter to explain the purpose of the study. The survey questionnaire was sent out by e-mail and started when participants provided informed consent. The complete questionnaire took about 15 to 20 minutes to complete. We sent three reminders, each 10 days apart.

Participants

All academic researchers, employed at an academic institution in Amsterdam between May 1st and July 18th 2017, were eligible to participate. This again included

PhD students. To be eligible for inclusion, a respondent had to be involved in research for at least one day per week. 1063 academic researchers completed the PPQr (59% women). 56% worked in biomedicine, 23% were from the department of social sciences and the remaining 21% from the departments of natural sciences and humanities. 49% were PhD candidates, 30% were postdoc or assistant professors and the remaining 21% were associates or full professors.

Results

A total of 7549 academic researchers were invited to participate in the study, of which 1063 completed the full PPQr (14%). First, we wanted to assess the internal structure of the PPQr by means of item-correlations and principal component analysis. Second, we aimed to assess whether the PPQr is reliable by computing Cronbach's alpha coefficients for each of the PPQr's three subscales.

To assess the internal structure of the PPQr, we conducted a principal components analysis. The three-component solution explained 50% of the variance, and the scree-plot also indicates a three-component solution. The pattern matrix showed that each item had the highest loadings on its own component, see Additional file 5. In addition, we conducted confirmatory factor analyses (CFA) which showed that a three-factor model fitted the data of the full sample satisfactorily, and that the same three factor model also fitted the data of each of the subgroups of men and women, four disciplines, and five academic ranks.

Corrected item-subscale correlations for Attitude ranged between .40 and .50. For Stress, this was slightly higher, between .43 and .60. For Resources item-subscale correlations were between .37 and .50. Cronbach's alphas were 0.80, 0.78, and 0.75 for Stress, Attitude, and Resources, respectively. We also calculated Cronbach's alphas for subgroups of men and women, four disciplines, and five academic ranks, but subgroup results did not substantially deviate from the full sample results. Correlations between the subscales were 0.46 between Stress and Attitude, 0.44 between Stress and Resources, and 0.39 between Attitude and Resources.

We conclude that the PPQr is a robust instrument to measure publication pressure in academic researchers.

Discussion

We aimed to improve the PPQ in order to accommodate concerns about its validity and created a revised version of the PPQ (PPQr). This new instrument (18 items) consists of three subscales: Publication Stress (6 items), Publication Attitude (6 items) and Publication Resources (6 items). After validating the PPQr in a convenience sample, we tested the reliability of the PPQr in an independent sample

We conclude that the PPQr is a valid instrument; correlations with both MBI subscales and relevant WDQ subscales are substantial and in the expected direction (all relevant *r*s > 0.4). Each of these subscales is reliable (all Cronbach's alphas > 0.7). The PPQr can be used to study publication pressure among academic researchers from all disciplinary fields and academic ranks.

This enables us to investigate the relationship between publication pressure and work stressors. The PPQr is strongly related to work pressure (correlations Resources and relevant WDQ subscales between .41 and .50), yet publication pressure seems at least in some ways to differ from 'classic' work pressure, as it was only marginally related to the Knowledge Characteristics subscale of the WDQ (see Table 2). Furthermore, subscale Resources underscores the relationship between publication pressure and job insecurity: a researcher with less resources is more likely to experience job insecurity (or conversely: a researcher with low job insecurity is more likely to perceive more resources). Stress is strongly associated with work-home interference: a researcher who experiences more work-home interference is more likely to experience publication stress (and vice versa).

Hierarchical regression analyses indicated that publication pressure was strongly related to burnout. Hence, a researcher who perceives higher publication pressure may be *more likely* to develop burnout symptoms. With the PPQr, this relation becomes even more apparent than with the PPQ, since its correlations with the MBI subscale 'emotional exhaustion' are stronger than those of the PPQ (r = .34 for the original PPQ and r = .62 for the PPQr).

However, in our sample, work-home interference was more strongly related to burnout (r = .73 and r^2 =.53, p <.001, see Appendices Table 4). This is to be expected, as work-home interference is known to be directly associated with burnout (40). Nevertheless, adding publication stress to the hierarchical regression model significantly increased the explained variance, emphasising its importance besides other burnout markers.

Alternative explanations

It could be that publication pressure is determined by factors currently not included in the PPQr; two particularly important ones being acquisition pressure and pressure from teaching duties. Along these lines, role conflict (the reasoning here is that since people have a limited amount of time and multiple tasks or responsibilities, when one task requires major attention, the other tasks suffer since there is simply no more time or attention left) is known to be a predictor of work stress. In this situation, the internal role conflict would regard academics striving to be both good researchers and good teachers. We did not measure role conflict in our study, yet it seems plausible that role conflict would lead to burnout and not so much publication pressure per se. We encourage future research into the relationship between evaluation criteria and role conflict in relation to publication pressure.

Another alternative explanation would be that publication pressure is mostly dependent on evaluation criteria as set by the institution of employment. To put it simply: a postdoc that needs to publish 10 papers a year will feel more publication pressure than a postdoc who is evaluated based on just 3 papers a year. A complicating factor herein is that it is nearly impossible to access individual evaluation criteria. Nevertheless, it would be fruitful in future research to study PPQr scores in relation to the amounts of papers a researcher is expected to publish within a specific timeframe, to attribute and interpret the score of the 3 subscales for an individual researcher and develop cut-off scores.

Finally, it could be that researchers with burn out symptoms experience more pressure and annoyance from the current publication system because of their symptoms. In this conceptualisation, burnout precedes publication pressure. Alternatively, since there is an abundance of research indicating that high job demands increase the risk of developing burnout symptoms, it could be that both publication stress and burnout are the result of excessive job demands or a related variable. We cannot exclude these possibilities based on our data and would encourage longitudinal investigation into this matter to confirm that publication pressure precedes burnout or vice versa.

Strengths

This study moves away from operationalizing publication pressure as an attitude or opinion of its severity to extended operationalization towards personally experienced pressure. Individual experience, not opinion, is one of the strongest driving forces of behaviour (41); someone can think that publication pressure pushes researchers beyond limits of responsible research, yet if that person herself does not lay awake at night because of her *H*-index, there seems little reason to suspect burnout is looming.

Secondly, our large sample (>1000 academic researchers) may increase the reliability of the results. Our sample consists of academic researchers from all academic ranks and disciplinary fields, which should indicate better generalisability.

Limitations

The most evident limitation is the use of a convenience sample for the pilot study. We recruited respondents by means of our personal network and social media. This can result in a selective pilot sample. Still, we found similar results in the reliability study, using an independent sample.

Secondly, the response rate (14%) for our reliability study is low. This could increase the chance of a response bias, which occurs when responders differ critically from non-responders. Statistics on female academics in The Netherlands indicate women make up 39% of the academic workforce, whereas 59% of our participants identified as female. Similarly, national statistics indicate 30% of academic researchers are currently enrolled as (non-biomedical) PhD candidates compared to 41% in our sample. Yet, this would

only indicate response bias if the PPQr items were understood differently depending on one's subgroup. Since the CFA model fit did not differ significantly between different subgroups, we conclude that this should not affect the validity of the PPQr results we present here.

Another limitation is that the current PPQr cannot be expressed in one total score, as was the case with its predecessor. We intended to make a total score to ease interpretation. However, upon reflection, it was unclear what that total score would express and hence we decided against it.

Conclusion

The PPQr is a valid and reliable measurement instrument. It covers the complex construct of publication pressure better than its predecessor and can measure publication pressure among researchers from all disciplinary fields. PPQr scores are strongly related to emotional exhaustion scores. The PPQr could also be beneficial for policy makers and research institutions to assess the degree of publication pressure in their institute. To sustain responsible research, institutions should invest in resources to combat the high demands, such as fostering an open atmosphere where difficulties can be discussed and where researchers have some freedom to decide what to study.

Acknowledgements

We would like to acknowledge master's student Jip Eleveld for her help in collecting the convenience sample data and her thesis that served as great inspiration for this manuscript.

References

1. Court S, Kinman G. Tackling Stress in higher education. 2008.
2. Anderson MS, Ronning EA, De Vries R, Martinson BC. The perverse effects of competition on scientists' work and relationships. Sci Eng Ethics. 2007;13(4):437–61.
3. Van Dalen HP, Henkens K. Intended and unintended consequences of a publish-or-perish culture: A worldwide survey. J Am Soc Inf Sci Technol. 2012;63(7):1282–1293.
4. Alberts B, Kirschner MW, Tilghman S, Varmus H. Rescuing US biomedical research from its systemic flaws. Proc Natl Acad Sci. 2014;111(16):5773–7.
5. Fanelli D. Do pressures to publish increase scientists' bias? An empirical support from US states data. PLoS One. 2010;5(4).
6. Levecque K, Anseel F, De Beuckelaer A, Van der Heyden J, Gisle L. Work organization and mental health problems in PhD students. Res Policy. 2017;46(4):868–79.
7. Tijdink J, Vergouwen A, Smulders Y. Publication pressure and burn out among Dutch medical professors: A nationwide survey. PLoS One. 2013;8(9).
8. Tijdink J, Smulders Y, Vergouwen A, de Vet H, Knol DL. The assessment of publication pressure in medical science; validity and reliability of a Publication Pressure Questionnaire (PPQ). Qual life Res An Int J Qual life Asp Treat care Rehabil. 2014;23(7):2055–62.
9. Tijdink J, Vergouwen A, Smulders Y. Emotional exhaustion and burnout among medical professors; A nationwide survey. BMC Med Educ. 2014;14(1):1–7.
10. Tijdink JK, Verbeke R, Smulders YM. Publication pressure and scientific misconduct in medical scientists. J Empir Res Hum Res Ethics. 2014;9(5):64–71.
11. Maslach C, Jackson SE. The measurement of experienced burnout. J Organ Behav. 1981;2(2):99–113.
12. Lee RT, Ashforth BE. A meta-analytic examination of the correlates of the three dimentions of job burnout. J Appl Psychol. 1996;81(2):123–33.
13. Gillespie NA, Walsh M, Stough C, Winefield AH, Dua J. Occupational stress in universities: Staff perceptions of the causes, consequences and moderators of stress. Work Stress. 2001;15(1):53–72.
14. Ganster DC, Rosen CC. Work stress and employee health: A multidisciplinary review. Vol. 39, Journal of Management. 2013. 1085–1122 p.
15. Vengoechea J, Moreno S, Ruiz A. Misconduct in medical students. Dev World Bioeth. 2008;8(3):219–25.
16. Donders NCGM, van der Gulden JWJ, Furer JW, Tax B, Roscam Abbing EW. Work stress and health effects among university personnel. Int Arch Occup Environ Health. 2003;76(8):605–13.
17. Taris TW. Is there a relationship between burnout and objective performance? A critical review of 16 studies. Work Stress. 2006;20(4):316–34.
18. Dyrbye LN, Massie FS, Eacker A, Harper W. Relationship between burnout and professional conduct and attitudes among US medical students. JAMA J Am Med Assoc. 2010;304(11):1173–80.
19. Karen SL, Anderson MS, Rosenberg L. Academic misconduct and values: The department's influence. Rev High Educ. 1995;18(4):393–422.

20. Kinman G, Jones F. A life beyond work? Job demands, work-life balance, and wellbeing in UK Academics. J Hum Behav Soc Environ. 2008;17(1–2):41–60.
21. Tytherleigh MY, Webb C, Cooper CL, Ricketts C. Occupational stress in UK higher education institutions: A comparative study of all staff categories. High Educ Res Dev. 2005;24(1):41–61.
22. Karasek RA, Theorell T. Healthy work: Stress, productivity and the reconstruction of working Life. New York: Basic Books; 1990.
23. Siegrist J. Adverse health effects of high-effort/low-reward conditions. J Occup Health Psychol. 1996;1(1):27–41.
24. Demerouti E, Nachreiner F, Bakker AB, Schaufeli WB. The job demands-resources model of burnout. J Appl Psychol. 2001;86(3):499–512.
25. Abouserie R. Stress, coping strategies and job satisfaction in university academic staff. Educ Psychol. 1996;16(1):49–56.
26. Boyd L. Exploring the utility of workload models in academe: a pilot study. J High Educ Policy Manag. 2014;36(3):315–26.
27. Karasek R, Amick B. The Job Content Questionnaire (JCQ): An instrument for internationally comparative assessments of psychosocial job characteristics. J Occup Health Psychol. 1998;3(4):322–55.
28. Aronsson G, Theorell T, Grape T, Hammarström A, Hogstedt C, Marteinsdottir I, et al. A systematic review including meta-analysis of work environment and burnout symptoms. BMC Public Health. 2017;17(1):1–13.
29. Guttman L. Introduction to facet design and analysis. Acta Psychol (Amst). 1959;130–8.
30. Oosterveld P. Questionnaire Design Methods. Nijmegen: Berkhout; 1996.
31. Cohen S, Kamarck T, Mermelstein R, Health J, Behavior S, Dec N. A Global Measure of Perceived Stress A Global Measure of Perceived Stress. J Heal Soc Behav Soc Behav. 1983;24(4):385–96.
32. Morgeson FP, Humphrey SE. The Work Design Questionnaire (WDQ): Developing and validating a comprehensive measure for assessing job design and the nature of work. J Appl Psychol. 2006;91(6):1321–39.
33. Vander Elst T, De Witte H, De Cuyper N. The Job Insecurity Scale: A psychometric evaluation across five European countries. Eur J Work Organ Psychol. 2014;23(3):364–80.
34. Geurts SAE, Taris TW, Kompier MAJ, Dikkers JSE, Van Hooff MLM, Kinnunen UM. Work-home interaction from a work psychological perspective: Development and validation of a new questionnaire, the SWING. Work Stress. 2005;19(4):319–39.
35. de Witte H, Pienaar J, de Cuyper N. Review of 30 years of longitudinal studies on the association between job insecurity and health and well-being: Is there causal evidence? Aust Psychol. 2016;51(1):18–31.
36. Streiner DL. Starting at the beginning: An introduction to coefficient alpha and internal consistency starting at the beginning. J Pers Assess. 2003;80(1):99–103.
37. Haven TL, Tijdink JK, Martinson BC, Bouter LM. Perceptions of research integrity climate differ between academic ranks and disciplinary fields - Results from a survey among academic researchers in Amsterdam.

38. Martinson BC, Thrush CR, Lauren Crain A. Development and validation of the survey of organizational research climate (SORC). Sci Eng Ethics. 2013;19(3):813–34.
39. Bouter LM, Tijdink J, Axelsen N, Martinson BC, ter Riet G. Ranking major and minor research misbehaviors: results from a survey among participants of four World Conferences on Research Integrity. Res Integr Peer Rev. 2016;1(17):1–8.
40. Blom V, Sverke M, Bodin L, Bergström G, Lindfors P, Svedberg P. Work-home interference and burnout: A study based on swedish twins. J Occup Environ Med. 2014;56(4):361–6.
41. Glasman LR, Albarracín D. Forming attitudes that predict future behavior: A meta-analysis of the attitude–behavior relation. Psychol Bull. 2006;132(5):778–822.

Additional files are available on the journal's website:
https://researchintegrityjournal.biomedcentral.com/articles/10.1186/s41073-019-0066-6

5

Chapter 5

Perceived publication pressure in Amsterdam: Survey of all disciplinary fields and academic ranks

This chapter has been published as:
Haven TL, Bouter LM, Smulders YM, Tijdink. Perceived publication pressure in Amsterdam: Survey of all disciplinary fields and academic ranks. *PLoS One. 2019;14(6):e0217931.*
https://doi.org/10.1371/journal.pone.0217931

Abstract

Publications determine, to a large extent, a researcher's possibility to stay in academia ("publish or perish"). While some pressure to publish may incentivise high quality research, too much publication pressure is likely to have detrimental effects on both the scientific enterprise and on individual researchers. Our research question was: What is the level of perceived publication pressure in the four academic institutions in Amsterdam, and does the pressure to publish differ between academic ranks and disciplinary fields? Investigating researchers in Amsterdam with the revised Publication Pressure Questionnaire, we find that a negative attitude towards the current publication climate is present across academic ranks and disciplinary fields. Postdocs and assistant professors (M = 3.42) perceive the greatest publication stress and PhD-students (M = 2.44) perceive a significant lack of resources to relieve publication stress. Results indicate the need for a healthier publication climate where the quality and integrity of research is rewarded.

Keywords: publication pressure, responsible conduct of research, research integrity, academic ranks, disciplinary fields

Introduction

The current state of academia is sometimes referred to as a system affected by hypercompetition (1–3). This goes hand in hand with strong emphasis on quantitative assessment of scientific output through journal impact factors, citation analyses and the H-index (4–6). Their number of publications, citations and grants determine to a large extent the status and recognition of academic researchers (7–10). Consequently, these indicators influence the recruitment, promotion and tenure appointments of researchers (11,12). This may in turn induce a high level of perceived publication pressure.

In line with Woolf (13), we define perceived publication pressure as the subjective pressure resulting from the feeling that one *has* to publish. In line with work stress literature, strong perceptions of pressure may provoke stress, but need not do so when one has many resources available to manage the pressure (14). Applied to publication pressure: Publication demands and attitudes towards the current publication climate determine the perceived pressure, yet pressure can be alleviated by resources like helpful co-authors, involved colleagues or supervisors, and a sense of academic competence (15).

Some degree of publication pressure can be an incentive to produce high quality scientific work (13,16). Yet, too much publication pressure may have detrimental effects on the scientific enterprise in general and on individual researchers in particular (17). Excessive publication pressure is associated with poor quality research (and teaching), a decreased willingness to share raw data, less involvement from researchers in public and policy issues, and less academic creativity (16,18–20). The perceived hypercompetition is thought to lead to less rigorous ("rushing into print") and less reliable science (1,21,22). Publication pressure is associated with a greater likelihood to engage in research misbehaviours (23–25). Lastly, publication pressure is associated with a disproportionate focus on positive and specular findings (21,22,26,27).

Publication pressure may also have detrimental effects on individual researchers. It is linked to a poor research climate and may render academic researchers emotionally exhausted (3,28). Previous research on publication pressure found junior researchers to experience more publication pressure compared to their senior counterparts (20,23). Studies investigating publication pressure thus far have mainly included academic researchers from particular disciplines like biomedicine, management and population studies, and included only a subset of academic ranks (16,20,23). This limits the generalizability of the degree to which researchers perceive publication pressure.

The current study aims to assess whether researchers from all academic ranks (including PhD students) and all disciplinary fields perceive publication pressure. This is important, as differences between academic ranks could signal the need for tailored interventions. Besides, comparing different disciplinary fields may enable us to determine fields that perceive less publication pressure. This may generate new insights

in the nature of publication pressure and possible protective factors. Our research question was: What is the level of perceived publication pressure in the four academic institutions in Amsterdam, and does the pressure to publish differ between academic ranks and disciplinary fields?

Materials and Methods

Ethical statement

Our study was ethically reviewed and approved by the Scientific and Ethical Review board of the Faculty of Behavioural and Movement Sciences (Vrije Universiteit Amsterdam).

Participants

All academic researchers in Amsterdam employed in research for at least one day per week at one of the four selected academic institutions (Vrije Universiteit Amsterdam, University of Amsterdam and the two Amsterdam University Medical Centers) were eligible to participate. This included PhD students, as in The Netherlands, PhD students are employees.

Procedure

First, we set up a data sharing agreement with all participating institutions to safely obtain the e-mail addresses of their researchers. Second, we sent an informational letter inviting all academic researchers in Amsterdam (n = 7465) to take part in our study. The informational letter contained links to the study protocol (S1 protocol) and the study's privacy policy (S1 appendix). In addition, we included a link to a short non-response questionnaire where we asked researchers to report their academic rank and gender, and enquired whether the reason for declining participation resulted from a sense that their data were not protected. For the full non-response questionnaire, see S2 Appendix.

A week later, researchers were invited to complete an online survey. The survey started with an informed consent statement followed by the inclusion check ("Are you currently employed in research for at least one day per week?") and ended with the demographic items about participants' academic rank (PhD student, postdoc, assistant professor, associate professor or full professor) and major disciplinary field: biomedicine (consisting of life and medical sciences), natural sciences, social sciences (including both social and behavioural sciences) or humanities (consisting of humanities, language, communication, law and arts). We used Qualtrics (Qualtrics, Provo, UT, USA) to create and distribute the survey, which took approximately 15 minutes to complete. We sent three reminders, each 10 days apart.

Instruments

We used the revised Publication Pressure Questionnaire (PPQr) to measure publication pressure (15). The PPQr is a valid and reliable instrument to measure publication pressure and consists of 3 subscales scored on a 5-point Likert scale ('Totally agree' = 5, 'Totally disagree' = 1). The Publication Stress subscale (6 items - Cronbach's α = .804) regards the stress a researcher experiences due to the feeling she/he has to publish and includes items such as "I feel forced to spend time on my publications outside office hours". The Publication Attitude (6 items - Cronbach's α = .777) subscale reflects researchers' attitudes towards publication pressure, for example: "Publication pressure harms science". Finally, the Publication Resources subscale (6 items - Cronbach's α = .754) consists of factors that can help prevent publication pressure (e.g. feelings of competence, freedom to choose topics of scientific investigation, involved colleagues). A typical item would be: "When working on a publication, I feel supported by my co-authors.". The full PPQr questionnaire can be found in S3 Appendix.

PPQr subscale scores are computed by taking the average of all items in the subscale. A higher score on all subscales means the researcher perceives publication stress, has a negative attitude towards the publication climate and perceives few publication resources to alleviate publication stress.

The survey contained two other instruments (Survey of Organizational Research Climate (29) and 60 major and minor research misbehaviours (30)), but those analyses will be part of another report see (31) and (32). The interrelations between these concepts will be reported in a separate future paper.

Statistical analyses

We preregistered our analyses on the Open Science Framework, see osf.io/w4t7u. To summarise: First, we calculated overall mean scores for all three subscales and stratified these for academic ranks and disciplinary fields. Second, we assessed whether there were differences between particular academic ranks or disciplinary fields using Bonferroni corrected *F*-tests and Mean Differences (*MD*) with 95% Confidence Intervals (*CI*). Third, we built multivariable regression models to test whether academic rank and disciplinary field were associated with PPQr subscale mean scores. In these regression models, we also looked for evidence of confounding and interaction. Estimates corrected for confounding are provided and instances of interaction were reported. All analyses were conducted using SPSS Statistics.

Results

Response rate and inclusion

From the 7548 researchers that were invited to participate, 30% (n = 2274) followed the link to the online survey. 1073 of the invitees filled in the PPQr (response rate = 14%). Demographic information is listed in Table 1. About 2% of the invitees filled in the non-response questionnaire. See Fig 1.

Table 1. Descriptives of participants, stratified by gender, academic rank and disciplinary field.

		Publication Stress	**Publication Attitude**	**Publication Resources**
	n	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
Males	441	3.10 (.79)	3.58 (.72)	2.08 (.63)
Females	632	3.29 (.79)	3.60 (.65)	2.30 (.61)
PhD students	503	3.18 (.80)	3.60 (.67)	2.44 (61)
Postdocs and assistant professors	318	3.42 (.74)	3.70 (.63)	2.12 (.55)
Associate and full professors*	216	3.03 (.82)	3.42 (.76)	1.80 (.54)
Biomedicine	603	3.16 (.79)	3.60 (.65)	2.24 (.61)
Natural sciences	119	3.12 (.80)	3.51 (.77)	2.04 (.68)
Social sciences	242	3.32 (.80)	3.60 (.71)	2.24 (.64)
Humanities	109	3.42 (.76)	3.58 (.68)	2.16 (.62)
Total participants	1073	3.22 (.80)	3.59 (.68)	2.21 (.63)

* 36 participants failed to disclose their academic rank.

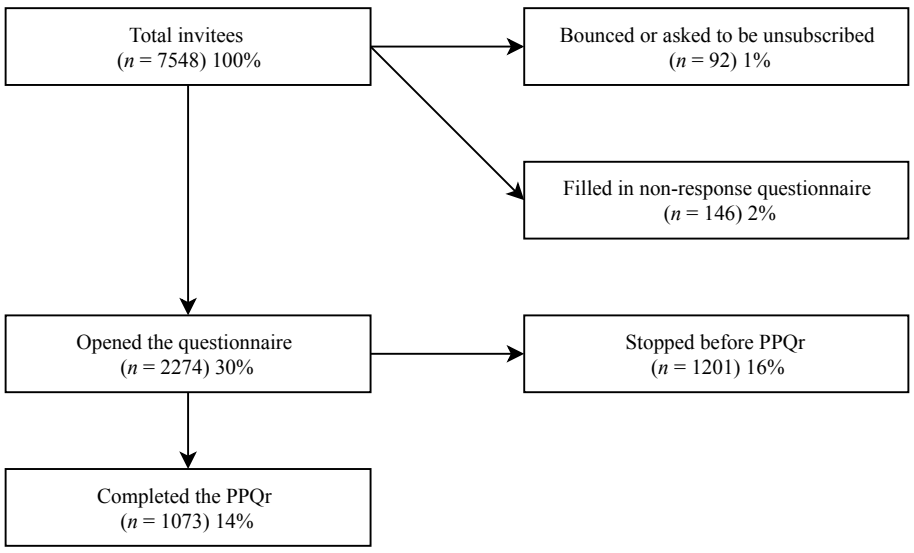

Fig 1. Overview of response rate.

Overall, we find academic researchers in our sample score highest on Attitude (*M* = 3.59). This indicates that their negative attitude towards the publication climate is substantial. There is on average a somewhat lesser degree of Publication Stress (*M* = 3.22) and a relatively small lack of Publication Resources (*M* = 2.21). Stratified and total sample mean scores can be found in Table 1.

Publication pressure by academic rank

Pairwise *Bonferroni* and confounding-corrected (disciplinary field and gender) mean differences between academic ranks indicate that postdocs and assistant professors perceive significantly more publication stress than both PhD students and associate and full professors. Also, both PhD students as well as postdocs and assistant professors have a more negative attitude towards the publication culture compared to full professors. Furthermore, PhD students perceive a significantly greater lack of resources than both postdocs and assistant professors as well as associate and full professors. Finally, postdocs and assistant professors perceive less resources than associate and full professors. See Fig 2. Crude and *Bonferroni* corrected mean differences between pairs of groups can be found in S1 Table. For crude and corrected association models between academic rank and the PPQr subscales, see S2 Table.

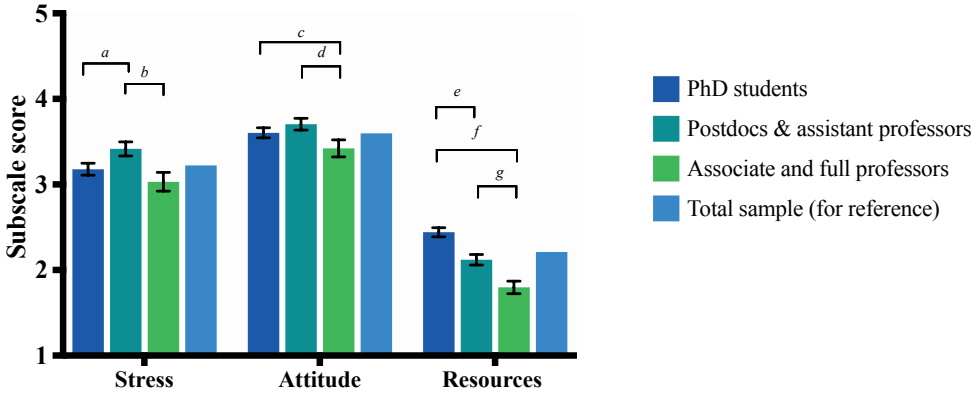

Fig 2. Differences between academic ranks in PPQr subscale scores.

Letters denote significant (α = .05) *Bonferroni* corrected mean differences (*MD*) between pairs of (denoted by the brackets) academic ranks and error bars express 95% Confidence Intervals (*CI*). *MDs* are corrected for confounders (disciplinary field and gender) if applicable. *N* = 1073.

a: *MD* = .237, *CI* = (.103, .371)
b: *MD* = .384, *CI* = (.219, .549)
c: *MD* = .181, *CI* = (.049, .314)
d: *MD* = .282, *CI* = (.139, .426)
e: *MD* = .322, *CI* = (.223, .421)
f: *MD* = .645, *CI* = (.532, .757)
g: *MD* = .322 *CI* = (.201, .444)

Publication pressure by disciplinary field

Pairwise *Bonferroni* and confounding-corrected (academic rank and gender) mean differences indicate that researchers in the humanities perceive more publication stress than both biomedicine researchers and researchers within the natural sciences. Researchers from the social sciences perceive more publication stress than their biomedical colleagues. There were no statistically significant differences between disciplinary fields on attitude scores. Finally, researchers in biomedicine as well as social sciences perceive a significantly greater lack of publication resources than researchers in the natural sciences. See Fig 3. Crude and Bonferroni corrected mean differences between pairs of groups can be found in S1 Table. For crude and corrected disciplinary field association models, see S3 Table.

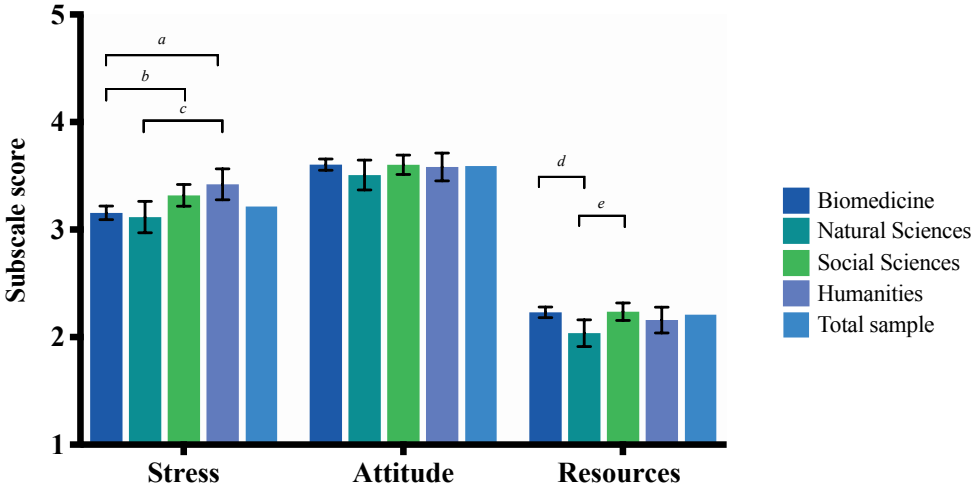

Fig 3. Differences between disciplinary field in PPQr subscale scores.

Letters denote significant (α = .05) *Bonferroni* corrected mean differences (*MD*) between pairs of (denoted by the brackets) disciplinary fields and error bars express 95% Confidence Intervals (*CI*). *MDs* are corrected for confounders (academic rank and gender) if applicable. N = 1073.

a: *MD* = .297, *CI* = (.080, .515)
b: *MD* = .202, *CI* = (.042, .363)
c: *MD* = .318, *CI* = (.040, .596)
d: *MD* = .204, *CI* = (.048, .359)
e: *MD* = .210, *CI* = (.036, .384)

Effect modification

We only found effect modification by disciplinary field concerning the differences between academic ranks' Publication Resources scores. Differences between PhD students and senior academic researchers in perceived Publication Resources are greater

in natural sciences compared to other disciplinary fields. Stratified results are displayed in Table 2.

Table 2. Effect modification from disciplinary field (natural sciences) in the relation between Publication Resources and academic rank[1].

Resources	*PhD students*	*Postdocs/assistant professors*	*Associate & full professors*
Natural sciences	2.36	1.99	1.44
Biomedical sciences	2.44	2.16	1.81
Social sciences	2.54	2.07	1.84
Humanities	2.33	2.25	1.98

[1] Scores stratified for academic rank and disciplinary field.

Effect sizes

We found 12 significant differences between pairs of groups and since we performed many statistical tests, it is likely that some of the significant differences are in fact due to chance. To provide the reader with some guidance on which effects are relevant, we calculated effect sizes of each difference. This analysis was not preregistered and thus should be considered exploratory. The effect sizes range from small to very large using Cohen's effect size criteria (33), see Table 3. To prevent overinterpreting small differences, we will focus further discussion on differences with an effect size of medium or above.

Table 3. Significant (p <.05) differences with corresponding effect sizes

Subscale	**Group vs. Group**	**Effect size[1]**	**Interpretation[2]**
Stress	PhD students vs. Postdocs < assistant professors	.31	Small
Stress	Postdocs & assistant professors < Associate & full professors	.50	Medium
Attitude	Associate & full professors < PhD students	.15	Small
Attitude	Associate & full professors < Postdocs & assistant professors vs.	.41	Small
Resources	Postdocs & assistant professors < PhD students	.55	Medium
Resources	Associate & full professors < PhD students	1.09	Very large
Resources	Associate & full professors < Postdocs & assistant professors	.59	Medium
Stress	Biomedicine < Humanities	.33	Small
Stress	Biomedicine < Social sciences	.20	Small
Stress	Natural sciences < Humanities	.38	Small
Resources	Natural sciences < Biomedicine	.32	Small
Resources	Natural sciences < Social sciences	.31	Small

[1] using Hedges' *G* computed as: *M1 – M2/SDpooled*

[2] Interpreted based on Cohen (33) where an effect size of .20 is defined as small, .50 is medium, .80 is large and 1.30 is very large.

Discussion

We assessed the level of perceived publication pressure in the four academic institutions in Amsterdam and whether the pressure to publish differed between academic ranks and disciplinary fields. Overall, there is a negative attitude towards the publication climate. Hence the 'publish or perish' mantra from the late 20th century may turn into 'publish *and* perish', since even when a researcher publishes reasonably, chances for tenure in academia may still be low (20,34,35). Below, we elaborate on the differences in effect sizes that were medium or above or on those where we found interaction effects (33).

Academic rank differences

Postdocs and assistant professors perceive the most publication stress and have the most negative attitude towards the current publication climate, which is in line with previous studies assessing perceived publication pressure in biomedicine and organisation science (20,23). This finding seems intuitive, as this particular group aims for a (tenured) position in academia and promotion criteria are, to a large extent, based on quantitative publication indicators. Associate and full professors already have an established position, and consequently may perceive less publication pressure. PhD candidates' likelihood of successfully defending their thesis is usually not dependent on their number of publications. This may explain why their publication pressure level is somewhat lower. Besides, some PhD students may not aspire to an academic career and will therefore presumably perceive less publication pressure than their colleagues who do wish to pursue such a career.

However, PhD candidates perceive the greatest lack of resources. This is both alarming and understandable. Arguably, PhD students are inexperienced at handling difficulties that may arise when working on a publication. The same holds for starting postdocs. Consequently, junior researchers could benefit most from supportive colleagues and supervisors. Unfortunately, mentoring may be suboptimal (30,36,37).

Disciplinary field differences

Differences between disciplinary fields were significant but small. Hence, we focus here on the interaction between disciplinary field and academic rank when it comes to perceived Resources. Researchers from the natural sciences perceive the most publication resources, which may be due to their typical organisation in (large) research teams where collaboration is vital for discovery. However, PhD students in the natural sciences perceive a lack of resources that is similar to PhD students from the other disciplinary fields. It may be that insufficient mentoring in the publication process makes them feel incompetent and insecure.

Strengths

This is the first study to comprehensively measure publication pressure with a validated measurement instrument. The three dimensions— stress, attitude and resources— are meaningful components when conceptualising publication pressure. Also, these three dimensions are sufficiently distinctive in the data reported here (15).

Second, this is the first study to investigate publication pressure across academic ranks and disciplinary fields. It can serve as a benchmark for future studies. We managed to include a substantial number of participants in our study, which increases the reliability of the differences found.

Limitations

Our study also has some limitations we would like to address. First, we have a relatively low completion rate (14%) which may be an indication of response bias, although our completion rate is similar to other web-based surveys (38). Only 2% of our invitees filled in the non-response questionnaire, which we consider to be too few to assess whether non-responders differed from responders. Perhaps invitees chose not to respond because they were too focused on their publications, leading to an underestimation. Relatedly, simply mentioning that our study investigated the publication culture could have prompted negative connotations with the publication culture, as it has not gone unnoticed in the public debate in The Netherlands.

To assess the representativeness of our sample, we first looked into the population characteristics. In our sample, 56% of completers indicated working in the biomedical field, whereas 53% of our invitees were employed at one of the Amsterdam University Medical Centers, indicating a small overrepresentation of biomedicine researchers.

Statistics on PhD students employed at both universities in Amsterdam indicated that PhD students make up 30% of the academic workforce, whereas PhD students formed 41% of our sample. Likewise, 44% of academic researchers in Amsterdam are female, yet women made up 57% of our sample, indicating overrepresentation of both PhD students and women.

However, we corrected for the potential gender bias by adjusting our estimates for confounding variables. Besides, we found no effect modification by gender. To conclude, it is unlikely that the selectivity of our sample biased our results.

To assess possible response bias, we conducted a wave analysis. We used late responders – those who responded after the last reminder – as a proxy for nonresponders and compared these to early responders – those who responded after the initial invitation – as described by Phillips' and colleagues (39). Differences were .13, .07 and .02 for Stress, Attitude and Resources, respectively. These differences were then multiplied by the proportion of non-responders, in our case 86%. Consequently, the non-response bias estimates were .11, .07 and .02 for Stress, Attitude and Resources, respectively. These were found to be small compared to the differences that we observed between

the subgroups which ranged from .18 to .65. It is therefore unlikely that non-response affected our conclusions.

Furthermore, the PPQr focuses exclusively on publication pressure. However, research is not conducted in a vacuum, and if teaching or other professional duties put excessive demands on a researcher, then naturally there will be less time left for publishing, which could lead to elevated levels of publication pressure. This can be labelled as role conflict: you are expected to meet different obligations, i.e. teaching, research, and professional duties, in a naturally limited amount of time (40). How much stress is due to *just* publication pressure is unclear (see also (15)).

Relatedly, universities have been subject to neoliberal and Taylorist reforms that were - in a nutshell - intended to make universities more competitive and were accompanied by an excessive focus on researchers' performance management, perhaps at the expense of traditional hallmarks of academia such as teaching and collegiality (41,42). A full review of Neoliberal and Taylorist reforms in academia is beyond the scope of this paper (the reader is referred to Lorenz' excellent paper (43) that includes specific examples of reforms in Dutch academia), but it seems feasible to reason that publication pressure is one of its consequences, although the exact relation has, to our knowledge, not been studied systematically.

Finally, since this is the first study conducted with the PPQr, it's rather difficult to interpret the absolute levels and differences in publication pressure we found.

Future research

Future work should aim to explore if the differences we found generalise internationally. Publication climates in the USA and Asian countries may be different as their funding systems greatly differ (2,44,45). However, the same could apply to closer examples such as Germany and Belgium, as their funding systems are also somewhat different from those in the Netherlands. Interestingly, a study using a previous version of the PPQ found Flemish biomedical researchers to experience more publication pressure than their Dutch colleagues (23). Furthermore, it will be informative to study publication pressure longitudinally to see if it is associated with burn-out and research misbehaviour. Finally, it would be intriguing to investigate qualitatively what it means for researchers to experience high publication pressure and how it impacts their academic work.

Conclusions

Taken together, our results indicate that publication pressure concerns researchers from all disciplinary fields and seems to be a particularly detrimental stressor for postdocs and assistant professors. In addition, PhD students perceive a significant lack of resources that may hamper their development into responsible researchers. The amount of resources is perceived to be better among researchers from the natural sciences, but PhD students in this disciplinary field would nevertheless also benefit from more

support from their senior colleagues. Our findings emphasize the need to move the debate forward towards a healthy publication climate, where researchers are incentivised to focus on the quality and the integrity of their publications and feel supported to conduct responsible research.

Acknowledgements

We would like to acknowledge all members of the steering committee of the Academic Research Climate Amsterdam project (Frans Oort, Gerben ter Riet, Hanneke de Haes, Guy Widdershoven and René van Woudenberg) for their continuous critical input. Frans Oort has been crucial in the creation of the PPQr. Peter van de Ven has been indispensable in discussing statistical tests and models. We also wish to acknowledge Gowri Gopalakrishna and Venetia Qendri for the independent re-analysis of the survey data.

References

1. Anderson MS, Ronning EA, De Vries R, Martinson BC. The perverse effects of competition on scientists' work and relationships. Sci Eng Ethics. 2007;13(4):437–61.
2. Alberts B, Kirschner MW, Tilghman S, Varmus H. Rescuing US biomedical research from its systemic flaws. Proc Natl Acad Sci. 2014;111(16):5773–7.
3. Tijdink JK, Schipper K, Bouter LM, Pont PM, De Jonge J, Smulders YM. How do scientists perceive the current publication culture? A qualitative focus group interview study among Dutch biomedical researchers. BMJ Open. 2016;6(2).
4. Gannon F. The Impact of the Impact Factor. EMBO Rep. 2000;1(4):292–3.
5. Hirsch JE. An index to quantify an individual's scientific research output. PNAS. 2005;102(46):16569–16572.
6. Dinis-Oliveira RJ, Magalhães T. The inherent drawbacks of the pressure to publish in health sciences: Good or bad science. F1000Research. 2016;4(1):419.
7. Bird S. Research ethics, research integrity and the responsible conduct of research. Sci Eng Ethics. 2006;12(3):56.
8. Walker RL, Sykes L, Hemmelgarn BR, Quan H. Authors' opinions on publication in relation to annual performance assessment. BMC Med Educ. 2010;10(1):2–6.
9. Anderson PA, Boden SD. Ethical Considerations of Authorship. SAS J. 2008;2(3):155–8.
10. Buela-casal G. Pathological publishing : A new psychological disorder with legal consequences ? Eur J Psychol Appl to Leg Contex. 2014;6(2):91–7.
11. Court S, Kinman G. Tackling Stress in higher education. 2008.
12. Glick W, Miller C, Cardinal L. Making a life in the field of organization science. J Organ Behav. 2007;28:817–35.
13. Woolf PK. Pressure to publish and fraud in science. Ann Intern Med. 1986;194:254–6.
14. Demerouti E, Nachreiner F, Bakker AB, Schaufeli WB. The job demands-resources model of burnout. J Appl Psychol. 2001;86(3):499–512.
15. Haven TL, Tijdink JK, De Goede MEE, Oort F. Personally perceived publication pressure - Revising the Publication Pressure Questionnaire (PPQ) by using work stress models. Res Integr Peer Rev. 2019;4(7):1–9.
16. Van Dalen HP, Henkens K. Intended and unintended consequences of a publish-or-perish culture: A worldwide survey. J Am Soc Inf Sci Technol. 2012;63(7):1282–1293.
17. Tijdink JK, Smulders YM, Vergouwen ACM, de Vet HCW, Knol DL. The assessment of publication pressure in medical science; validity and reliability of a Publication Pressure Questionnaire (PPQ). Qual Life Res. 2014;23(7):2055–62.
18. DeAngelis CD. Professors not professing. J Am Med Assoc. 2004;292(9):1060–1.
19. Blumenthal D, Campbell EG, Gokhale M, Yucel R, Clarridge B, Hilgartner S, et al. Data witholding in genetics and the other life sciences: Prevalences and predictors. Acad Med. 2006;81(2):137–45.
20. Miller AN, Taylor SG, Bedeian AG. Publish or perish: academic life as management faculty live it. Career Dev Int. 2011;16(5):422–45.

21. Fanelli D. Do pressures to publish increase scientists' bias? An empirical support from US states data. PLoS One. 2010;5(4).
22. Ioannidis JPA. Why most published research findings are false. PLoS Med. 2005;2(8):0696–701.
23. Tijdink JK, Verbeke R, Smulders YM. Publication pressure and scientific misconduct in medical scientists. J Empir Res Hum Res Ethics. 2014;9(5):64–71.
24. Bedeian A, Taylor S, Miller A. Management science on the credibility bubble: Cardinal sins and various misdemeanors. Acad Manag Learn Educ. 2010;9(4):715–25.
25. Bouter LM. Commentary: Perverse incentives or rotten apples? Account Res. 2015;22(3):148–61.
26. Munafò MR, Nosek BA, Bishop DVM, Button KS, Chambers CD, Percie Du Sert N, et al. A manifesto for reproducible science. Nat Hum Behav. 2017;1(1):1–9.
27. Harzing A-W. Publish or Perish: A proposal. Ann Intern Med. 1986;104:261–2.
28. Tijdink J, Vergouwen A, Smulders Y. Publication pressure and burn out among Dutch medical professors: A nationwide survey. PLoS One. 2013;8(9).
29. Martinson BC, Thrush CR, Lauren Crain A. Development and validation of the survey of organizational research climate (SORC). Sci Eng Ethics. 2013;19(3):813–34.
30. Bouter LM, Tijdink J, Axelsen N, Martinson BC, ter Riet G. Ranking major and minor research misbehaviors: results from a survey among participants of four World Conferences on Research Integrity. Res Integr Peer Rev. 2016;1(17):1–8.
31. Haven TL, Tijdink JK, Martinson BC, Bouter LM. Perceptions of research integrity climate differ between academic ranks and disciplinary fields: Results from a survey among academic researchers in Amsterdam. PLoS One. 2019;14(1):e0210599.
32. Haven TL, Tijdink JK, Pasman HR, Widdershoven G, Riet G, Bouter LM. Researchers' perceptions of research misbehaviours : a mixed methods study among academic researchers in Amsterdam. Res Integr Peer Rev. 2019;4(25):1–12.
33. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale, N.J.: Erlbaum; 1995.
34. Csiszar A. The catalogue that made metrics, and changed science. Nature. 2017;551(7679):163–5.
35. Glausiusz J. Tenure derailed. Nature. 2019;565:525–7.
36. Levecque K, Anseel F, De Beuckelaer A, Van der Heyden J, Gisle L. Work organization and mental health problems in PhD students. Res Policy. 2017;46(4):868–79.
37. Panger G, Tryon J, Smith A. Graduate Student Happiness & Well-Being Report. 2014.
38. Cook C, Heath F, Thompson R. A meta-analysis of response rates in web-or internet-based surveys. Educ Psychol Meas. 2000;60(6):821–36.
39. Phillips AW, Reddy S, Durning SJ. Improving response rates and evaluating nonresponse bias in surveys: AMEE Guide No. 102. Med Teach. 2016;38(3):217–28.
40. Kahn RL, Wolfe DM, Quinn RP, Rosenthal RA. Organizational stress: Studies in role conflict and ambiguity. New York, SE.: Wiley; 1964.
41. Shore C. Beyond the multiversity: Neoliberalism and the rise of the schizophrenic university. Soc Anthropol. 2010;18(1):15–29.
42. Bal E, Grassiani E, Kirk K. Neoliberal individualism in Dutch universities: Teaching and learning anthropology in an insecure environment. Learn Teach. 2014;7(3):46–72.

43. Lorenz C. If you're so smart, why are you under surveillance? Universities, neoliberalism, and New Public Management. Crit Inq. 2012;38(3):599–629.
44. Zare RN, Winnacker EL. China's science funding. Science (80-). 2011;334(6055):433.
45. Vaesen K, Katzav J. How much would each researcher receive if competitive government research funding were distributed equally among researchers? PLoS One. 2017;12(9):4–6.

Supporting infomation is available on the journal's website:
https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0217931


5


6

Chapter 6

Researchers' perceptions of research misbehaviors: a mixed methods study among academic researchers in Amsterdam

This chapter has been published as:
Haven TL, Tijdink JK, Pasman HR, Widdershoven G, Riet G, Bouter LM. Researchers' perceptions of research misbehaviours : a mixed methods study among academic researchers in Amsterdam. *Res Integr Peer Rev. 2019;4(25):1–12. https://doi.org/10.1186/s41073-019-0081-7*

Abstract

Background: There is increasing evidence that research misbehaviour is common, especially the minor forms. Previous studies on research misbehaviour primarily focused on biomedical and social sciences, and evidence from natural sciences and humanities is scarce. We investigated what academic researchers in Amsterdam perceived to be detrimental research misbehaviours in their respective disciplinary fields.
Methods: We used an explanatory sequential mixed methods design. First, survey participants from four disciplinary fields rated perceived frequency and impact of research misbehaviours from a list of 60. We then combined these into a top five ranking of most detrimental research misbehaviours at the aggregate level, stratified by disciplinary field. Second, in focus group interviews, participants from each academic rank and disciplinary field were asked to reflect on the most relevant research misbehaviours for their disciplinary field. We used participative ranking methodology inducing participants to obtain consensus on which research misbehaviours are most detrimental.
Results: In total, 1080 researchers completed the survey (response rate: 15%) and 61 participated in the focus groups (3 to 8 researchers per group). Insufficient supervision consistently ranked highest in the survey regardless of disciplinary field and the focus groups confirmed this. Important themes in the focus groups were insufficient supervision, sloppy science, and sloppy peer review. Biomedical researchers and social science researchers were primarily concerned with sloppy science and insufficient supervision. Natural sciences and humanities researchers discussed sloppy reviewing and theft of ideas by reviewers, a form of plagiarism. Focus group participants further provided examples of particular research misbehaviours they were confronted with and how these impacted their work as a researcher.
Conclusion: We found insufficient supervision and various forms of sloppy science to score highly on aggregate impact throughout all disciplinary fields. Researchers from the natural sciences and humanities also perceived nepotism to be of major impact on the aggregate level. The natural sciences regarded fabrication of data of major impact as well. The focus group interviews helped to understand how researchers interpret 'insufficient supervision'. Besides, the focus group participants added insight into sloppy science in practice. Researchers from the natural sciences and humanities added new research misbehaviours concerning their disciplinary fields to the list, such as the stealing of ideas before publication. This improves our understanding of research misbehaviour, or 'questionable research practices' beyond the social and biomedical fields.

Keywords: *Research misbehaviour, research integrity, disciplinary fields, academic ranks, research misconduct, survey, focus groups*

Background

Most researchers think of themselves as honest and consider their work to be conducted with integrity (1–3). In spite of this, there is increasing evidence that researchers misbehave quite frequently in their work (4–6). Aside from the widely recognized misconducts of falsification, fabrication and plagiarism (henceforth: FFP) there is little evidence on what are perceived to be the most detrimental research misbehaviours (7–9). Besides, it is becoming increasingly clear that research misbehaviours that may seem minor compared to FFP could have a substantial aggregate impact since they occur much more frequently than the 'deadly sins' (10–13).

A meta-analysis of 21 surveys investigating research misbehaviour found that about 2% of researchers admitted to falsification or fabrication. About 34% of participants admitted to Questionable Research Practices (QRP) (4). QRPs embody a large class of research misbehaviours, such as deleting outliers without disclosure. However, since 14 of the 21 studies included in the meta-analysis focused on biomedical researchers, it is unclear whether these proportions generalise to other disciplinary fields.

Similarly, when pooling the results of 17 studies investigating plagiarism, 1.7% of participants admitted to plagiarism (14). However, 10 of those studies used a biomedical sample. Hence, these results may not represent all sciences or the humanities. This also begs the question whether the research misbehaviours that participants were asked about were actually relevant to their own research, as some QRPs may be field or discipline-specific.

We investigated whether the research misbehaviours that are perceived detrimental vary across disciplinary fields. We distinguished four major disciplinary fields in our study: biomedical sciences, natural sciences, social sciences and the humanities. Since FFP are relatively rare, we focus on research misbehaviours that are detrimental on the aggregate level. To get a sense of which research misbehaviours were most detrimental at the aggregate level, we also take the frequency of the research misbehaviour into account. Hence, our study aims to assess what academic researchers in Amsterdam perceive to be detrimental research misbehaviours on the aggregate level in their respective disciplinary fields.

Methods

Design

We used a mixed methods sequential explanatory quantitative first design (15). This implies that our study had two phases: 1) a quantitative phase in which we collected survey data, and 2) a qualitative phase in which we conducted focus group interviews to deepen our understanding of the survey responses. See figure 1.

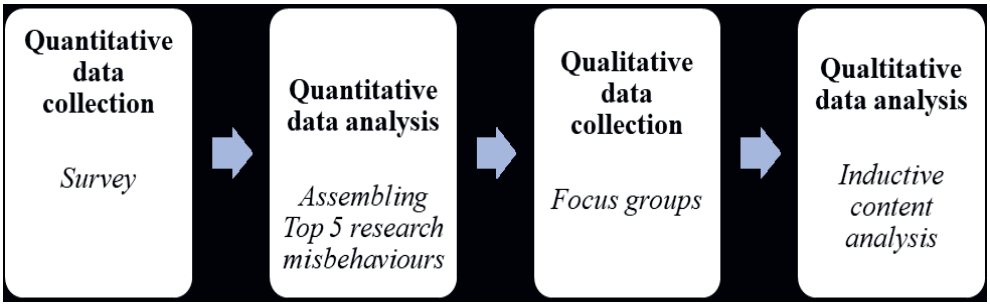

Figure 1. Overview of study design and analysis.

Ethical statement

The Scientific and Ethics Review board of the Faculty of Behavioural and Movement Sciences, Vrije Universiteit, Amsterdam reviewed and approved our study.

Participants

Participants consisted of academic researchers with at least a 0.2 FTE research appointment at the Vrije Universiteit Amsterdam, University of Amsterdam or the Amsterdam University Medical Centers and included PhD candidates.

Materials

We presented participants with research misbehaviours from a list of 60 major and minor misbehaviours as composed by Bouter et al. (11). For a thorough description of the development of the list, the reader is referred to Bouter et al. (11). The list can be found in additional file 1.

In brief, they compiled an extensive list of over 100 research misbehaviours based on the existing literature on research misbehaviours. After removing duplications, 60 items remained which were tested for comprehensibility on 15 researchers. These 60 items were then distributed among keynote speakers and oral presenters of the World Conference on Research Integrity for review. Finally, the list of 60 was used in an invitational workshop at the 4th World Conference on Research Integrity (2015) which provided final input for the phrasing of the items and the relevant response scales. The list was developed and used by us in English.

We used two response scales from the initial list: frequency and impact, respectively. We altered these response scales slightly by specifying the time frame or unit respondents had to keep in mind when reading the items. The impact response scale, "How often will this misbehaviour occur?", was changed into (italics stress our changes): "How often have you observed the behaviour stated above *in the last three years*?". This question had to be answered in reference to respondents' main disciplinary field. Answer options were 1 ("Never"), 2 ("Once or twice") and 3 ("Three times or more"). The impact response

scale, "If it occurs, how large will its impact be on the validity of knowledge?", was changed into "If you were to observe this behaviour, how large would its impact to be on the validity of *the findings of the study at issue*?". Responses ranged from 1 ("Negligible") to 5 ("Enormous").

Quantitative data collection procedure

We contacted the deans and rectors from the participating institutions with a request to contact their academic researchers. The institutions shared the contact details of their researchers on the basis of a formal data sharing agreement. To explain the study's aim, we sent all academic researchers in Amsterdam (n = 7548) an information letter. This letter also included a hyperlink to the privacy policy and the study protocol on our project website (see additional files 2 and 3 and www.amsterdamresearchclimate.nl). One week later we sent an invitational email to all researchers. Participants had to give informed consent and confirm that they were actually involved in research for at least one day per week on average (inclusion check) at the beginning of the survey. We used Qualtrics (Qualtrics, Provo, UT, USA) to build the survey.

To reduce the overall length of the survey and decrease the risk of participant fatigue (16), participants were randomly presented 20 out of 60 items from the list by Bouter et al (11). To preclude order effects, the order of presentation of the 20 items was also randomised.

The survey ended with three demographic items: participants' academic rank (PhD student, postdoc, assistant professor, associate professor or full professor), disciplinary field (biomedicine, natural sciences, social sciences and humanities) and gender (male or female). In the remainder of this paper, we distinguish three main groups of academic ranks: PhD students; postdocs and assistant professors; and associate and full professors.

The survey consisted of three parts, one of which was the list of 60 research misbehaviours described here. The remainder comprised two instruments, one about the research climate for integrity (17) and another about the degree of perceived publication pressure (18). The data described here extend our previous findings (17) by identifying the research misbehaviours that are perceived to impact the research climate most.

Quantitative data analysis

We preregistered our analyses on the Open Science Framework, see https://osf.io/x6t2q/register/565fb3678c5e4a66b5582f67. Here we explain the main analyses briefly. First, we calculated the five most frequent and five most impactful research misbehaviours per academic rank and disciplinary field. Second, although falsifying data, fabricating data or committing plagiarism are most detrimental to science, they are relatively rare and therefore it is not useful to overemphasize the importance of FFP. To get a sense of which research misbehaviours were most detrimental at the aggregate level, we followed Bouter et al. (11) and multiplied the impact score of each research misbehaviour with its

perceived frequency. In particular, we use the product score (multiplication) of impact and frequency as a proxy for aggregate impact throughout this manuscript. This metric ranged from 1 (negligible impact/never observed this) to 15 (enormous impact/observed this more than three times). We present these stratified top 5 rankings of detrimental research misbehaviours on the aggregate level below.

Finally, we carried out exploratory analyses to statistically assess whether the top 5 was actually a good representation of impactful research misbehaviours at the aggregate level. These analyses were not preregistered and should be treated as exploratory. Our reasoning was as follows: if a research misbehaviour could have been on #1 on the ranking, it means the research misbehaviour has substantial impact. We thus assessed the bias-corrected bootstrapped 95% confidence intervals around the mean estimates. If there was any overlap between the confidence intervals, we concluded that this research misbehaviour could have also been ranked first. If this was the case, we adjusted the rankings. Second, we used those new rankings to inspect whether there were any differences between disciplinary fields, seeing if the confidence intervals around a mean estimates overlapped between disciplinary fields.

Qualitative data collection

We extended the survey results with focus group interviews. Our aim was twofold. First, we wanted to know whether researchers recognised the top 5 research misbehaviours we identified based on the survey as relevant for their disciplinary field. Second, if they did not recognise (some of) the research misbehaviours, we gave participants of the focus group interviews the opportunity to present and discuss other research misbehaviours that they considered (more) relevant to their disciplinary field.

We organized focus groups with researchers from three academic ranks and four disciplinary fields. These focus groups took place at the Vrije Universiteit, therefore we only invited researchers from the Vrije Universiteit and the Amsterdam UMC (location VUmc) as they were most conveniently located.

We recruited researchers in three ways. First, we wrote to heads of department and asked them to provide e-mail addresses of potentially interested researchers. Second, we used our network of colleagues that work on unrelated topics. Third, we randomly selected researchers from the different academic ranks and disciplinary fields and invited them via e-mail where we explained the purpose of the focus group and asked them to participate. When an invitee abstained from participation (abstaining from participation was mostly due to conflicting schedules, lack of time or other reasons), we invited a new researcher, and so on until we reached a minimum of 4 confirmations per focus group. Note that it could thus be the case that the focus group participants had also participated in the survey that was disseminated nine months prior to the start of the focus groups. Yet, we have no information to quantify this as we did not ask about it specifically.

In total, we conducted 12 focus group interviews between March 2018 and May

2018 with 61 researchers. To encourage participants to speak freely, the groups were homogenous for academic rank and disciplinary field, see table 1.

Table 1. Overview of academic researchers from Vrije Universiteit Amsterdam and Amsterdam UMC location VUmc per focus group.

Academic rank	**Disciplinary field***			
	Biomedicine	*Natural sciences*	*Social sciences*	*Humanities*
PhD students	5 (5)	4 (0)[E]	4 (3)[E]	6 (5)[E]
Postdocs and assistant professors	5 (4)	3 (0)	7 (3)[E]	5 (5)[E]
Associate and full professors	4 (0)	4 (0)	4 (1)[E]	7 (3)[E]
Total	14 (9)	11 (0)	13 (7)	18 (13)

* In brackets is the number of female researchers.
[E] denotes focus groups that were conducted in English.

A facilitator (TH or JT) led the focus groups, accompanied by an observer who made notes and ensured audiotape recording. We constructed a topic guide to direct the focus group interviews (see additional file 4) where we presented participants with the aggregated impact top 5 of research misbehaviours that we had found in the survey among researchers from their disciplinary field. We then asked participants to add new research misbehaviours that were, in their opinion, at least as relevant to their disciplinary field. As a restriction, we asked all researchers to focus on things they had actually experienced or observed, instead of something they had only heard of or read about.

We used a participative ranking method to structure the focus group discussion about the research misbehaviours. The procedure of the participative ranking method involved three steps. First, participants were presented with the 5 research misbehaviours that ranked highest on aggregate impact on post-its. Second, they were asked to reflect on the relevance of these behaviours for their disciplinary field and prompted to add new behaviours that we may have missed but that participants considered more relevant for their disciplinary field. All research misbehaviours were written down on post-its. Finally, participants were asked to reach consensus over a ranking of all the research misbehaviours. For that, we had created a provisional severity continuum/scale that ranged from 'Minor' to 'Major'. When participants agreed on where each post-it had to be placed on the severity scale, we ended the exercise. In total, this took between 20 and 35 minutes. The remaining results of the focus groups will be part of another report. For an elaborate description of participative ranking methodology, the reader is referred to the guide by Ager, Stark & Potts (19).

Qualitative data analysis

We read the transcripts and started open coding using ATLAS.ti 8.3.0 for Mac Version. If the transcripts were in Dutch, we assigned English codes for consistency and translated quotations. We used inductive content analysis to analyse the transcripts as it is a good method for systematically describing and understanding complex phenomena (20) and it helps to reduce rich data to meaningful concepts that capture the phenomenon of interest (21).

The themes reported below are based on the qualitatively ranked research misbehaviours according to severity as well as the transcripts of the focus group conversations. Specific research misbehaviours, e.g. "reviewing without feedback, harsh reviewing, reviewers not up to scratch with developments" were clustered into broader issues, e.g. "sloppy reviewing". For issues to be identified as emerging themes, the issue had to be related to the research question that involved *research* misbehaviours. Therefore, some issues that focused on political intricacies or personal integrity were disregarded. Moreover, it should be either mentioned multiple times, or during the conversation be discussed as important and powerful.

Team members (JT, TH, GW and RP) independently identified themes and these were discussed to achieve consensus and thereby increase reliability. See additional file 5 for our code tree. Finally, we identified appropriate quotes to illustrate each theme.

Results

Quantitative results

Ninety-two e-mail addresses were no longer in use and 146 researchers filled in the non-response questionnaire. Hence 7310 potential respondents were left, of which 1080 researchers completed the 60 items. Survey completion rate was 15% (see figure 2). First, we present the quantitative top 5 of detrimental research misbehaviours on the aggregate level per disciplinary field. Second, we provide the relevant themes from the focus groups that shed more light on what these research misbehaviours mean and illustrate these with quotes.

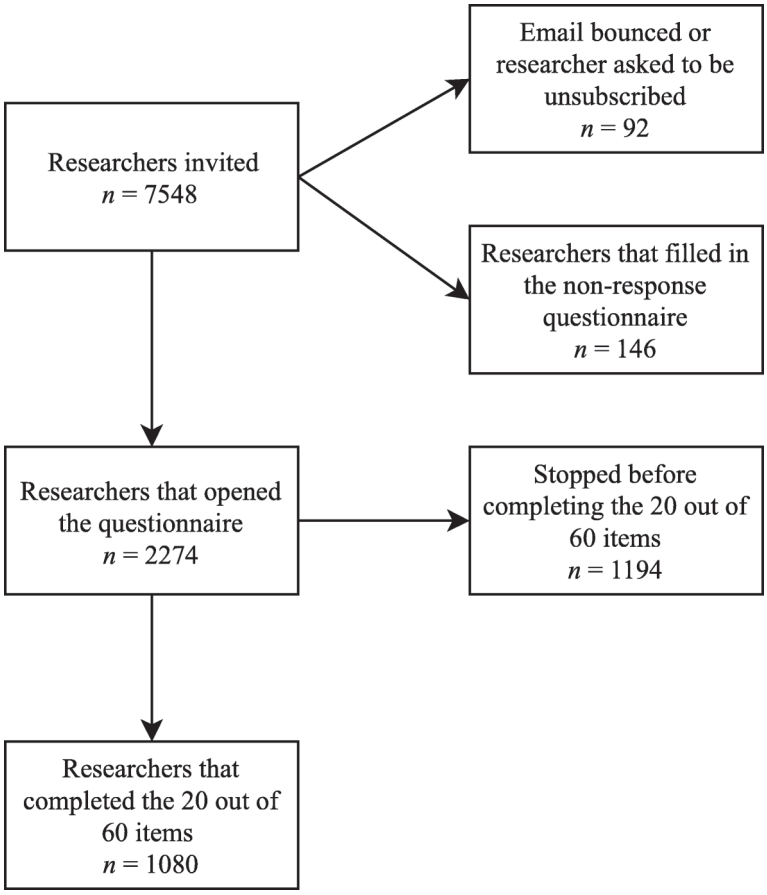

Figure 2. Overview of survey response rate.

Disciplinary fields

A detailed description of the top 5 most frequent and most impactful research misbehaviours per disciplinary field can be found in additional files 6 and 7. The top 5 detrimental research misbehaviours on the aggregate level stratified per rank can be found in additional file 8. Finally, a stratified ranking of all 60 items can be found in additional file 9.

Briefly, the misbehaviour 'fabrication of data' qualified as the most impactful for the validity of the study in all disciplinary fields. Most frequent research misbehaviours differed somewhat. Biomedical researchers perceived listing an author that doesn't qualify for authorship to be most frequent. According to natural sciences researchers and social sciences researchers, insufficient supervision was most frequent. Researchers in the humanities perceived selective citation to be most frequent. Humanities researchers

rated the presentation of grossly misleading information in a grant application as having most impact.

In this paper, we focus on the top 5 of most detrimental research misbehaviours on the aggregate level per disciplinary field, see table 2.

Table 2. Top 5 detrimental research misbehaviours on the aggerate level by disciplinary field.

	Top 5 research misbehaviours per disciplinary field with *M (SD)*							
	Biomedicine		**Natural sciences**		**Social sciences**		**Humanities**	
#1	Insufficiently supervise or mentor junior co-workers	7.02 (3.63)	Insufficiently supervise or mentor junior co-workers	7.72 (4.13)	Insufficiently supervise or mentor junior co-workers	6.95 (3.78)	Insufficiently supervise or mentor junior co-workers	6.76 (3.84)
#2	Choose a clearly inadequate research design or using evidently unsuitable measurement instruments	6.04 (3.16)	Not report clearly relevant details of study methods	6.95 (3.43)	Not publish a valid 'negative' study	6.54 (3.98)	Use published ideas or phrases of others without referencing	6.69 (3.69)
#3	Let own convictions influence the conclusions substantially	5.99 (3.17)	Insufficiently report study flaws and limitations	6.64 (3.41)	Let own convictions influence the conclusions substantially	5.86 (2.95)	Selectively cite to enhance own findings or convictions	6.17 (3.25)
#4	Give insufficient attention to the equipment, skills or expertise which are essential to perform the study	5.64 (3.32)	Let own convictions influence the conclusions substantially	6.38 (3.27)	Choose a clearly inadequate research design or using evidently unsuitable measurement instruments	5.77 (3.38)	Choose a clearly inadequate research design or using evidently unsuitable measurement instruments	6.11 (3.37)
#5	Keep inadequate notes of the research process	5.62 (2.96)	Give insufficient attention to the equipment, skills or expertise which are essential to perform the study	6.26 (3.48)	Give insufficient attention to the equipment, skills or expertise which are essential to perform the study	5.71 (3.3)	Unfairly review papers, grant applications or colleagues applying for promotion	6.03 (4.15)

Exploratory analyses

The following analyses were not preregistered and should be treated as exploratory. We wanted to assess the precision of our mean estimates in Table 2. In what follows, we use bias-corrected bootstrapped 95% confidence intervals around the mean estimates.

In biomedicine, 'Insufficient supervision' ranked #1 and inspection of the confidence intervals indicated that no other misbehaviour could be ranked highest. There was no overlap between the confidence interval around the mean estimate for 'Insufficient supervision' and the confidence intervals of the research misbehaviours listed second and third. For the natural sciences, the confidence interval around 'Insufficient supervision' overlapped with confidence intervals up to misbehaviours ranked twelfth. The top 12 for natural sciences can be found in additional file 10. Besides sloppy science, the top 12 for natural sciences also listed data fabrication (#7) and nepotism (#11). In the social sciences, the confidence interval around 'Insufficient supervision' overlapped with the confidence intervals up to the misbehaviour ranked sixth, see additional file 10. The confidence interval around 'Insufficient supervision' in the humanities overlapped with research misbehaviours up to rank #12. Hence, the top 12 for the humanities can be found in additional file 10. Besides sloppy science, the top 12 for researchers in the humanities included nepotism (#6).

To see if the updated rankings differed between disciplinary fields, we again inspected the confidence intervals around the mean estimates. Biomedical sciences perceived 'Insufficient supervision' to have the greatest impact on the aggregate level, but this was not different from other fields. For the natural sciences, 'Not report clearly relevant details of study methods' ranked second (*CI*: 5.93 – 7.93). However, this rank differed significantly from the two other main disciplinary fields, i.e. the natural sciences perceived this to have a greater impact on the aggregate level than both biomedical researchers (#12, *CI*: 4.69 – 5.43) and researchers in the humanities (#51, *CI*: 2.88 – 3.97). In addition, insufficient attention to the expertise to perform the study (#5, *CI:* 5.23 – 7.36) ranked higher on aggregate impact in for natural sciences compared to the humanities (#36, *CI*: 3.03 – 4.9). Lastly, the presentation of grossly misleading information in a grant application (#9, *CI*: 4.54 – 6.5) as of greater impact than researchers in the social sciences (#47, *CI*: 3.11 – 4.00) and the biomedical sciences (#36, *CI*: 3.76 – 4.22).

For the social sciences, not publishing a negative study ranked second (CI: 5.71 – 7.29) and social science researchers were significantly more concerned about this than their colleagues in the humanities (#25, *CI*: 3.5 – 5.00). In addition, insufficient attention to the expertise to perform the study (#5, *CI:* 5.06 – 6.42) ranked higher on aggregate impact in for social sciences compared to the humanities (#36, *CI*: 3.03 – 4.9). Also, 'Reporting an unexpected finding as being hypothesized from the start' (#6, *CI*: 4.94 – 6.25) was perceived as having a greater impact on the aggregate level by social science researchers compared to researchers in the natural sciences (#34, *CI*: 3.24

– 4.83) and the biomedical sciences (#17, *CI:* 4.28 – 4.92).

Researchers in the humanities indicated selective citation to please editors, reviewers and colleagues (#5, *CI*: 5.13 – 7.03) to have more impact on the aggregate level compared to biomedical researchers (#23, *CI*: 4.11 – 4.78). Lastly, researchers in the humanities perceived the use of published ideas or phrases of others (#12) as of greater impact than biomedical researchers (#49, *CI*: 3.29 – 3.85) and the natural sciences (#36, *CI:* 3.09 – 5). All other comparisons between fields were nonsignificant, see additional file 11.

Qualitative results

From our qualitative analysis, the majority of research misbehaviours fell into one of three broad categories: issues around peer review, sloppy conduct of research and insufficient supervision. To better understand what sort of research misbehaviours researchers from a particular disciplinary field were confronted with and how these impacted their research, we zoomed in on themes that were more specific for a disciplinary field or that received more attention in their discussions. We present these themes per disciplinary field and, where possible, we identified quotes as illustrations, see table 3 below. The rankings of research misbehaviours per focus group be found in additional file 12.

Biomedicine: delaying reviewers, sloppy reporting and insufficient supervision

Biomedical researchers were vexed about inflexible reviewers that either delayed the publications of their findings or that were unresponsive to valid counterarguments in rebuttal letters when the manuscript challenged the mainstream view in the field. This made it particularly hard to publish negative research findings, whereas focus group participants agreed that this was pivotal for knowledge to progress.

Another hindrance for knowledge to progress was that authors drew (wrong) conclusions based on little solid argumentation or seemed to interpret the data as it suited them. This was especially pertinent when only the most positive findings were reported, that then led to replication problems because the positive result was likely obtained by chance.

Insufficient supervision was a concern that participants recognised but they also indicated that a PhD student is expected to ask for help when in need. In addition, the supervisor can make the PhD student aware of existing time pressures, but this should be realistic and not indicate that PhD students are not allowed to take holidays. Finally, it was generally agreed upon that little supervision is not a sufficient condition for irresponsible research, yet it could increase the chances of a PhD student conduct irresponsible research.

Natural sciences: Review misconduct and no team spirit

Natural scientists brought up the topic of review misconduct. The misconduct takes the form of the editor or reviewer stealing ideas when put forth for publication or in a grant proposal. Reviewers or editors could either postpone publication and quickly publish the idea themselves or they could reject the manuscript and publish its ideas elsewhere. A similar story holds for grant proposals.

Natural science researchers also noted that insufficient communication and supervision may damage team relations and some researchers may fail to put their success into context, claiming that the success is only theirs.

Social sciences: sloppy reviewing, sloppy design and statistics, and insufficient supervision

Social science researchers often encountered reviewers that demanded to be cited, which is obviously not the purpose of the review. Furthermore, they encountered reviewers that were not up to scratch with developments in the field. Lastly, some had experience with reviewers that failed to declare a conflict of interest as they had a previous relationship with the authors, revealing nepotism in publication review.

Another concern was the sloppy methods where researchers referred to conducting an underpowered study or failing to report a non-replication. Related was the use of 'harking' (hypothesizing after results are known), where supervisors encouraged their PhD students to present an unexpected finding as being hypothesised from the start (22). Other examples involved collecting more data when results were almost significant or just pressing PhD students into 'finding' an effect in the data, even when probably no actual effect was there.

Finally, concerns were voiced about insufficient supervision of PhD students. More senior researchers noted that PhD students were being held for their academic projects at a very early stage in their career, when a PhD student is still learning what academic research involves. Sometimes supervisors took advantage of their PhD students, either by demanding co-authorships without a justification or by mentally intimidating their PhD students.

Humanities: uncritical reviewing, mediocre research and scarce supervision

Uncritical reviewing was a concern of researchers from the humanities. That could involve a reviewer reviewing without specific comments, or reviewers that just accept a paper because of the authority of the author. This could be due to the fact that peer reviewing is not valued high enough by the scientific community. Another form of uncritical reviewing regarded failure to filter out fake papers that were clearly a hoax. Participants connected this to the fact that some fields lack clear publication criteria that a reviewer can use to judge a manuscript's potential.

A second worry regarded mediocre research, which could mean research that is not value free, opaque or hastily written up, repetitive and inflating small findings. A related research misbehaviour was the stealing of original ideas from colleagues but also stealing ideas from PhD or master students and publishing it without (even) acknowledging them.

Finally, humanities researchers noted scarce supervision could lead to fraud. 'Scarce' could be in terms of quantity; there are very few postdocs and hence there is no day-to-day supervision of PhD students. 'Scarce' could also refer to the quality of supervision, such as when supervisors do not take their responsibility seriously, or when supervisors who are actually not an expert on the topic of the PhD student are assigned to be their supervisor.

Table 3. Quotations per disciplinary field to illustrate the content of the research misbehaviour themes.

	Biomedicine
Theme	Quote
Sloppy reporting	"Take things that are reported as a decrease of 80% in tumour rate. Well, when you attempt to repeat the experiment you get a 60% decrease so obviously their 80% was the most positive result from all the times they tried…" – *PhD candidate*
Insufficient supervision	"If you have a PhD candidate and you completely throw her in at the deep end, surely you increase the chance of irresponsible research" – *Full professor*
Inflexible reviewers	"So everything that is novel or different, it requires an lot of effort to get that accepted in the, in the journals, due to most likely also rigid reviewers" – *Assistant professor*
	Natural sciences
Theme	Quote
Review misconduct	"I had it once with a journal editor who was being really difficult about a publication of mine. And then he managed to get his own publication [with the same idea] in before mine" - *Full professor*
Team spirit missing	"Research is no one man show, you have to teach them [PhD candidates] to also let go, it is not just theirs. The same holds for what I do, it is not just mine, it is a team effort..." – *Assistant professor*
	Social sciences
Theme	Quote
Sloppy reviewing	"You're on a grant review panel and you're judging someone whom you have a personal or professional relationship with. You're an editor of a journal and you don't recuse yourself for a conflict of interest with the author of a paper" – *Associate professor*
Sloppy methods and statistics	"What is so horrible about these strategies is, post-hoc storytelling, salami slicing, is how you win the game, this is how you become a professor, this what you should do. Some professors even tell you, like: this is what you should do" – *Postdoctoral researcher*
Insufficient supervision	"Supervisors exploiting their PhD students. I think that can be sort of extended into any type of harassment; sexual, personal, mental harassment, whatever it is. Also about any type of power relationship that there is and… that he demands co-authorships, that supervisors say… I want to be on this paper, I am on this paper, not as a question but, you know, as a statement..." – *postdoctoral researcher*
	Humanities
Theme	Quote
Uncritical reviewing	"What you see is that, there is no review culture, in which the standards of what constitutes good and bad publications are adequately present, to filter out actual hoaxes" - *Full professor*
Lack of supervision	"I have a PhD student who got sent to me from abroad… I said well, when did you last speak to your supervisor? And he said no, no, no, because you can answer my questions better, the last couple of months I didn't, because I was saving it up for you… While the actual supervision who will get… the credits is actually not an expert." – *Assistant professor*

* Brackets added by the authors.

Discussion

This mixed-method study, involving a survey followed by focus groups, aimed to develop insight into what academic researchers in Amsterdam from different disciplinary fields considered to be the most detrimental research misbehaviours. There are a few important takeaways from our study. First, based on the survey results, we found insufficient supervision and various forms of sloppy science to score highly on aggregate impact throughout all disciplinary fields. Researchers from the natural sciences and humanities also perceived nepotism to be of major impact on the aggregate level. The natural sciences regarded fabrication of data of major impact as well. The focus group interviews helped us to understand how researchers interpret 'insufficient supervision'. Besides, the focus group participants added insight into sloppy science in practice. Second, researchers from the natural sciences and humanities added new research misbehaviours concerning their disciplinary fields to the list, such as the stealing of ideas before publication. This improves our understanding of research misbehaviour, or 'questionable research practices' beyond the social and biomedical fields.

When comparing our findings to the literature, it is important to keep in mind that our findings are not prevalence estimates. Equating the self-reported proportion of a research misbehaviour with its prevalence has been criticised, see Fielder & Schwarz (23). Moreover, in our survey we asked respondents to report how often they had *witnessed* a particular research misbehaviour, not how often they had engaged in such behaviour themselves. We then combined this with the degree of impact respondents assigned to that item to obtain the 'aggregate impact'. Because our aggregated impact metric is the product of impact (1-5) and frequency (1-3), one may wonder if we deliberately assigned impact more weight. Although this is true for the absolute score, this is not the case for the ranked aggregate impact product scores since the rank of a particular research misbehaviour does not change after recoding the impact scale.

Somewhat surprising is the consistent recognition of insufficient supervision and mentoring. We would like to reiterate that we regard insufficient supervision a research misbehaviour in itself. Like many other research misbehaviours, insufficient supervision describes non-compliance with one of the professional norms in academic research (adequate mentoring).

Yet, it seems plausible that insufficient supervision could, in some cases, lead to the supervisees unintentionally engaging in sloppy science because they were not socialised well into responsible conduct of research (24). However, we believe that the influence of insufficient supervision may go further. If a supervisor fails to create a safe learning climate, this could lead to situations where PhD students do not feel confident to share their concerns about a mistake (e.g. in the data-analysis) or to oppose their supervisor's interpretation. Similarly, Roberts and colleagues (25) put forth the speculation that when the supervisor creates an environment where only spectacular outcomes

are valued, supervisees may engage in sloppy science because that yields the desired outcomes. Nevertheless, in our study we did not investigate the possible reasons for research misbehaviours and investigating this would require a different research design.

The amount of literature on supervision and mentoring differs between disciplinary fields. Mentoring received extensive attention in medicine (26,27) and substantial attention in psychology (28). Mentoring and supervision have primarily been used as tools to foster diversity by encouraging minority groups to stay in science and engineering fields (29,30), but received little attention in themselves. One exception is a study by Green & Bauer (31) that linked mentoring to science students' success. In the humanities, mentoring was coined as a way to improve the workplace culture (32). Interestingly, in our study, participants from the humanities expressed concerns about the lack of supervision altogether, or a supervisor who is in fact not an expert in the field. Natural sciences researchers recognised this, but added that bad mentoring or a supervisor mentoring too many PhD candidates can make group relations sour and ultimately slow down research.

Strengths

Our study may be the first that investigates research misbehaviours and includes researchers from different disciplinary fields and all academic ranks. It is noteworthy that the different methods we used (quantitative survey and qualitative focus groups) led to similar results as both survey and focus group participants recognised sloppy science and insufficient supervision as relevant.

Additionally, our quantitative results largely confirm the findings by Bouter and colleagues (11). Their population consisted of visitors of the World Conference of Research Integrity, but apparently both groups identify insufficient supervision and sloppiness as problems in contemporary academia.

Limitations

There are some study limitations to bear in mind. We had considerable non-response. However, response bias is not a necessary consequence of a low response rate as long as the respondents are representative for the population (33). We assessed the representativeness of our sample in two ways. First, we looked at our population that consisted of academic researchers in Amsterdam from two universities and two university medical centres. Those two university medical centres comprised 53% of the population. Biomedical researchers constituted 56% of our sample, indicating a small overrepresentation. Second, we compared our sample to national statistics on researchers in The Netherlands. As there are no national statistics on academic researchers in biomedicine, we filtered biomedical researchers out of our sample for this comparison. National statistics indicate that 32% of researchers work in the natural sciences, 41% work in social sciences and 27% in the humanities. In our sample, we find 25% of

researchers to work in the natural sciences, 51% in the social sciences and 23% in the humanities. This indicates a moderate overrepresentation of the social sciences researchers and a slight underrepresentation of researchers in the natural sciences and humanities.

In addition, a high number of respondents that started answering the survey questions stopped before completing the 20 items. Before respondents were presented a random selection of 20 randomized items, they completed the Survey of Organisational Research Climate (henceforth: SOuRCe© (34)). The number of participants that 'started' the survey included all researchers that opened the survey, even those who decided not to participate. In total, 18% of our invitees completed the SOuRCe© and the later dropout rate of 3% during a survey questionnaire lies within the normal range (35).

A further limitation is that we presented participants with a random selection of 20 research misbehaviours because we feared that presenting them the full list of 60 would be too time-consuming. This type of design is sometimes called missingness by design, as all participants have missing values for some items. Based on similar surveys in the field, we estimated our response rate to be at least 15%. Since our population consisted of 7548 researchers, 15% of them answering 1/3 of our items would mean at least 300 responses per item. Initially we expected more than 300 responses would be sufficient to compute reliable standard deviations, standard errors and confidence intervals.

Unfortunately, a quick glance at the width of the standard deviations in Table 2 revealed that the distribution of our scores was not normal. In fact, more than 90% of the aggregated impact variables have a skewed distribution. Consequently, we must be careful in the interpretation of the top 5. The ranking is purely based on point estimates. In fact, labelling the ranking as a top 5 may be dangerous as "top" suggest that the #1 misbehaviour ranks absolutely higher than #2. Based on our explorative analyses, it can be concluded that this only holds for biomedicine, see additional file 11. The top 5 presented in Table 2 simply lists 5 research misbehaviours that were impactful on the aggregate level and one should not overinterpret differences in the places on the list.

Another limitation regards the interpretation of aggregate impact. Participants did not rate the research misbehaviours to have major impact on the aggregate level, but we used the product of the perceived frequency of a research misbehaviour and the potential impact on the validity as a proxy for aggregate impact. Hence, we labelled these scores as 'aggregate impact' scores. The validity of this metric has no exact (mathematical) justification but is intuitively similar to e.g. the well-known QALY (Quality-Adjusted LifeYear) metric, which multiplies the subjective quality score of a state of living by the time spent in that state (36). In the focus groups, we explicitly asked whether research misbehaviours had actual impact. As the focus groups in general confirmed the results of the survey, our notion of 'aggregate impact' is supported by the qualitative findings.

Furthermore, since the list of 60 research misbehaviours is not formally validated, it remains possible that survey items were unclear to participants. Nevertheless, the

list was tried out at length through workshops and other types of informal review. Yet, especially researchers from the natural sciences and humanities mentioned research misbehaviours that seemed missing or at least substantially different from the list of 60, such as referees or editors that abuse their power to steal original ideas. Properly assessing the relevance of these new items would require translating the qualitative data into items and a representative sample from all disciplinary fields. To facilitate such an attempt, we provide an updated list of research misbehaviours (additional file 13) in which items are reformulated, included as explanatory examples or added as new research misbehaviours. Validation of such a list could be an avenue for further research.

Finally, note that we explicitly asked respondents to focus on research misbehaviours that they had witnessed themselves, so this could decrease the generalisability of our findings so that they might not even apply to the population of academics in Amsterdam. Nevertheless, since sloppy science and insufficient supervision were recognised by academic researchers across disciplinary fields, it seems plausible that these research misbehaviours concern researchers outside Amsterdam as well.

Implications

Since we found insufficient supervision to be recognised across fields, it may be worth exploring interventions that foster responsible supervision and mentoring. The connection between mentoring and responsible research may seem novel. Yet, Whitbeck (37) described an innovative type of group mentoring created to strengthen supervisors in discussing research integrity and to support research groups in comprehending the variety of integrity challenging situations they may encounter. More recently, Kalichman & Plemmons (38,39) described a workshop curriculum for supervisors and faculty to convey responsible research in the actual research environment. Training programs like these are a step forward in making responsible supervision the norm.

Conclusion

We found insufficient supervision and various forms of sloppy science to score highly on aggregate impact across disciplinary fields. Researchers from the natural sciences and humanities also perceived nepotism to be of major impact on the aggregate level. The natural sciences regarded fabrication of data of major impact as well. The focus group interviews helped us to understand how researchers interpreted 'insufficient supervision'. Researchers from the natural sciences and humanities added new research misbehaviours concerning their disciplinary fields to the list, such as the stealing of ideas before publication. This improves our understanding of research misbehaviour, or 'questionable research practices' beyond the social and biomedical fields.

Acknowledgements

We would like to thank all members of the steering committee of the Academic Research Climate Amsterdam project (professor F. Oort, professor Y. Smulders, professor H. de Haes and professor R. van Woudenberg) for their critical comments during presentations on the study design or draft communications regarding the study or interpretation of the data. Besides, we would like to acknowledge Peter van de Ven and Birgit Lissenberg-Witte for their statistical support. In addition, Michiel de Boer has been indispensable in discussing the exploratory analyses. Finally, Natalie Evans's excellent proofreading of this manuscript has significantly improved its overall quality.

References

1. Veldkamp CLS, Hartgerink CHJ, van Assen MALM, Wicherts JM. Who believes in the storybook image of the scientist? Account Res. 2017;24(3):127–51.
2. Mahoney MJ. Psychology of the scientist: An evaluative review. Soc Stud Sci. 1979;9(3):349–75.
3. Mahoney MJ. Scientist as subject: The psychological imperative. Cambridge, MA SE: Ballinger Publishing Company; 1976.
4. Fanelli D. How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. PLoS One. 2009;4(5):e5738.
5. Martinson BC, Anderson MS, de Vries R. Scientists behaving badly. Nature. 2005;435(7043):737–8.
6. Kornfeld DS. Perspective: Research misconduct: The search for a remedy. Acad Med. 2012;87(7):877–82.
7. Godecharle S, Nemery B, Dierickx K. Guidance on research integrity: No union in Europe. Lancet. 2013;381(9872):1097–8.
8. Bedeian A, Taylor S, Miller A. Management science on the credibility bubble: Cardinal sins and various misdemeanors. Acad Manag Learn Educ. 2010;9(4):715–25.
9. National Academies of Sciences, Engineering and M. Fostering Integrity in Research. 2017.
10. Bouter LM. Commentary: Perverse incentives or rotten apples? Account Res. 2015;22(3):148–61.
11. Bouter LM, Tijdink J, Axelsen N, Martinson BC, ter Riet G. Ranking major and minor research misbehaviors: results from a survey among participants of four World Conferences on Research Integrity. Res Integr Peer Rev. 2016;1(17):1–8.
12. Steneck N. Fostering integrity in research: Definition, current knowlege, and future directions. Sci Eng Ethics. 2006;12(1):53–74.
13. De Vries, Raymond; Anderson, Melissa; Martinson B. Normal misbehavior: Scientists talk bbout the ethics of research. J Empir Res Hum Res Ethics. 2006;1(1):43–50.
14. Pupovac V, Fanelli D. Scientists admitting to plagiarism: A meta-analysis of surveys. Sci Eng Ethics. 2015;21(5):1331–52.
15. Creswell JW. Research design: Qualitative, quantitative, and mixed method approaches. 4th ed. Los Angeles: Sage; 2014.
16. Graham JW, Taylor BJ, Olchowski AE, Cumsille PE. Planned missing data designs in psychological research. Psychol Methods. 2006;11(4):323–43.
17. Haven TL, Tijdink JK, Martinson BC, Bouter LM. Perceptions of research integrity climate differ between academic ranks and disciplinary fields: Results from a survey among academic researchers in Amsterdam. PLoS One. 2019;14(1):e0210599.
18. Haven TL, Bouter LM, Smulders YM, Tijdink. Perceived publication pressure in Amsterdam – survey of all disciplinary fields and academic ranks. PLoS One. 2018;14(6):e0217931.
19. Ager A, Stark S, Potts A. Participative ranking methodology: A brief guide. Mailman School of Public Health. 2010.
20. Krippendorff K. Content analysis: An introduction to its methodology. Beverly Hills SE: Sage Publications; 1980.

21. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs. 2008;62(1):107–15.
22. Kerr NL. HARKing: Hypothesizing after the results are known. Personal Soc Psychol Rev. 1998;2(3):196–217.
23. Fiedler K, Schwarz N. Questionable research practices revisited. Soc Psychol Personal Sci. 2016;7(1):45–52.
24. Anderson MS, Horn AS, Risbey KR, Ronning EA, De Vries R, Martinson BC. What do mentoring and training in the responsible conduct of research have to do with scientists' misbehavior? Findings from a national survey of NIH-funded scientists. Acad Med. 2007;82(9):853–60.
25. Roberts GC, Kavussanu M, Sprague RL. Mentoring and the impact of the research climate. Sci Eng Ethics. 2001;7(4):525–37.
26. Sambunjak D, Straus SE, Marušić A. Mentoring in academic medicine: A systematic review. J Am Med Assoc. 2006;296(9):1103–15.
27. Liénard JF, Achakulvisut T, Acuna DE, David S V. Intellectual synthesis in mentorship determines success in academic careers. Nat Commun. 2018;9(1):4840.
28. Forehand R. The art and science of mentoring in psychology: a necessary practice to ensure our future. Am Psychol. 2008;63(8):744–55.
29. Kendricks, Kimberly D.; Nedunuri, K. V.; Arment AR. Minority student perceptions of the impact of mentoring to enhance academic performance in STEM disciplines. J STEM Educ Innov Res. 2013;14(2):38–46.
30. Wilson ZS, Holmes L, DeGravelles K, Sylvain MR, Batiste L, Johnson M, et al. Hierarchical mentoring: A transformative strategy for improving diversity and retention in undergraduate STEM disciplines. J Sci Educ Technol. 2012;21(1):148–56.
31. Green SG, Bauer TN. Supervisory mentoring by advisers: Relationships with doctoral student potential, productivity, and commitment. Pers Psychol. 1995;48(3):537–62.
32. Metzger AM, Petit A, Sieber S. Mentoring as a way to change a culture of academic bullying and mobbing in the humanities. High Educ Futur. 2015;2(2):139–50.
33. Cook C, Heath F, Thompson R. A meta-analysis of response rates in web-or internet-based surveys. Educ Psychol Meas. 2000;60(6):821–36.
34. Martinson BC, Thrush CR, Crain AL. Development and validation of the Survey of Organizational Research Climate (SORC). Sci Eng Ethics. 2013;19(3):813–134.
35. Manfreda KL, Batagelj Z, Vehovar V. Design of web survey questionnaires: Three basic experiments. J Comput Commun. 2002;7(3).
36. Sassi F. Calculating QALYs, comparing QALY and DALY calculations. Health Policy Plan. 2006;21(5):402–8.
37. Whitebeck C. Group mentoring to foster the responsible conduct of research. Sci Eng Ethics. 2001;7(4):541–58.
38. Plemmons DK, Kalichman MW. Mentoring for responsible research: The creation of a curriculum for faculty to teach RCR in the research environment. Sci Eng Ethics. 2018;24(1):207–26.
39. Kalichman MW, Plemmons DK. Intervention to promote responsible conduct of research mentoring. Sci Eng Ethics. 2018;24(2):699–725.

Additional files

Additional files are available on the journal's website:
https://researchintegrityjournal.biomedcentral.com/articles/10.1186/s41073-019-0081-7

7

Chapter 7

Explaining variance in perceived research misbehavior

Under review

Abstract

Background: Concerns about research misbehavior in academic science have sparked interest in the factors that may explain such misbehavior. Often, three clusters of factors are distinguished: individual factors, climate factors and publication factors. Our research question was: to what extent can individual, climate and publication factors explain the variance in frequently perceived research misbehaviors?
Methods: From May 2017 until July 2017, we conducted a survey study among academic researchers in Amsterdam. The survey included three measurement instruments that we previously reported individual results of and here we integrate these findings.
Results: 1298 researchers completed the survey (response rate: 17%). Results showed that individual, climate and publication factors combined explained 32% of variance in perceived frequency of research misbehavior. Individual factors explained 7%, climate factors explained 22% and publication factors 16%.
Conclusions: Our results suggest that perceptions of the research climate play a substantial role in explaining variance in research misbehavior. This suggests that efforts to improve departmental norms might have a salutary effect on research behavior.

Key words: research climate, research misbehavior, publication pressure, research integrity

Background

There has long been concern about research misbehavior in academic science (1–4). Research misbehavior includes a broad array of behaviors, some of which may invalidate research results, some that damage trust in science, and others that may deny credit to those to whom credit is due in ways that may hamper their career progression, possibly leading to their exit from the scientific workforce and the loss of highly talented individuals (5). These behaviors range in "severity" or "seriousness" from research misconduct (fabrication, falsification and plagiarism, henceforth RM) to "lesser" forms of misbehavior usually termed questionable or detrimental research practices (henceforth: QRP) (6). These behaviors also differ in their level of intentionality and may be just negligent or reckless, or conscious deviations from the standards for good quality research, with a purpose other than finding true answers.

Explanations for why researchers misbehave can generally be grouped into three clusters of potentially explanatory factors: those at the level of the individual, factors arising from the organization in which researchers go about their work, and forces that may act upon individual researchers from beyond their immediate workplace - such as the commonly referenced "publish or perish" pressure (7–10).

Examples of individual-related factors are gender and academic rank. Examples of climate factors are perceptions of research-related norms and fairness of supervision, and the quality of resources available to support researchers in their work. Examples of publication system factors are the perceived publication stress among academic researchers and their attitudes towards the current publication system governing academic research.

Previous research has found that male researchers were overrepresented when reviewing RM reports and that junior researchers also seem to be more likely to report QRPs or RM. In addition, researchers are supposedly more likely to misbehave in a climate where they feel treated unjustly and perceive heavy competition. Lastly, RM and QRPs have been associated with high perceived publication pressure (11–13).

Objectives

In this paper, we integrate our previously published findings (14–16) that used measurement instruments that are, at best, proxies for these complex phenomena to see what share of variance in QRPs and RM these three groups of factors account for. We work from the assumption that in a poor-quality research climate with high publication pressure, researchers should be more likely to observe research misbehavior. Our research question is: to what extent can individual, climate and publication factors explain the variance in frequently perceived research misbehaviors?

Methods

Study design

We use a cross-sectional survey design.

Ethics

The Scientific and Ethical Review board of the Faculty of Behavior & Movements Sciences (Vrije Universiteit Amsterdam) approved our survey questionnaire containing three different instruments, approval number: VCWE-2017-017R1.

Participants

Participants were academic researchers employed at two universities in Amsterdam (Vrije Universiteit Amsterdam and University of Amsterdam) and two academic medical centers (i.e., Amsterdam University Medical Centers, location AMC and VUmc). In order to be eligible for participation, respondents had to be employed in research work for at least one day per week. We included PhD candidates, as they are formally employed by Dutch institutions. A full description of our recruitment procedure can be found elsewhere (15).

Variables

The survey questionnaire consisted of three instruments (Survey of Organizational Research Climate, henceforth: SOURCE© (17), the revised Publication Pressure Questionnaire, henceforth: PPQr (18), 20 randomly drawn research misbehaviors from a list of 60 QRPs and RM (5)) and three demographic items (gender, academic rank and disciplinary field). For an overview of the different subscales and items that we used as proxies for the individual, climate and publication factors, see Table 1.

Setting

Between May 2017 and July 2017, we conducted a survey study among academic researchers in Amsterdam. We used Qualtrics (Qualtrics, Provo, UT, USA) to design the survey. The survey started after participants indicated informed consent. The survey included three measurement instruments that we previously reported individual results of and here we integrate these findings.

Study size

We invited the complete population of interest; no specific sample size calculations were made prior to data collection.

Table 1. Overview of instruments used in survey questionnaire.

Construct of interest	Instrument (reference), *interpretation*	# items, (# subscales)	Reliability of scores
Individual factors	Gender (male*/female) Academic rank (PhD student*, postdoc or assistant professor, associate or full professor) Disciplinary field (biomedical sciences, natural sciences, social sciences and humanities*)	3	*Not applicable.*
Climate factors	SOURCE© (17), *The higher the subscale score, the more positive the perceptions of the research climate.*	28 (7)	Cronbach's α ranges from .81 to .87
Publication factors	PPQr (18), *The higher the subscale score, the more negative the perceptions of the publication system.*	18 (3)	Cronbach's α ranges from .75 to .80
Research misbehaviors	List of QRPs and RM (5)	60	Generalizability coefficients for perceived frequency and perceived impact are 0.80 and 0.89 respectively.

[1] If available. * reference category- to ease interpretation, we chose the group with the highest or the lowest score.

Bias

The greatest source of potential bias in our design is response bias, which is why we sent multiple reminders and advertised our study in university newsletters and on the intranet. Still, the choice to participate in a study related to research integrity and misbehavior is presumably not random.

Quantitative variables

Explanatory variables are the demographic characteristics of the participant (we refer to these as individual factors, as they regard characteristics of the individual), SOURCE© subscales, and PPQr subscales.

Outcome variables are (1) perceived frequency (never observed/observed)[1] and (2) perceived impact, the product score of perceived frequency and impact on validity that we henceforth denote as perceived impact[2]. We use perceived impact because focusing on perceived frequency alone may result in a model that explains more trivial trespasses only. We took the square root of this perceived impact score for normalization purposes.

To give the reader an indication of the overall frequency of perceived misbehavior, we

1 This remains an imperfect measure of how often misbehavior actually occurs, as it relies on whether respondents report observing the behaviour in question.

2 We reasoned that the impact of the misbehavior increases as the behavior is more frequently perceived and is assigned a greater impact on the validity. We used the term impact on the aggregate level in another paper, but believe the term perceived impact is more succinct.

calculated percentages of the three possible frequencies. To get a sense of the reliability of our outcome measures, we calculated generalizability coefficients, based on the theory of generalizability developed by Cronbach and colleagues (19). The generalizability coefficient is a function of variance components and can also be estimated with incomplete data.

Statistical methods

For the calculation of the specific subscales and explanatory variables, we refer to the previously published studies (5,17,18) on our survey data. Outcome variables are perceived frequency and perceived impact of misbehavior. Each participant responded to 20 items, randomly selected out of a set of 60 items. As a result, participants responded to different sets of items.

We applied multilevel logistic regression analysis to the perceived frequency item scores and multilevel linear regression analysis to the perceived impact item scores, with items nested within respondents, and the characteristics of the participants as the higher-level variables. Perceived frequency item scores were dichotomized, as the third response option was hardly used (0 = not observed, 1 = observed). The concept of explained variance is not defined in multilevel logistic regression. However, as our application items are first level units and our respondents are second level units, the estimated intercept variance represents between-subject variance (20). We can compare intercept variance in the empty model with intercept variance in models that include explanatory variables, and use unity minus the proportional reduction in intercept variance as an index of explained variance.

Our approach comprised four steps: first, we analyzed the influence of each explanatory variable on the two outcome variables individually. Second, we used a stepwise procedure to assess which cluster of explanatory variables explained most variance (cluster 1, individual factors = gender, academic rank and disciplinary field, cluster 2, climate factors = 7 SOURCE© subscales and cluster 3, publication factors = PPQr subscales). Third, we employed a hierarchical model where we consecutively added the explanatory variables in their clusters – starting with cluster 1 – to assess how much cumulative variance was explained. Finally, we inspected the relationships between the different explanatory variables with Pearson's correlation.

Results

Response rate

We obtained 7548 e-mail addresses of active academic researchers in Amsterdam of which 83 were no longer in use. Some researchers explicitly declined participation (n = 109) and 1298 researchers completed at least one subscale from the SOURCE©, which was sufficient to use their responses in our models, yielding a response rate of 17%.

Descriptive data

Demographic information can be found in Table 2.

Table 2. Participants' demographic information[1].

Gender	*Male*	441
	Female	632
Academic rank	*PhD students*	503
	Postdocs and assistant professors	318
	Associate and full professors	216
Disciplinary field	*Biomedical sciences*	603
	Natural sciences	119
	Social Sciences	242
	Humanities	109

[1] 225 participants did not indicate their demographic information or stopped prematurely.

Outcome data

Percentages of each frequency for all 60 QRPs and RM (as well as for the SOURCE© and PPQr) can be found in Additional file 1.

Main results

We assessed the association of each explanatory variable with both the perceived frequency measure and the perceived impact measure. An overview of these results can be found in Table 3. Note that these are all separate univariate multilevel regression analyses with a single variable in the model (not corrected for any confounders). Individual factors explain between 0-5% of the variance in perceived frequency of research misbehaviors, climate factors explain between 5-18% and publication factors explain between 1-15% of the variance in frequency of research misbehaviors.

Table 3. Explanatory variables for perceived research misbehaviors, univariate analyses.

	Perceived frequency of misbehavior				Perceived impact of misbehavior			
Explanatory variable	β	SE[a]	*p-value*[b]	%Index of Expl.Variane[c]	β	SE[a]	*p-value*[b]	%Expl.Variane[c]
Individual factors								
Female vs male	-.159	.065	.015	.69%	-.018	.035	.609	0%
Academic rank				6.06%				1.08%
Postdoc/assistant professor vs PhD students	.469	.072	<.001		.160	.039	<.001	
Associate/full professor vs PhD students	.617	.081	<.001		.082	.044	.064	
Disciplinary field				.25%				.13%
Biomedical sciences vs humanities	.127	.109	.245		.021	.058	.720	
Natural sciences vs humanities	.110	.139	.430		.041	.074	.581	
Social sciences vs humanities	.035	.121	.771		.065	.064	.312	
Climate factors								
RCR Resources	-.309	.039	<.001	6.01%	-.137	.021	<.001	3.77%
Regulatory Quality	-.463	.050	<.001	5.26%	-.039	.027	.145	.95%
Departmental Norms	-.565	.044	<.001	14.67%	-.243	.024	<.001	8.09%
Integrity Socialization	-.318	.039	<.001	6.31%	-.121	.021	<.001	3.17%
Supervisor-supervisee Relations	-.425	.040	<.001	10.57%	-.168	.022	<.001	5.08%
Integrity Inhibitors	-.542	.039	<.001	17.46%	-.271	.021	<.001	12.98%
Departmental Expectations	-.319	.037	<.001	7.56%	-.160	.020	<.001	5.77%
Publication factors								
Publication Attitude	.580	.045	<.001	14.75%	.284	.024	<.001	11.74%
Publication Resources	.185	.051	<.001	1.44%	.137	.027	<.001	2.28%
Publication Stress	.323	.039	.039	6.39%	.157	.021	<.001	4.92%

[a] = standard error of the β, [b] = associated *p* value [c] = percentage of explained variance by the independent variable at issue.

When using perceived impact as outcome variable, individual factors explain 1% of variance, climate factors between 1-13% and finally publication factors between 2-12% of variance in perceived impact of research misbehaviors.

We added the explanatory variables in their respective clusters and then followed up with a hierarchical model where we consecutively added the clusters, see Table 4. Individual factors as a cluster explain 7% of variance in perceived frequency of research misbehaviors, climate factors as a cluster explain 22% of variance and publication factors as a cluster explain 16% of variance in perceived frequency of research misbehaviors. Individual factors as a cluster explain 1% of variance in perceived impact of research misbehavior, the cluster of climate factors explains 14% and the cluster of publication factors explains 12% of variance in perceived impact of research misbehaviors.

Table 4. Explained variance of clusters of factors using hierarchical mixed modelling.

	Perceived frequency of misbehavior				**Perceived impact of misbehavior**			
Clusters added[1]	*Index of expl. var*[0]	*Cum. explained variance*[1]	*Diff. model fit*[2] *(df)*	*Signif. model fit*[3]	*Expl. var*[0]	*Cum. explained variance*[1]	*Diff. model fit*[2] *(df)*	*Signif. model fit*[3]
Individual factors[a]	6.74%	6.74%	74.1 (*6*)	<.001	1.18%	1.18%	18.1 (*6*)	.001 <*p*< .01
Climate factors[b]	22.22%	31.64%	358.2 (*7*)	<.001	14.10%	15.66%	205.7 (*7*)	<.001
Publication factors[c]	15.85%	34.21%*	32.5 (*3*)	<.001	12.28%	18.42%*	37.6 (*3*)	<.001

[0] = this is the explained variance when only one group of factors is analyzed, i.e. just the climate factors explain 22.22% of variance perceived frequency of research misbehaviors.

[1] = the explained variance here is the cumulatively explained variance. Since the models are hierarchical, factors are added consecutively, i.e. the explained variance is 31.64% when both individual as well as climate factors are added to the model.

[2] = difference in model fit, model fit here is the difference between the -2 Log likelihood of the previous model, i.e. 74 is the difference between the intercept-only model and the model with individual factors added, etc.

[3] = significance of model fit is contrasted with the previous model; the row above or a model with no parameters (vs. individual factors).

[a] = gender, academic rank and disciplinary field, b = SOURCE subscales, c = PPQr subscales.

* Note that the explained total is less than the sum of its parts because there is some overlap in the variance that publication and climate factors explain.

We followed up with a hierarchical model where we consecutively added the clusters. The clusters of individual factors and climate factors combined explain 32% of the variance in perceived frequency of research misbehaviors. Adding all three clusters to the model, hence including publication factors, explains 34% of the variance in perceived frequency of research misbehaviors. When using perceived impact as the outcome variable, individual and climate clusters combined explain 16% of variance in variance

of perceived impact of research misbehaviors. Finally, adding all three clusters to the model explains 18% of variance in perceived impact of research misbehaviors.

Other analyses

Note that publication factors explain little additional variance when climate factors are already in the model, which prompts questions about the relationship between the different explanatory variables. To assess why adding publication factors last to the model had only a marginal effect on the cumulative increase in variance, we calculated Pearson correlation coefficients between the individual factors and the publication factors [see Additional file 2]. We also inspected correlations between the climate factors and publication factors [see Additional file 3]. Overall, correlations between climate and publication factors range from between .27 and .53 and the more positive a participant's perception of the research climate, the less negative that participant's perception of the publication system.

Discussion

Key results

We investigated the extent to which variances in research misbehavior can be explained by individual, climate and publication factors. Overall, individual, climate and publication factors combined explain 34% of variance in perceived frequency of research misbehavior and 18% in perceived impact of research misbehavior. The cluster accounting for the greatest percentage of explained variance is the research climate, 22% and 14% in perceived frequency and perceived impact of research misbehavior, respectively. Publication pressure is the second greatest explanatory variable, accounting for 16% of variance in perceived frequency and 12% of variance in perceived impact of research misbehavior. Individual factors are the smallest cluster, explaining 7% of variance in perceived frequency and 1% in perceived impact.

Interpretation

We found academic rank to play the greatest role within the cluster of individual factors. Previous research coined explanations for the association between academic rank and research misbehavior, including the idea that junior researchers are less familiar with responsible research practices (8), or, when under pressure to perform, they would potentially compromise their ethics (16). However, our results indicate that senior researchers observed significantly more research misbehavior. Hence, perhaps junior researchers are more honest in their self-reporting but when asked about the behavior of others, senior researchers are equally critical of their colleagues.

We found no effect of gender and in fact the influence of individual variables

(such as gender) for research misbehavior has received criticism. For example, Kaatz, Vogelman & Carnes (21) pointed out that males being overrepresented among those found guilty of misconduct and evidence from other areas finding men more likely to commit fraud, are insufficient to conclude that male researchers would be more likely to engage in research misconduct. Besides, Dalton & Ortegren (22) found that the consistent finding that women respond more ethically than men was greatly reduced when controlling for social desirability. The authors note that this does not indicate males and females respond equally ethically, but simply that the differences in ethical behavior may be smaller than initially assumed.

We found the cluster of climate factors to have the greatest share in explaining research misbehavior, which is similar to the findings of Crain and colleagues (23), who found that especially the Integrity Inhibitors subscale (a scale that measures the degree to which integrity inhibiting factors are present, such as the pressure to obtain funding and whether there is suspicion among researchers) was strongly related to engaging in research misbehavior in their sample of US scientists. A high score on the Departmental Norms (the extent to which researchers value norms regarding scholarly integrity in research, such as honesty) subscale was negatively associated with engaging in research misbehavior. When reviewing the individual subscale effects in our study, these two subscale scores are most strongly associated with perceived frequency as well as with perceived impact. Bearing in mind that we focused on perceptions of engagement in research misbehavior by others in the direct environment and not on research misbehavior by the respondent him- or herself, we still think it is reasonable to believe that we observed a similar pattern. In addition, using a large bibliographic sample based on retracted papers, Fanelli, Costas and Larivière (24) reported that academic culture affects research integrity, again emphasizing the importance of this cluster.

Broadly speaking, the relationship we observed aligns with existing literature that investigates unethical behavior in organizations (25). A meta-analysis by Martin and Cullen (26) found that unethical behavior (among which they considered lying, cheating and falsifying reports) was associated with what is called an instrumental climate where individual behavior is primarily motivated by self-interest (27). Related, Gorsira et al. (28) found that when employees perceived their work climates to be more ethical, they were less likely to engage in corrupt behavior and vice versa.

Maggio and colleagues (12) used the previous version of the Publication Pressure Questionnaire and found publication pressure to account for 10% of variance in self-reported research misbehavior among researchers in health professions' education. This is similar to our findings, although the authors focused on self-reported misbehaviors, whereas we focused on perceptions of engagement in research misbehavior by others in the direct environment. In addition, we used a slightly different set of research misbehaviors and we have investigated researchers from other disciplinary fields as well. Nevertheless, both study results indicate that in an environment where perceived

publication pressure is high, the likelihood of researchers reporting research misbehavior will be larger compared to an environment with low publication pressure.

Holtfreter and colleagues (29) used a list of criminological factors that have been associated with research misconduct and asked academic researchers in the US to indicate which factor they thought contributed most to research misconduct. Regardless of their disciplinary field, researchers reported that the stress and strain to perform (among which was the pressure to publish) was the main cause for research misconduct. Holtfreter and colleagues only distinguished two clusters of factors: 'bad apples' (similar to our individual factors) and 'bad barrels', comprising both climate and publication factors. That said, the strain items are rather similar to our publication pressure items, supporting the idea of publication pressure as a factor contributing to research misconduct.

Note that we do not claim that individual, climate and publication factors are independent. We found, for instance, publication pressure to account for 16% of variance in perceived frequency when added as the first variable. However, when climate factors are already in the model, the cumulative increase of explained variance when adding publication pressure is only 2%, which seems intuitive, since it could be that publication factors influence climate factors, such as when increased publication pressure leads to authorship disputes that in turn potentially damage the research climate in particular research groups (13). A related line of reasoning could be that publication pressure may arise as a function of how one's department and departmental expectations for "productivity" are set up, or may arise at a higher organizational level, to the extent that publication expectations are set or influenced by decision makers above the department level.

Generalizability

Our study's sample included researchers from different academic disciplines and academic ranks. The findings thus bear relevance to a broad group of academic researchers. Besides, relying on previously validated and repeatedly employed instruments such as the SOURCE© (17) and PPQr (18) should substantiate the validity of our findings.

Limitations

We should acknowledge a number of weaknesses in our study. Firstly, a response rate of 17% is arguably low. That said, it is not lower than other recent surveys that are considered valid (30). In addition, a low response rate in itself does not indicate a response bias. In another study, we tried to estimate response bias in our sample using a wave analysis and found early responders to be similar to late responders (14). Also, when looking at demographic characteristics, such as academic rank, our respondents seemed similar to the population (15), reducing the concern that our sample is biased, at least with respect to those dimensions. In conclusion, with our response rate, we cannot

exclude the possibility of response bias, but we have some reason to believe it should not influence our results substantially.

Secondly, our outcome variables regard perceived misbehavior by others, whereas many studies into misbehavior focus on self-reports of misbehavior by the respondent, including some of the literature we cited. Interestingly, whereas self-reported rates of misbehavior by the respondent have decreased over time, perceptions of the frequency of misbehavior by others have remained more stable (31). Nevertheless, perceptions of misbehavior measurements may be artificially inflated in situations where various responders have witnessed the same incident. Besides, people are generally more critical when reporting on others' misbehavior (and more lenient when it regards their own), also known as the Mohammed Ali effect (31), which could artificially inflate reported perceptions. Hence, our data may overestimate the actual frequency of perceived research misbehavior. Relatedly, as we measured all outcome and explanatory variables through subjective self-report, the correlations between these variables may be inflated by common-method bias (32). It seems reasonable to say that perceptions carry credible evidence about the 'true' prevalence of research misbehavior and its explanatory variables, although surveying perceptions is by no means conclusive.

Thirdly, the assumption that is implicit in our work is that when participants reported on what research misbehaviors they observed in their field of study, they were largely reporting on what they had observed in their own research setting. Although we do not think this is an unreasonable assumption, we nevertheless want to acknowledge that we could not test it explicitly in our survey.

Fourthly, it is a characteristic of multiple regression that the more explanatory variables within a cluster, the larger the explained variance. This should be kept in mind, as our clusters have different numbers of explanatory variables within them.

Finally, our results are cross-sectional in nature so we have to refrain from any causal conclusions.

Conclusions

Our results suggest that researchers' perceptions of the research climate as well as researchers' perceptions of publication pressure play a significant role in explaining research misbehavior. Especially the norms that govern research practices in a department and the extent to which integrity inhibiting factors such as suspicion were present, explained a large proportion of variance. Finally, it was not so much researchers' publication stress but more their attitudes towards the current publication system that played a substantial role. Note that these proportions of explained variance decreased when using impact as outcome, but the results pattern remained the same. This suggests that efforts to improve departmental norms might have a salutary effect on behavior.

References

1. De Vries R, Anderson MS, Martinson BC. Normal misbehavior: scientists talk about the ethics of research. J Empir Res Hum Res Ethics. 2006;1(1):43–50.
2. Martinson BC, Anderson MS, de Vries R. Scientists behaving badly. Nature. 2005;435(7043):737–8.
3. Stroebe W, Postmes T, Spears R. Scientific misconduct and the myth of self-correction in science. Perspect Psychol Sci. 2012;7(6):670–88.
4. Steneck N. Fostering integrity in research: Definition, current knowledge, and future directions. Sci Eng Ethics. 2006;12(1):53–74.
5. Bouter LM, Tijdink J, Axelsen N, Martinson BC, ter Riet G. Ranking major and minor research misbehaviors: results from a survey among participants of four World Conferences on Research Integrity. Res Integr Peer Rev. 2016;1(17):1–8.
6. Medicine), NASEM (National Academies of Sciences, Engineering A. Fostering Integrity in Research. Washington, D.C.; 2017.
7. Sovacool BK. Exploring scientific misconduct: Isolated individuals, impure institutions, or an inevitable idiom of modern science? J Bioeth Inq. 2008;5(4):271–82.
8. Bogner A, Menz W. Science Crime. The Korean cloning scandal and the role of ethics. Sci Public Policy. 2006;33(8):601–12.
9. George SL. Research misconduct and data fraud in clinical trials: Prevalence and causal factors. Int J Clin Oncol. 2016;21:15–21.
10. Neill US. Publish or perish, but at what cost? J Clin Invest. 2008;118(7):1–2.
11. Tijdink JK, Verbeke R, Smulders YM. Publication pressure and scientific misconduct in medical scientists. J Empir Res Hum Res Ethics. 2014;9(5):64–71.
12. Maggio L, Dong T, Driessen E, Artino A. Factors associated with scientific misconduct and questionable research practices in health professions education. Perspect Med Educ. 2019;8(2):74–82.
13. Tijdink JK, Schipper K, Bouter LM, Pont PM, De Jonge J, Smulders YM. How do scientists perceive the current publication culture? A qualitative focus group interview study among Dutch biomedical researchers. BMJ Open. 2016;6(2).
14. Haven TL, Bouter LM, Smulders YM, Tijdink. Perceived publication pressure in Amsterdam – survey of all disciplinary fields and academic ranks. PLoS One. 2019;14(6):e0217931.
15. Haven TL, Tijdink JK, Martinson BC, Bouter LM. Perceptions of research integrity climate differ between academic ranks and disciplinary fields: Results from a survey among academic researchers in Amsterdam. PLoS One. 2019;14(1):e0210599.
16. Haven TL, Tijdink JK, Pasman HR, Widdershoven G, Riet G, Bouter LM. Researchers' perceptions of research misbehaviours : a mixed methods study among academic researchers in Amsterdam. Res Integr Peer Rev. 2019;4(25):1–12.
17. Martinson BC, Thrush CR, Crain AL. Development and validation of the Survey of Organizational Research Climate (SORC). Sci Eng Ethics. 2013;19(3):813–34.

18. Haven TL, Tijdink JK, De Goede MEE, Oort F. Personally perceived publication pressure - Revising the Publication Pressure Questionnaire (PPQ) by using work stress models. Res Integr Peer Rev. 2019;4(7):1–9.
19. Cronbach LJ, Nageswari R, Gleser GC. Theory of generalizability: A liberation of reliability theory. Br J Stat Psychol. 1963;16:137–63.
20. Snijders TAB, Bosker RJ. Multilevel analysis: An introduction to basic and advanced multilevel modeling. London: Sage; 1999. 266 p.
21. Kaatz A, Vogelman PN, Carnes M. Are men more likely than women to commit scientific misconduct? Maybe, maybe not. MBio. 2013;4(2):3–4.
22. Dalton D, Ortegren M. Gender Differences in Ethics Research: The Importance of Controlling for the Social Desirability Response Bias. J Bus Ethics. 2011;103(1):73–93.
23. Crain LA, Martinson BC, Thrush CR, Crain AL, Martinson BC, Thrush CR. Relationships Between the Survey of Organizational Research Climate (SORC) and Self-Reported Research Practices. Sci Eng Ethics. 2013;19(3):835–50.
24. Fanelli D, Costas R, Larivière V. Misconduct policies, academic culture and career stage, not gender or pressures to publish, affect scientific integrity. PLoS One. 2015;10(6):1–18.
25. Treviño LK, den Nieuwenboer NA, Kish-Gephart JJ. (Un)Ethical Behavior in Organizations. Annu Rev Psychol. 2014;65(1):635–60.
26. Martin KD, Cullen JB. Continuities and extensions of ethical climate theory: A meta-analytic review. J Bus Ethics. 2006;69(2):175–94.
27. Simha A, Cullen JB. Ethical climates and their effects on organizational outcomes: Implications from the past and prophecies for the future. Acad Manag Perspect. 2012;26(4):20–34.
28. Gorsira M, Steg L, Denkers A, Huisman W. Corruption in Organizations: Ethical Climate and Individual Motives. Adm Sci. 2018;8(1):4.
29. Holtfreter K, Reisig MD, Pratt TC, Mays RD. The perceived causes of research misconduct among faculty members in the natural, social, and applied sciences. Stud High Educ. 2019;1–13.
30. Groves R. Nonresponse rates and nonresponse bias in household surveys: What do we know about the linkage between nonresponse rates and nonresponse bias? Public Opin Q. 2006;70(5):646–75.
31. Fanelli D. How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. PLoS One. 2009;4(5):e5738.
32. Podasakoff PM, MacKenzie SB, Lee J-Y, Podsakoff NP. Common method biases in behavioral research: A critical review of the literature and recommended remedies. J Appl Psychol. 2003;88(5):879–903.

Additional information is available via https://osf.io/q45ev/

8

Chapter 8

Researchers' perceptions of a responsible research climate – a multi focus group study

This chapter has been published as: Haven, T., Pasman, H.R., Widdershoven, G. *et al*. Researchers' Perceptions of a Responsible Research Climate: A Multi Focus Group Study. Sci Eng Ethics (2020).
https://doi.org/10.1007/s11948-020-00256-8

Abstract

Introduction: The research climate plays a key role in fostering integrity in research. However, little is known about what constitutes a responsible research climate. We investigated academic researchers' perceptions on this through focus group interviews.
Methods: We recruited researchers from the Vrije Universiteit Amsterdam and the Amsterdam University Medical Centre to participate in focus group discussions that consisted of researchers from similar academic ranks and disciplinary fields. We asked participants to reflect on the characteristics of a responsible research climate, the barriers they perceived and which interventions they thought fruitful to improve the research climate. Discussions were recorded and transcribed at verbatim. We used inductive content analysis to analyse the focus group transcripts.
Results: We conducted 12 focus groups with 61 researchers in total. We identified fair evaluation, openness, sufficient time, integrity, trust and freedom to be mentioned as important characteristics of a responsible research climate. Main perceived barriers were lack of support, unfair evaluation policies, normalization of overwork and insufficient supervision of early career researchers. Possible interventions suggested by the participants centered around improving support, discussing expectations and improving the quality of supervision.
Discussion: Some of the elements of a responsible research climate identified by participants are reflected in national and international codes of conduct, such as trust and openness. Although it may seem hard to change the research climate, we believe that the realisation that the research climate is suboptimal should provide the impetus for change informed by researchers' experiences and opinions.

Key words: research climate, responsible conduct of research, research integrity

Introduction

Breaches of research integrity have inspired studies into what drives researchers to engage in questionable research practices or research misconduct (1). Whereas initial explanations focused on the individual level, it has become increasingly apparent that the organizational research climate plays a key role in fostering integrity in research (2–5).

In this study, we define the organizational research climate as: "the shared meaning organisational members attach to the events, policies, practices and procedures they experience and the behaviours they see rewarded, supported, and expected." (6,7). Organisational culture, in contrast, can be defined as "the shared basic assumptions, values, and beliefs that characterise a setting…" (6) (p. 362). In this paper, we therefore focus on the shared meaning researchers attach to the policies, practices and behaviours they associate with a responsible research climate, reasoning that it is easier to intervene on behaviour or policies, compared to intervening on values and beliefs.

Interest in researchers' practices and behaviours in relation to research integrity can be traced back to Robert Merton's (8) scientific norms of disinterestedness, universalism, communality and organized skepticism and the emergence of the field of science and technology studies (STS). Later Zuckerman (9) built on Merton's norms and connected a failure to uphold them to various forms of scientific fraud. For example, failure to uphold communality would lead to plagiarism. However, Merton's norms have been criticised, with some questioning whether they are unique to science (10), whether we need specific norms to describe scientific good practice (11), and some contesting the idea of science as pursuing universal goals (10).

Part of the critique came from researchers arguing that studying researchers' behaviour should not be done by theorising using an outsiders' perspective. Instead, to understand researchers' behaviour, STS scholars argued that science had to be studied from within (12). These researchers, most notably Latour (13), immersed themselves inside the laboratory to study first-hand the research climate as a phenomenon "defined by local rules and local knowledge" (14) (p. 112). Recently, a Swiss research team investigated the local knowledge among Swiss biomedical scientists about research integrity, something they defined by referring to scientists' personal responsibility to be honest and objective (15). Our research question is inspired by this bottom-up approach, specifically those rules related to a responsible research climate.

There is already some evidence, mostly quantitative, that the research climate can foster or undermine research integrity (16–18). For example, in a research climate where competition and suspicion among peers prevail, researchers seem to be more inclined to misbehave (19–21). Conversely, in a research climate where new members were socialised into responsible research practices, researchers report less research misbehaviour (22,23).

Existing codes of conduct for research integrity are aspirational when it comes to

creating an environment conducive to research integrity, stating, for example, "Research institutions and organisations promote awareness and ensure a prevailing culture of research integrity." (ALLEA, 2017, p. 5) or "Institutions provide a working environment that promotes and safeguards good research practices." (Netherlands Code of Conduct for Research Integrity, 2018, p. 20). Yet it is not clear what this would look like in practice. With this study, we aim to explore researchers' perception of a responsible research climate through focus group interviews. Specifically, we looked into three questions: 1) What are key characteristics of a responsible research climate?; 2) What are the barriers to the creation of a responsible research climate?; and 3) Which interventions alleviate barriers and improve the research climate where necessary?

Methods

Ethical approval

All procedures performed in this study involving human participants were in accordance with the ethical standards of the institutional and/or national research committee (The Scientific and Ethical Review board of the Faculty of Behavioural and Movement Sciences (Vrije Universiteit Amsterdam), approval Number: VCWE-2017-017R1) and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards

Participants

We included researchers that worked at the Vrije Universiteit Amsterdam and the Amsterdam University Medical Centre, location VUmc. The inclusion criterium was that the participants had to work in research for at least one day per week. Our recruitment strategy was threefold. We approached heads of department to ask for interested researchers, used our collegial network and randomly invited researchers by email to invite them to participate in our study. We aimed to recruit researchers for 4 discipline-specific focus groups (i.e. biomedicine, natural sciences, social sciences and the humanities) that were homogenous for 3 different academic ranks (PhD students, postdocs or assistant professors, and associate or full professors).

Procedure

After confirmation of participation, participants received the information letter (Online resource 1) that included a link to our privacy policy (Online resource 2) as well as the informed consent form via email (Online resource 3). We sent these again one week prior to the focus group. The focus groups were conducted between March and May 2018. To ensure that participants felt safe to speak freely, we conducted the focus groups with researchers from similar ranks only. We conducted the focus group

in English if there were participants that were not fluent in Dutch. We presumed that all researchers were proficient in verbalising themselves in English since English is considered the lingua franca of academia.

A moderator guided the focus group (JT or TH) discussion and an observer (internship student ES) made notes about the process and its content. The focus group started with a brief description of the project and its goals, as well as the goal of the focus group. Possible questions could be asked prior to signing the informed consent form.

We started the focus groups with a general task in which we asked participants to reflect on the responsible research climate by writing down three characteristics of a responsible research climate, discuss them with their neighbour and then share their insights with the group. We then put forth the question which barriers participants perceived for a responsible research climate and enquired which interventions could help to overcome these barriers. More detailed information can be found in our topic guide in Online resource 4. The focus groups took 90 minutes on average.

Within 10 days of the focus groups, we sent participants a short summary of the discussion and asked them for corrections to increase reliability (member check) (24). Participants agreed or provided minor suggestions that we incorporated prior to analyses. Recordings were transcribed by a transcription company under a data processing agreement. We used ATLAS.ti 8.3.0 for Mac for the data analysis.

Analysis

We used inductive content analysis to analyse the focus group transcripts. Inductive content analysis helps to bring down complex discussions to meaningful themes of interest (25). Two team members (TH and JT) analyzed and coded the transcripts independently. Individual analyses were then contrasted and discussed with two other team members (RP and GW) until we achieved consensus (26).

Themes had to be relevant to our research questions. Specifically, we created separate coding schemes to visualize the characteristics of the responsible research climate (see Online resource 5). We did the same for the barriers for responsible research as well as interventions to improve the research climate where necessary (Online resource 6).

In our analyses, we focused on the research *climate* as defined in our introduction. We tried to code as openly as possible and looked for concrete behaviours, policies or practices related to each characteristic or theme. We acknowledged beforehand that the concrete behaviours that flow from a particular theme may be different for different disciplinary fields or academic ranks.

Results

Descriptive information

We conducted 12 focus groups with 61 researchers across four different disciplinary fields and three academic ranks, see table 1. In total, 36% was recruited through heads of departments, 7% through our own connections and the remaining 57% was randomly selected.

Table 1. Demographic information of focus groups participants.

Academic rank **Disciplinary field**	PhD student	Postdocs and assistant professor	Associate and full professor
Biomedical sciences	$5^{/5}$	$5^{/4}$	$4^{/0}$
Natural sciences	$4^{E/0}$	$3^{/0}$	$4^{/0}$
Social sciences	$4^{E/3}$	$7^{E/3}$	$4^{E/1}$
Humanities	$6^{E/5}$	$5^{E/5}$	$7^{/3}$

^{E} = focus group was conducted in English; other focus groups were conducted in Dutch. $^{/x}$ = number of female participants.

Responsible research climate

In the introduction of the focus groups, we asked participants to reflect on characteristics of a responsible research climate. Based on their discussions, we identified 6 characteristics that are presented in decreasing order of frequency: fair evaluation, openness, sufficient time, integrity, trust and freedom, respectively.

In what follows, we first describe the characteristic in general and how participants thought the characteristic could foster responsible research. We then elaborate on what sort of behaviors, policies or practices participants provided. Finally, we note whether the characteristic was described differently depending on participants' academic rank or disciplinary field. Illustrative quotes per characteristic can be found in Table 2.

Fair evaluation

Participants expressed that responsible conduct of research is partly fostered through fair evaluation of research and researchers' performance. A fair evaluation was thought to be conducive to a responsible research climate as it could encourage researchers to perform responsible research and put less emphasis on citation criteria. In particular, this meant that research institutions should have sound policies about talent development, selection and promotion. Participants noted that there should be a formal evaluation system that is tailored to a researcher's academic rank. More senior participants indicated that this meant assessing whether they mattered to the field and whether the papers they had written demonstrably had 'impact'. The PhD candidate participants indicated that fair evaluation meant that their PhD thesis should not require three papers published

in high-impact factor journals. Instead, evaluation of a PhD thesis should revolve around the quality of their research, independent of whether the results were positive or spectacular.

Relatedly, participants underlined that fair evaluation should include team-based performance evaluation and that teams need a diverse set of players. Responsible research evaluation then becomes more of a team endeavour. Participants in the biomedical sciences emphasised that studies are hardly ever done by one person. Participants from the humanities emphasised that team-based evaluation would allow them to appreciate that one colleague that is terrific at grant writing or that other colleague who is very skilled at writing international papers without perceiving them as competitors.

Openness

Participants used the term openness as an umbrella concept that involved open research practices, open communication between colleagues and openness to collaborations. Participants noted that openness about the conduct of research is a cornerstone of responsible research climate and involves conducting and reporting the research as openly as possible to transparently inform the reader about the results. Participants from the social and biomedical sciences mentioned multiple examples of openness that revolved around transparent ways of conducting research where 'the research trail' is traceable and verifiable. In addition, participants noted that openness included sharing data, methodology or codes where possible. Early career researchers said that there should be more openness in reporting as well. Early career researchers often felt as if important details were left out to prevent others from replicating the findings in question.

In addition, participants described that openness meant being open to colleagues in the department or research group. This included communication about expectations but also openness about mistakes that were made, so that these could be handled appropriately.

Besides, participants emphasized that in a responsible research climate, researchers are more open to collaboration. This form of openness could allow researchers to work with others that may broaden their professional horizon. Participants perceived that collaborations were often only initiated when there was some form of personal gain. In terms of behavior, openness to collaboration meant that researchers would actively seek out collaboration, especially interdisciplinary collaboration, to solve more complex problems as a collective.

Sufficient time

A dominant perception among focus group participants was that responsible research requires sufficient research time. To participants' discontent, this research time was often overshadowed by other tasks (i.e. clinical, teaching or administrative duties). When time is short, research conduct could become 'sloppy'. Participants acknowledged

that good quality research takes time: time to keep up with developments in the field, to think, to read a paper, to produce a thorough manuscript or grant application review or to supervise students.

In addition, early career researchers emphasized that time is needed to make mistakes, learn and improve so that they can become responsible researchers. The participating senior researchers noted that they perceived difficulties prioritising their research, whereas in a responsible research climate, research is no hobby that one does on the side but one's main focus.

Integrity

Third, the topic of integrity pervaded many focus group discussions. In terms of behavior, integrity meant that researchers reflect critically on their own work. Furthermore, participants considered integrity to be conducive to a responsible research climate as it would involve department leaders encouraging their staff to develop their moral, and not just their scientific, competences. Lastly, to our participants, this includes having the right type of intrinsic motivation: the desire to do good research and to pursue the truth.

In addition, senior participants noted that there should be role models that conduct their work and supervision with integrity. This means there should be good examples, from starting with a good research question to being fair in scientific attributions and including all researchers that deserve authorship on the author list.

Trust

According to the majority of the participants, trust is crucial in contemporary research with (inter)national and interdisciplinary collaborations. However, participants noted that trust is not sitting back and blindly relying on one's peers. Instead, trust has to be sustained by actively holding each other accountable. One way to do this is through actively checking-in with peers and collaborators on how their work is going. This way, researchers can collectively hold expectations of good research practices in high regard. Senior researchers emphasized that in a responsible research climate, they should be able to trust those working below them to do their work with utmost care. Likewise, junior researchers stressed that they needed to trust that their supervisors know where the research projects are going.

Freedom

Our participants indicated that freedom is vital for a responsible research climate. Regarding behavior, participants noted that there should be freedom to disagree with the existing scientific paradigm and to engage in scholarly debate. Freedom also meant that PhD students were encouraged to not passively accept their supervisor's view, instead they should be encouraged to, when appropriate, challenge their supervisor's

view. When the research climate does not allow for this, false views may perpetuate for longer than needed. In addition, researchers from the humanities and natural sciences referred to the academic freedom to have autonomy in deciding which topics to study.

Table 2. Illustrative quotes regarding the characteristics of a responsible research climate.

Characteristic	Quotation
Fair evaluation	"I think an evaluation in which you can excel in one of the topics and don't have to excel in all of them, so either you are required to have average scores on all topics that would be okay, or you can excel on a few of these and then perhaps not excel so much in others. I'm really fed up with all the boxes that have to be ticked and the list is getting longer and longer, and there's no priorities there." – Assistant professor, social sciences "I think more evaluations on a team base, so everybody has his own [strength], you have the one [colleague] that's on the media, you have a colleague who's writing grants and you have a colleague who's writing more international journal papers or something like that." – Postdoctoral researcher, social sciences
Openness	"…data sharing… allowing your data to be re-analyzed by other individuals to confirm the results that you have reported in your publication, and also allowing other researchers to incorporate those data into their own meta-analyses or allowing it to inform their research questions." – full professor, social sciences "Openness is also that you feel open to discuss with others, if you feel that they are, maybe not mindfully, but they are doing things in a slightly different or wrong way in your opinion. That you can discuss this with the other person, without him or her feeling attacked by this. So that there is really an atmosphere of okay, we just trying all the best that we can and if somebody is doing something slightly wrong, it's not a problem. We just work it out and we go on and we continue to do it better." – Associate professor, biomedical sciences
Sufficient time	"The essence for conducting sound research is to have time and this time increasingly shrinking due to many disruptions and the loss of support staff" – Full professor, natural sciences "What is a research climate that I can work in responsibly without getting completely stressed out or anything? That means having time to think and write, because often the teaching time uses up all the research time so being able to protect that time." – Assistant professor, humanities "And the second is time for research, we now have research time of less than a day a week and in that time, I can hardly read papers since in that time, I also need to supervise my students. When am I supposed to do my own research? I find that very unsatisfying. I enjoy teaching students, but I would like to define myself as a researcher, not primarily an administrator or a teacher" – Associate professor, humanities
Integrity	"Integrity is more something for yourself. So, you need to approach research with integrity and be the first to doubt your own research results" – Full professor, natural sciences "My point was mainly that there is an atmosphere where the professor gives the good example. But a good example is also truly listening, to people, to the data, to convey the attitude that the nothing but the truth matters" – Associate professor, biomedical research
Trust	"To trust both the one working beneath you as well those above you and to assume that they conduct good research and that they claim something for a reason" – PhD student, biomedical sciences "Scientific progress may look immense but it arises because we are with so many and all those baby steps eventually lead to giant leaps forward. And here reputation is of utmost importance, because you are in this international network in which you don't always see what others have done in a different lab with their students, so if I try to repeat what a colleague from abroad has done and it does not work, then that trust is gone." – Assistant professor, natural sciences
Freedom	"You need to have freedom in choice of topic… Ultimately you are best at judging what has potential in your area of expertise, what will lead to success or changes…" – Full professor, natural sciences

Barriers

Participants described four main barriers to the creation of a responsible research climate: lack of support, unfair evaluation policies, normalization of overwork, and insufficient supervision. We describe them in order of frequency. Illustrative quotes per theme can be found in Table 3.

First, participants perceived a lack of support from their research institution, which included bureaucracy. Participants reported an excessive administrative burden and talked about administrative systems that worked inefficiently. Some professors connected this to the university policy that had centralised all forms of support. In effect, this resulted in support staff that were generalists, who were often unable to help as promptly as the former department secretary could. Instead of simply delegating a task, professors sometimes spent half their day in requesting lab supplies whereas they felt that their time would be better spent on research.

Second, participants expressed concerns about unfair evaluation policies where the main focus was on publication quantity instead of on the quality of their scientific work. Related to this were concerns about the emphasis on impact factors, as one natural sciences professor put it: "It is not that I don't want my paper to appear in Nature, but the reason that I would want it, is mainly due to the perception of policy makers or the university board, whereas there are other journals that are actually better, or at least, I would personally be happier for my paper to appear in one of those. You are pushed by people with little understanding of what you are doing, they now determine that you have to get published in Nature." This discontent was mirrored in a discussion among biomedical sciences professors, who referred to this as the "impact fetish".

In addition, participants in the humanities perceived that they had to excel in everything with little facilitation from their research institute in supporting them to develop professionally. Assistant professors said that they were on teaching positions without sufficient hours which meant that, in effect, teaching "ate up" their research time.

Third, participants talked about how their research climate normalized overwork and how everything had to go fast. Many described how it seemed to be expected of them to do the bulk of their research in their own time and how their working week often consisted of 60 hours or more. Besides, participants felt stressed when they were not available as they noted their colleagues expected them to respond during the weekends and in the evenings. In other words: participants reported that it seemed the norm to be available all the time. Participants noted that this perpetuation of overwork was not conducive for a responsible research climate, as overworked researchers risk conducting their work less thoroughly than is desirable. Besides, some participants shared that they had experienced burnout symptoms, but that they were too afraid to lose their job and hence did not raise the issue.

Finally, participants noted that supervision of junior researchers often seemed

suboptimal. PhD students described situations where colleagues were given three years of time to finish their thesis, but that their supervisor expected the quality of a four-year thesis. Besides, PhD students were aware that the odds of a future academic career were low, but reported that their supervisors devoted no attention to a possible career outside academia. Senior participants expressed concerns about the lack of role models who convey responsible research, whereas more junior participants underlined how to do research is mostly conveyed by the supervisor. Both PhD students and senior researchers noted that there was little guidance on how to become a good supervisor or a role model that fosters a responsible research climate.

Table 3. Illustrative quotes about barriers to achieving a responsible research climate.

Theme	Quotation
Lack of support	"The teaching is pretty badly organized in general, so you have a lot of administrative burden aside from what you do with the students, which should be taken away", Assistant professor, social sciences "Inconsequent technical support. Within our faculty, there has been a lot of changes in the technical support. For example, in our group, we had a technician that was the expert on all the lasers that we are using. And he was in the group for fifteen years and he knew everything about it. So, if any PhD was coming in new, they were trained by him. Then later on, they could do their own experiments. But at least this guy was always there and he was the person to train new people. And later on… there were less grants, so less people in the group eh, this person moved to another university." – PhD student, natural sciences
Unfair evaluation	"A lot of it is dependent on the evaluation if you're a good researcher, a particular evaluation of output across the different departments, trying to sort of find a scale that compares them all, which now has become the impact factor, is the holy grail. Impact factor based, so we, we align everybody along the same scale, the, regardless of sort of the history, the discipline, the, the nature, I guess the ease of publishing in the availability of journals and journal space etcetera eh, how, how you wanna see that. So that's, that's where you see that one thing has made it very difficult at the department level, is to manage your own policy eh, make your own decisions and, and make judgements on what you think is good and bad research and how you want to incentivize research…" – Full professor, social sciences
Normalization of overwork	"I once said to my supervisor that I had pain in my arms and neck from working. She said, yeah, I've had that for twenty years already so…" – Postdoctoral researcher, social sciences "the norm is that you are available full time, that is the norm. So, if you deviate from that [norm]… you don't feel well or you think maybe I am going to get into trouble because I am not, now I am not responding" – Assistant professor, humanities

Insufficient supervision	"If you're only trained by your promoter and your co-promoter, you get one world view about how you should publish things. And that's not really the only world view that's out there. And it's then just up to your supervisor how flexible he is and his world view where he only accepts different approaches to research. So, I think in that way it can be really limiting to not have any experience with other research groups" – PhD student, natural sciences "There's this huge amount of PhD students because they are cheap, and then the story is that you educate them and outside it would be helpful outside of academia, but from many PhDs perception I have the feeling that that's not true." – Postdoc, social sciences "I mean, now you can do a course to learn about teaching. But in the past, nobody learned how to teach. It was just something expected and based on tradition rather than the idea that you, well you might actually learn how to teach. So, I think this is the same with supervision…" PhD student, humanities

Interventions

Finally, participants discussed several interventions to overcome the beforementioned barriers to a responsible research climate. We identified three themes related to interventions that participants thought to be conducive to a responsible research climate: improve researchers' support, discuss expectations, and improve quality of supervision. Illustrative quotes per theme can be found in Table 4.

Improve support of researchers

First, participants mentioned that researchers could be supported in different ways: by decreasing the administrative burden, by sound research evaluation policies and by creating formal research time. For example, participants emphasized the need to diminish administrative hassles by investing in support staff. They noted that support staff hours should be included in grant applications.

Additionally, participants discussed the creation of evidence-based research evaluation policies that acknowledge scientific excellence in different ways. Concretely, participants noted that funding volume and number of publications should not be the sole criteria for promotion. Instead, professors and group leaders should be consulted regarding whom they thought suitable and why. Additionally, other criteria such as public outreach, outstanding teaching qualities and cross-disciplinary collaborations should also be considered. Lastly, participants explored the idea of team-based evaluations so that individual researchers need not excel at everything on their own.

Besides, participants discussed creating formal research time. This included constructing periods in the academic year where researchers had no teaching duties and their administrative duties were kept at a minimum. For clinicians, this meant their supervisors needed to support them in protecting their research time whilst specializing. Participants stressed that this required a change in attitude among researchers: instead of treating research "as a hobby" that you do on the side, researchers needed to take pride

in their research time.

Discuss expectations

Also, participants noted that it should be more accepted to set limits about what to do. Participants contended that it is healthy to do something else than work. They noted that in order to change the current existing climate, team members should be transparent and openly discuss expectations regarding (un)availability. Participants stressed that this discussion needed to involve PhD students. Relatedly, PhD supervisors had to be aware about the (unintended) expectations they might convey when sending PhD students emails in the evening or over the weekend.

Improve quality of supervision

Lastly, participants underlined that to improve the quality of supervision, it should be formalized what is expected of a PhD supervisor. Participants in the biomedical sciences coined the idea of creating a discipline-specific manual listing supervising duties. Other participants mentioned peer support groups for supervisors where they could discuss supervision-related dilemmas with their peers. In addition, participants emphasized that there should be training modules available for PhD supervisors that focus on cultivating responsible role model behaviors.

Table 4. Illustrative quotes about interventions to improve the research climate.

Theme	Quotation
Improve support	"In terms of say, the respect for professors to make decisions versus the authority to make [hiring] decisions… And it's a really negative signal of trust, if you are not being seen as the one who can actually best think about hiring decisions, about promotion decisions, about what task you want to whom." – Full professor, natural sciences
Discuss expectations	"I'm aware of colleagues who are very conscious about when they are sending emails. Opposite to what you're saying, they work on the weekends but they make sure not to send their PhDs replies on weekends or in the evenings because they don't want to get that message across." – Assistant professor, social sciences
Improve supervision	PhD student 1: "there are courses for principal investigators on how to supervise PhD students but they all don't have time… PhD student 2: or you should make it compulsory, that they have to repeat the course each year or something… PhD student 3: yes, and if you don't pass, you are not allowed to be a supervisor!" – PhD students, biomedical sciences "For example, I once took a course about supervising PhD students. Well, at that point the part on integrity was really small, it has increased somewhat in recent years. But there is a natural role for good supervision in your training as a researcher or clinical professional" – Associate professor, biomedical sciences

Discussion

This focus group study investigated academic researchers' perceptions of a responsible research climate, which barriers these researchers perceived in fostering a responsible research climate, and which interventions they considered beneficial for improving the research climate when necessary. In what follows, we reflect on our findings, connect them to existing literature, briefly consider the differences between academic ranks and disciplinary fields, and discuss the strengths and weaknesses of our study.

Connection with research integrity

It is important that the barriers we describe are not necessarily leading to an irresponsible research climate. They are interconnected factors that may hamper research integrity in various (and often indirect) ways, below we critically examine the barriers our participants discussed.

Take normalisation of overwork, there are different ways in which this could hamper the development of a responsible research climate, but it need not do so per se. To our participants, a systematic state of overwork could increase the chance that researchers engage in sloppy science (unintentionally, but in an overworked state, a researcher is less likely to notice errors, inconsistencies or flaws) or give in to temptation (when a researcher is overworked and frustrated, she may be more likely to incorrectly round off the obtained p-value).

Another example would be unfair evaluation, it is almost never the case that a system is unfair to everyone, rather it is unfair because it favours some over others. Hence researchers working on eye-catching topics may thrive in an evaluation system based on impact factors. But for many of our participants, it hampered a responsible research climate because the "impact fetish" steered researchers away from supervising or peer review, research-related activities that are also important (27).

Relatedly, readers may have experienced a lack of support somewhere in their academic career, but was this associated with an irresponsible research climate? To some of our participants, it was a form of research waste: valuable grant money is put towards a professor's salary with the idea that he or she uses that time to coordinate a study, not with the idea that the professor's days are spent on ensuring lab supplies. For junior researchers who were not charged with running a lab, the lack of support meant that they struggled to learn the right skills required to do their research, as the lab support staff who were extremely well-versed in complicated lab techniques got laid-off. For those outside a lab, the lack of support meant that it was hard to find sufficient time to sit down and write a sound paper. According to them, academic writing is not something one can do in between various administrative tasks, it requires time to engage with a topic, find the right words to succinctly convey the findings and integrate different perspectives. The idea here would be that if senior staff are forced to spend rather large

stretches of their day on administrative tasks, this may negatively impact the quality, rigor and thereby integrity of their work.

Finally, many PhD students may, in hindsight or currently, describe their supervision as suboptimal. Can this not just be interpreted as PhD students complaining about their superiors, as employees in other work environments will do from time to time? It is our understanding that participants recognised the challenges of good mentorship and that they saw a clear role for the supervisor in conveying responsible conduct of research through role modelling and responsible supervision (28). This is not to say that without a responsible role model, PhD students would go astray. Yet, with an irresponsible role model, it may be more likely that PhD students internalise flawed research practices.

Responsible Research Climates

The term, responsible research climate may remind readers of a field called Responsible Research and Innovation (RRI) (29) that has gained attention thanks to the European Commission's emphasis on the topic in its Horizon 2020 funding calls. Where RRI focuses on public engagement and societal relevance of research, RCR concerns behaviours that influence the validity of, and the trust in, research (World Conferences on Research-Integrity, 2020). Nevertheless, some studies that looked into factors that hindered or facilitated RRI can be illuminating for RCR as these studies looked at the research climate as well.

A case study of what RRI looks like in practice that was conducted at two Dutch research universities (Wageningen University and Radboud University Nijmegen) around the same time as our study allows us to compare and contrast our results (31). One of the main barriers to a fruitful uptake of RRI that their interviewees reported was the mismatch between researchers' wishes (e.g. to conduct research that is relevant for society) and the way in which they were formally evaluated and rewarded that emphasised publications and grants (31). This resembles what our participants described as unfair evaluation, although our participants did only occasionally mention societal relevance.

The RRI case study listed the "autonomy-oriented academic culture" (p. 58) at Radboud as a barrier for implementing compulsory research integrity training for senior researchers. This could be interpreted as a clash between two of our characteristics of a responsible research climate: freedom on the one hand (especially among "anti-hierarchical" senior researchers (p. 58)) versus integrity on the other hand. What this example illustrates is that creating a responsible research climate requires great care as predominantly focusing on one characteristic may come at the expense of another.

Comparison with existing literature

The characteristics that our participants associated with a responsible research climate may not surprise the reader. Here we compare and contrast our findings with existing literature, in an attempt to show that at times, new or different perspectives that

can be revealed.

Openness among colleagues and openness to collaboration have previously been mentioned as conducive to responsible research. Based on a survey and focus groups among U.K. academics, Joynson and Leyser reported openness and collaboration to be pivotal for high quality research (Joynson & Leyser, 2015).

Related, Munafò and colleagues reported a variety of ways to make research more open and thereby more responsible (32). In line with the goals of open science, they encouraged open data, open software and open materials, which was something our participants discussed too. However, there are also differences: Munafò and colleagues (2017) also elaborate on the benefits of initiatives such as registered reports, preprints or preregistration, whereas these initiatives were not mentioned in our focus group discussions. This indicates that certain open science related initiatives may be more accepted by the scientific community overall than others, perhaps because open data or open software are more broadly considered to be relevant compared to, for example, preregistration.

Finally, scholars that study the organisational climate have consistently emphasised the importance of fairness of internal processes, such as promotion and evaluation (6,33). It is thought that working in a climate where one is treated fairly, one is more likely to abide by organisational rules and procedures. Applied to the research climate: when researchers feel fairly evaluated, they may be less likely to cut corners (34).

Finally, improving supervision has previously been discussed as something that institutions should provide clearer guidelines for (35). Our participants extended this with concrete examples of how to formalize what is expected of a supervisor by means of a supervision manual or through training programs that focus on responsible supervision. Guidance on good mentoring is not uncommon in The Netherlands, as indicated by a qualitative study amongst PhD students and PhD supervisors by Maastricht University (36,37). They recommend making expectations mutually explicit and emphasized clear and constructive feedback. Furthermore, Leiden University has a best practices manual for supervision (38) where they translate commitments into concrete actions that both the PhD supervisor and the PhD student can take. One possible avenue would be to incorporate good mentorship into the reward system, making it a scientific activity that is valued in its own right, as described in the recently released Hong Kong Principles for Assessing Researchers (27). Principle 5 reads: "Value a range of other contributions to responsible research and scholarly activity, such as peer review for grants and publications, mentoring, outreach, and knowledge exchange." (p. 11). These contributions could be incorporated into the talent development policies that our participants discussed, but their possible effects on the research climate should be examined with due scrutiny.

Future studies examining these intervention effects (here: incorporating supervision into the evaluation criteria) should be explicit about where it is they expect change to occur. For example, would they expect a change in the perceptions of the research climate

among PhD students or a decrease in dropout among PhD students? It is beyond the scope of this study to define a precise way to measure this or any other effect. We also note that evaluating the effects of interventions to promote RCR can be challenging, a Cochrane review of 31 educational intervention studies to promote RCR found hardly any effect (39).

Differences between disciplinary fields and academic ranks

In our analyses, we focused on broad characteristics and we did occasionally find differences in the way in which a particular characteristic was operationalised. When reviewing the possible differences between academic ranks, the idea of teaching-free periods pervaded discussions among postdocs and assistant professors and was less pronounced among focus groups with PhD students, possibly because not all PhD students have formal teaching duties. Similarly, some participating biomedical researchers noted that it was not teaching that 'ate up' their research time, it was that their clinical duties were prioritised regardless. Yet, these biomedical researchers too contended that good research practices benefit from sufficient time. All in all, we take these differences to be a matter of degree and believe that the characteristics we identified could bear relevance for researchers across academic ranks and disciplinary fields.

Strengths

This is the first study to systematically investigate perceptions of a responsible research climate across academic ranks and disciplinary fields (that is, across the biomedical sciences, natural sciences, social sciences and the humanities). Most of the published focus group studies about good research practices thus far focused on a particular disciplinary field, e.g. social or biomedical science (19,20,40). Because our sample included a diversity of academic ranks and disciplinary fields, we hope that our results are relevant for researchers regardless of their specific academic rank or disciplinary field.

We believe we were able to unpack what constitutes a responsible research climate by tangibly describing how participants characterised a responsible research climate. This, together with the interventions that were brought up by our participants, can provide a start for an evidence-based debate about fostering a responsible research climate.

Weaknesses

In light of our results, there are some limitations that need to be addressed. First, we aimed to recruit more participants and especially more female participants, particularly among the more senior ranks and in the natural sciences. This unequal composition reflects a national finding: Whereas Dutch universities have taken measures to increase the percentage of women in higher ranks of academia, the percentage of women gets lower with higher functions. We therefore acknowledge the possibility that female-specific barriers were overshadowed or that some barriers were more pronounced for

female researchers than we could find out due to our focus groups' gender composition.

Second, this study did not include participants from the technical and engineering sciences. That leaves the option open that we have missed key characteristics of responsible research for technical scientists and engineers. However, this is a sample-specific deficit since Amsterdam does not have technical or engineering science faculties.

Lastly, we focused mainly on behaviours, policies and practices which provide tangible results but may oversimplify certain issues. In addition, focusing on these behaviours does not allow one to infer the drivers of those same behaviours based on the same data. All in all, our data regard *which* characteristics researchers associate with a responsible research climate but not *why* they associate certain characteristics with a responsible research climate.

Conclusion

It may be hard to change the research climate. Alike the meteorological climate, an analogy that we borrow from the creators of the SOuRCe© (41), the research climate is influenced by different factors, such as individual researchers and the system governing academic science. But to stick to this analogy, it is the realisation that there is something wrong with the climate that can spark behavioural change. The interventions that we listed could give that behavioural change concrete shape. We hope that future research will explore their feasibility and effectiveness.

Acknowledgements

TH would like to acknowledge Elias Anttila and David Ludwig for their helpful literature suggestions. TH would also like to acknowledge the support of internship student Eveline Schipper during the collection of the data.

References

1. Levelt Committee, Noort Committee, Drenth Committee. Flawed science: The fraudulent research practices of social psychologist Diederik Stapel. 2012.
2. Bouter LM. Commentary: Perverse incentives or rotten apples? Account Res. 2015;22(3):148–61.
3. Sovacool BK. Exploring scientific misconduct: Isolated individuals, impure institutions, or an inevitable idiom of modern science? J Bioeth Inq. 2008;5(4):271–82.
4. Steneck NH. Institutional and individual responsibilities for integrity in research. Am J Bioeth. 2002;2(4):51–3.
5. Casadevall A, Fang FC. Reforming science: Methodological and cultural reforms. Infect Immun. 2012;80(3):891–6.
6. Schneider B, Ehrhart MG, Macey WH. Organizational climate and culture. Annu Rev Psychol. 2013;64(1):361–88.
7. Wells JA, Thrush CR, Martinson BC, May TA, Stickler M, Callahan EC, et al. Survey of organizational research climates in three research intensive, doctoral granting universities. J Empir Res Hum Res Ethics. 2014;9(5):72–88.
8. Merton RK, Storer NW. The sociology of science: Theoretical and empirical investigations. Chicago SE: University of Chicago Press; 1973. 605 p.
9. Zuckerman H. Deviant behavior and social control in science. In: Sagarin E, editor. Deviance and Social Control. Beverly Hills, CA: SAGE Publications; 1977. p. 89.
10. Sismondo S. An introduction to science and technology studies. 2nd ed. Chichester, West Sussex, U.K.: Wiley-Blackwell; 2010. 60 p.
11. Schmaus W. Fraud and the norms of science. Sci Technol Hum Values. 1983;8(4):12–22.
12. Knorr Cetina K. Laboratory studies: The cultural approach to the ttudy of science. In: Handbook of Science and Technology Studies. 1995. p. 140–66.
13. Latour B. Science in action: How to follow scientists and engineers through society. 7th print. Cambridge, Mass. SE: Harvard University Press; 1997. 274 p.
14. Jasanoff S, Markle GE, Peterson JC, Pinch T. Handbook of science and technology studies. Thousand Oaks, Calif.: Sage Publications; 1995. 820 p.
15. Shaw D, Satalkar P. Researchers' interpretations of research integrity: A qualitative study. Account Res. 2018;25(2):79–93.
16. Martinson BC, Anderson MS, Crain AL, De Vries R. Scientists' perceptions of organizational justice and self-reported misbehaviors. J Empir Res Hum Res Ethics. 2006;1(1):51–66.
17. Edwards MA, Roy S. Academic research in the 21st century: Maintaining scientific integrity in a climate of perverse Incentives and hypercompetition. Environ Eng Sci. 2017;34(1):51–61.
18. Karen SL, Anderson MS, Rosenberg L. Academic misconduct and values: The department's influence. Rev High Educ. 1995;18(4):393–422.
19. De Vries R, Anderson MS, Martinson BC. Normal misbehavior: scientists talk about the ethics of research. J Empir Res Hum Res Ethics. 2006;1(1):43–50.

20. Anderson MS, Ronning EA, De Vries R, Martinson BC. The perverse effects of competition on scientists' work and relationships. Sci Eng Ethics. 2007;13(4):437–61.
21. Joynson C, Leyser O. The culture of scientific research. F1000Research. 2015;4:1–11.
22. Crain LA, Martinson BC, Thrush CR. Relationships between the Survey of Organizational Research Climate (SORC) and self-reported research practices. Sci Eng Ethics. 2013;19(3):835–50.
23. Anderson MS, Louis KS, Earle J. Disciplinary and departmental effects on observations of faculty and graduate student misconduct. J Higher Educ. 1994;65(3):331.
24. Meadows LM, Morse JM. Constructing evidence within the qualitative project. J Qual Res. 2005;106–12.
25. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs. 2008;62(1):107–15.
26. Lincoln YS, Guba EG. Naturalistic inquiry. Beverly Hills, Calif. SE.: Sage Publications; 1985.
27. Moher D, Bouter L, Kleinert S, Glasziou P, Sham MH, Barbour V, et al. The Hong Kong principles for assessing researchers: Fostering research integrity. PLoS Biol. 2020;18(7):e3000737.
28. Bird SJ. Mentors, advisors and supervisors: their role in teaching responsible research conduct. Sci Eng Ethics. 2001;7(4):455–68.
29. Burget M, Bardone E, Pedaste M. Definitions and conceptual dimensions of Responsible Research and Innovation: A literature review. Sci Eng Ethics. 2017;23(1):1–19.
30. World-Conferences-on-Research-Integrity. Mission of the WCRIF [Internet]. 2020 [cited 2020 May 11]. Available from: https://wcrif.org/foundation/mission
31. Van der Molen F, Consoli L, Ludwig D, Magnaghten P. Responsible Research and Innovation in Practice: Report from national case study, The Netherlands. 2018.
32. Munafò MR, Nosek BA, Bishop DVM, Button KS, Chambers CD, Percie Du Sert N, et al. A manifesto for reproducible science. Nat Hum Behav. 2017;1(1):1–9.
33. Gorsira M, Steg L, Denkers A, Huisman W. Corruption in Organizations: Ethical Climate and Individual Motives. Adm Sci. 2018;8(1):4.
34. Martinson BC, Crain LA, De Vries R, Anderson MS. The importance of organizational justice in ensuring research integrity. J Empir Res Hum Res Ethics. 2010;5(3):67–83.
35. Kornfeld DS. Perspective: Research misconduct: The search for a remedy. Acad Med. 2012;87(7):877–82.
36. Woolderink M, Van Der Boom H, Putnik K, Klabbers G. The voice of PhD candidates and PhD supervisors. A qualitative exploratory study amongst PhD candidates and supervisors to evaluate the relational aspects of PhD supervision in the Netherlands. Int J Dr Stud. 2015;10:217–35.
37. Van der Boom H, Klabbers G, Putnik K, Woolderink M. It takes two to tango. 2013.
38. Leiden/University. Best Practices for PhD Supervision [Internet]. [cited 2020 May 7]. Available from: https://www.universiteitleiden.nl/binaries/content/assets/geesteswetenschappen/pdfs/best-practices-for-phd-supervision.pdf
39. Marusic A, Wager E, Utrobicic A, Sambunjak D, Anderson MS, Rothstein HR. Interventions to prevent misconduct and promote integrity in research and publication. Cochrane Database Syst Rev. 2013;2013(2).

40. Tijdink JK, Schipper K, Bouter LM, Pont PM, De Jonge J, Smulders YM. How do scientists perceive the current publication culture? A qualitative focus group interview study among Dutch biomedical researchers. BMJ Open. 2016;6(2).
41. Martinson BC, Nelson D, Hagel-Campbell E, Mohr D, Charns MP, Bangerter A, et al. Initial results from the Survey of Organizational Research Climates (SOuRCe) in the U.S. department of veterans affairs healthcare system. PLoS One. 2016;11(3):1–18.

Online resources available on the journal's website:
https://link.springer.com/article/10.1007/s11948-020-00256-8


9

Chapter 9

Discussion

It has been nearly four years since I started this PhD project and only knew about the Stapel case remotely. In this final chapter, I want to summarize the main findings, connect them to existing trends and other factors related to research misbehavior and integrity, touch upon some methodological limitations of my work, draw lessons for future research, revisit the different disciplinary fields in the context of major research misbehaviors, and conclude with a brief list of recommendations for fostering a responsible research climate.

Main findings

In **chapter 2**, we learned that there is more detailed information needed about cases like Stapel's to meaningfully apply the theories intended to shed light on these instances of research misconduct. As for the research climate for integrity, we learned that perceptions thereof differ between disciplinary fields, with researchers working in the natural sciences having a more positive perception overall. Senior researchers also have a more positive perception of the research climate than assistant professors, postdocs or PhD students (**chapter 3**). Assistant professors and postdocs perceived more publication pressure (**chapter 5**), but PhD students perceived the greatest lack of resources, as measured with the revised Publication Pressure Questionnaire (**chapter 4**). When we surveyed all academic researchers and specifically asked focus group participants which of the research misbehaviors they actually perceived and were most concerned about, we learned that insufficient supervision and various forms of sloppy science are researchers' greatest concerns (**chapter 6**). This sparked questions about the extent to which the research climate and publication pressure explain variance in research misbehavior, where we found that both factors together with someone's academic rank jointly explain a substantial proportion (32%). Most notably, the research climate explained 23% of variance in perceived research misbehavior (**chapter 7**). Finally, we asked researchers to describe the characteristics of a responsible research climate and learned that, according to our participants, a responsible research climate arises where fair evaluation, openness, integrity, trust and freedom thrive and where researchers are given sufficient time to do their work (**chapter 8**).

Existing trends

I noted in the introduction that there were two ideas about research integrity that paved the way for this PhD project. First, there was reason to be more concerned about QRPs as their prevalence is estimated to be much higher than FFP's prevalence (1–3). To some extent, the same picture appears from the current study. To be clear:

Our participants all recognized and emphasized the seriousness of FFP, but the items about the FFP were ranked low on perceived frequency and when asked to reflect on misbehaviors they were concerned about to hamper the research climate, FFP was not the main focus. Compare this with the long lists of misbehaviors that researchers added, some even new to us, such as changing a manuscript due to pressure from editors or reviewers in a grossly incorrect way in order to get it published or the use of research grants for unintended purposes.

The second recognition was that it is meaningful to extend our focus to the 'barrel', as opposed to solely focusing on 'bad apples' (4–7). This view is echoed in this dissertation, with perceptions of the research climate playing a major role in explaining variance in perceived research misbehavior.

Relationships with other factors

The research climate and publication pressure, however, are not the only factors that may contribute to promoting research integrity or preventing research misbehavior. Let's revisit our fruit analogy. We have the apples, the barrels and the fruit market, or as George (2016) wrote: "three broad general narratives about three primary contributing factors — individual traits, institutional issues, and structural problems in science itself" (8) (p. 19). The measurement instruments we used are at best proxies for these complex sets of factors at the institutional and structural level. Below I briefly review some evidence of other, sometimes related, factors that have been linked to research misbehavior. I also note trajectories worth exploring further in relation to our findings.

Appels

First, this study did not focus extensively on the individual (the 'apple'). Hence we did not investigate trait-theory, the idea that an individual's traits make the person more prone to research misbehavior. There is some evidence that narcissism or Machiavellistic personality traits might be associated with research misbehavior (9,10). Here it stands to reason that in a department with researchers scoring high on these traits, researchers will be more likely to perceive that as a climate where competition and suspicion among colleagues are high. However, it is not clear if poor-quality climates attract specific personalities, specific personalities determine the quality of the climate, or whether the two are associated in some other way.

Related, we did not empirically study rational choice theory, where it is presumed that researchers make a rational decision to misbehave, carefully outweighing the costs and benefits (11–13). Some argue that the association between the impact factor of the journal (high benefits) and amount of retractions (as an indication of misbehavior (14,15)) can be taken as evidence that rational choice theory is a meaningful explanation

of research misbehavior (11,16,17).

A more positive approach that lies somewhere in between individuals and their climates is virtue-ethics (18). In contrast to conveying RCR as a set of rules that scientists have to obey, the notion of virtue-ethics developed by Pennock (2015) has it that RCR should be seen as something that naturally follows from academic research' aims. Virtue-ethicists argue that we should direct our attention to how we can strengthen these virtues for researchers, traits that make researchers *better* researchers, e.g. the trait of curiosity or intellectual honesty (for other science-related virtues, see (19)). They believe training or cultivating these virtues contributes to creating a culture of research integrity. It seems intuitive to me that cultivating these virtues could contribute to the quality research climate, as virtuous researchers presumably act in ways that are, for example, more intellectually honest (19). If the climate is then mostly made up of honest researchers, the perceptions of researchers working in this research climate are presumably rather fair and positive, but I am not aware of any empirical studies putting this to the test.

Barrels

Other factors would best be classified as part of the barrel, such as mentoring (20–22). We touched upon mentoring and supervision as these as they were part of the items in our survey questionnaire and appeared in our focus group discussions, but there is more to be said about their influence on research misbehavior. Especially since there is research from other areas showing that people may cheat to benefit others or their organization (23,24).

For mentoring, it seems natural to assume that better quality mentoring will be associated with responsible research, whilst a poor-quality mentoring relationship would be risk factor for questionable research practices (as when mentoring is mostly focused on survival in science, it may up the odds of engaging in research misbehavior (21)).

It would be interesting to measure the perceptions of this relationship among both supervisors and supervisees, as a relationship's quality is determined by both parties. Lombarts and colleagues developed the Systematic Evaluation of Teaching Qualities (SETQ) system for this purpose to measure the relationship between medical doctors and their trainees (25). Another starting point is the Mentoring the Responsible Conduct of Research Scale (MRCR), where one subscale focused on mentors modelling RCR in psychology (26). Adaptation of these scales to academic science (i.e. beyond a particular discipline) would be useful to measure whether interventions to improve mentoring produce their desired effect.

Outside the research integrity debate, the role of leadership in relation to ethical climates has been studied. Some of these studies looked into how leadership tries to empower employees, or what the influence of their communication style is (27), which may be worthwhile factors for exploring in relation to responsible research climates too.

Fruit markets

Pressure for funding (28) and job insecurity (29,30) could be classified structural problems in science ('fruit market'). Both are readily linked to publication pressure: to get a stable job, one needs sufficient funding, and to get sufficient funding, one is a better position with an impressive list of publications.

The realization that high publication pressure may be a problem to the integrity of academic research may prompt attempts to decrease publication pressure. Late 2019, The Association of Universities in the Netherland, Netherlands Federation of University Medical Centres, Royal Netherlands Academy of Arts and Sciences, Dutch Research Council (NWO), Netherlands Organization for Health Research and Development released a position paper (31) advocating a new way to evaluate and recognize academic researchers. This position paper states that, among other things, academic's assessment should be less focused quantity and more on quality. This was translated by NWO by introducing narrative resumes (32) where academics who apply for funding are only allowed to describe 10 'key outputs' (instead of a long list of published articles, and key outputs can be things other than scientific publications, too).

This is an interesting step that prompts questions, such as: Would the freedom that researchers now experience lead to more or less publications, what do we understand better-quality publications to be, and how will this affect the position of Dutch researchers in an international academic landscape? I do not claim to have the answer to any of these questions, only to note that there is more to be learned about fostering research integrity than has been covered in this dissertation.

Methodological limitations and lessons for future research

Let me start by saying that this study has some strong points, most notably its diverse sample of researchers across academic ranks and disciplinary fields. Besides, the use of validated measurement instruments is important, and particularly the revised publication pressure where our data can be used as a benchmark for others[1]. We looked at three categories of explanatory factors, individual, climate, and systematic, in conjunction. Lastly, I believe the combination of both quantitative and qualitative methods add value to the work discussed, where we did our utmost best to be thorough by preregistering our quantitative work and extensively discussing our qualitative work among the different researchers that contributed to it.

There are some methodological limitations of this dissertation that deserve attention. Where possible, I connect these to lessons for future research. I divide this section into quantitative and qualitative limitations.

1 The PPQr is currently used to evaluate publication pressure in Indonesia, see (65)

Quantitative

First, a recurring threat to the validity of our survey findings, is the possibility of a response bias (33). Less than 20% of our population of academic researchers in Amsterdam as employed in research in 2017 at the Amsterdam University Medical Centers, University of Amsterdam and Vrije Universiteit Amsterdam took part in our surveys. It is not the small number of participants (n = 1073) per se that leads to invalid results, as a response bias refers to a bias in the findings because respondents differed in some critical way from non-responders (34).

We tried to assess the extent of a response bias by comparing the demographics of our participants with those we could find about the population (35,36). In addition, we conducted a wave analysis, where we compared early responders to late responders (here the reasoning is that late responders, who needed multiple reminders, are probably similar to non-responders) (33). Both provided a mildly assuring picture: Respondents did not differ very much in term of academic rank, gender or disciplinary field compared with national statistics and the differences in mean scores between early and late respondents were small.

It is hard to tackle the problem of response bias, because many of us are flooded with surveys and the decision to participate in a survey on research integrity is presumably not random. Some recommend sending a non-response survey, an ultra-brief survey that enquires the reason for not responding or some demographic characteristics (36). In our case, less than 1% completed this non-response survey, arguably too low to draw conclusions from.

An often-cited reason for non-response is the length of the questionnaire (37). This inspired us to use *missingness-by-design* (38) in presenting participants with only 20 randomly selected misbehaviors from the list of research misbehaviors. Future research should take the general decreasing tendency (39–42) to participate in surveys into account and design smart and short questionnaires to decrease the chance of a response bias.

Second, both the SOURCE© (43) that we used to obtain a view of the research climate and the list of major and minor misbehaviors (44) were designed with mainly quantitative biomedical and social sciences research in mind. This raises questions about whether the questions asked were suitable for researchers in other disciplinary fields, especially those who use qualitative methods or don't do empirical research (like mathematicians and philosophers for instance).

The designers of the SOURCE© (43) took their biomedical orientation into account by adding a response option "not relevant to my research" (so when presented with an irrelevant item, respondents could indicate it did not apply to their work). If respondents indicated this was the case for more than half of the items in a scale, their mean score was not considered when calculating group means, reasoning that they have too little experience with the construct in question for their scores to be meaningful.

When presented with a misbehavior from the 60 items (44) irrelevant to their research, respondents could indicate that they never observed this. But that may still result in irritation among respondents and begs the question whether there were no *other* misbehaviors they were confronted with that we failed to ask about. This was the main reason why the focus groups looked into this.

For future research, it could be relevant to conduct focus groups first with researchers from the natural sciences and humanities prior to sending out a survey to assess the prevalence of research misbehavior (for an example of this approach, see (45)). Instead of only presenting focus groups participants 5 misbehaviors, they could then discuss the list of 60 in-depth and use participants input to build the survey logic in such a way that participants *first* indicated their disciplinary field (this was now done last) and then *only* saw the misbehaviors that were of relevance to them (both from the list and the ones their peers had added).

This brings us to the larger question of how to assess frequencies of research misbehavior in the most valid way. Asking participants to admit their own misbehavior or that of their peers both have their disadvantages (1). Research in fields that have investigated human misbehavior in other areas (i.e. corruption (46)) uses combined data, meaning they combine self-report, reports on others, known cases and other data (47) and generally utilize a more privacy-sensitive means of enquiring admittance rates, such as randomized response and related approaches (48,49).

Lastly, our data regard perceptions from academic researchers in Amsterdam, The Netherlands, yet it is unclear to what extent they generalize nationally, let alone internationally. When presenting this work at international conferences, the main themes seemed to resonate with the audience, but that is anecdotal evidence at best.

Qualitative

The major qualitative limitation is that we 'measured' the research climate indirectly. We discussed and analyzed participants' perceptions of the research climate but we did not *observe* their research climates ourselves. Future research should maybe consider an anthropology-inspired fly-on-the-wall approach (for an example of this approach in the field of research integrity, see (50)) where the researcher immerses him or herself in participants' day-to-day research climate.

As alluded to above, we presented focus group participants with 5 research misbehaviors that were most detrimental according to researchers in their disciplinary field on the aggregate level. These misbehaviors were put on the table as a stepping stone for discussion among participants regarding the research misbehaviors they considered most relevant — and that they had observed in their own research settings (to prevent them from contemplating about misdeeds in some faraway place). The research misbehavior that they added, in their own words, proved notoriously difficult to compare with our existing list of 60 items and resulted in many discussions within the research team of

whether an item was novel or actually the same thing phrased differently (e.g. is stealing an idea from a grant proposal sufficiently covered by the concept of plagiarism?). In line with the earlier remark, if it is the goal to find out whether some misbehavior is missing, future research should bring all misbehaviors to the table.

Finally, there was less concrete guidance on best open science practices for qualitative scholars. During the time I worked on this dissertation, I wrote a paper exploring the possibility of preregistering qualitative research together with Leonie van Grootel (51) and conducted a Delphi to attain consensus on what items to include in a qualitative preregistration form (52). Future research should look into the best way to preregister qualitative work that is sensitive to the inherent flexibility of qualitative research.

General suggestions future research

I would like to note two general suggestions for future research in the field of research integrity at large. First, as a field that studies responsible conduct of research, it is our duty to do this to the best of our ability. This means no longer surveying participants with questionnaires that are hastily constructed, and hence contain dubious or double-barrowed questions (53). A first step would be attaining consensus on how we unpack the constructs that are often measured in relation to research integrity. This in turn can ensure that enable valid contrasting and comparing, because there is general understanding of what these constructs mean (or even better: how to measure them). More broadly, if we want our results to be taken as evidence for policies, we need to be critical of that evidence, and agree, at least in some form, on how we evaluate that evidence (54).

Second, and related, our decision to study the research climate because it would allow us to deduce interventions reflects a general trend to design interventions to foster integrity in research. But these interventions should be evaluated with rigorous designs, a good example is the work of Plemmons and colleagues (55). We risk creating a never-ending list of pilot-interventions that seem promising with respect to a small, often self-selected population, but with little clue as to whether these interventions can be scaled up, whether they work better compared to no intervention, or whether they work on some larger population too[2]. Evaluating interventions to improve any human organization is notoriously difficult, but if we care about improving academic research, it may be worth the effort.

Disciplinary differences

We have repeatedly emphasized differences between disciplinary fields in perceptions of the research integrity climate, perceived publication pressure and aggregate impact of

2 I am not claiming promising pilot interventions are a bad thing, just that they are only a first step.

research misbehaviors. For example, 'Use published ideas or phrases of others without referencing' appeared in the top 5 of most detrimental research misbehaviors on the aggregate level among humanities researchers, but it was not so much an issue among researchers in biomedicine. This finding could be an informative conversation-starter in a debate among humanities researchers, but would presumably not lead to a relevant debate in a biomedical laboratory.

Another example was that researchers in the natural sciences, besides PhD students, researchers perceived less publication pressure. An intervention would then focus on strengthening the publication resources for PhD students and not focus on full professors (after all, you don't want to waste time on designing an intervention for a problem they don't perceive exists, or they do not seem bothered by).

Yet, there were no differences in the *overall* frequency of perceived research misbehaviors between disciplinary fields, but that was when taking all research misbehaviors, both major and minor, as a whole. What picture appears when we focus only on major misbehaviors that all researchers condone, regardless of their field, namely fabrication, falsification, and plagiarism. What percentage of researchers admitted having perceived this behavior in our sample? This concerns the following percentages:

	Biomedicine	**Natural sciences**	**Social sciences**	**Humanities**	**Total**
Falsification	15%	19.5%	14%	17%	15%
Fabrication	4%	17%	4%	9%	6%
Plagiarism	27%	38%	30%	43%	30%

Questions were asked in reference to the last three years.

Fanelli's (1) meta-analyses indicated 14.2% (95% CI: 9.9-19.7) of researchers knew about a colleague that falsified data. We observed a similar rate, ranging from 14 to 19.5%, depending on the field. For fabrication, Fanelli's meta-analysis showed 12.3% (95% CI 8.4-17.7) of researchers who knew of a case. Most of the studies reported in Fanelli's (1) meta-analyses used social sciences or biomedical researchers, so it would be fairest to contrast those columns with Fanelli's data. Interestingly, the proportion of researchers who knew of a case of fabrication is lower in our sample, with the natural sciences and humanities reporting in line with Fanelli's (1) meta-analytic findings (i.e. within the 95% CI) whereas only 4% of biomedical and social sciences reported having observed fabrication.

There is no strong agreement about the scope of major misconduct (56). Some have argued that changing results under the pressure of a sponsor should be included (57). Similarly, some argued that turning a blind eye to a breach of research integrity by others is itself research misconduct (20,56). I think it is safe to assume that researchers, regardless of their disciplinary field, will agree that these two behaviors are completely unacceptable. Let us review these percentages in our sample:

	Biomedicine	Natural sciences	Social sciences	Humanities	Total
Modify results or conclusions due to pressure of a sponsor	10%	9%	6%	8%	9%
Turn a blind eye to putative breaches of research integrity by others	25%	23%	29%	29%	26%
Questions asked in reference to the last three years.					

Again, we see only small differences between disciplinary fields, for example, 6% of social sciences' researchers observed this strong conflict of interest versus 10% among biomedical scientists. Only 23% of natural sciences researchers witnessed someone turning a blind eye, versus 29% of researchers in the social sciences and the humanities.

My tentative conclusion is: when it comes to *major* research misbehaviors, there are some bad apples in every field.

What can these findings teach us about a responsible research climate?

I noted above that the factors studied here don't tell the complete story, but do we need them at all for responsible conduct of research? If we go back to Steneck's conceptualization of RCR, as "conducting research in ways that fulfil the professional responsibilities of researchers, as defined by their professional organizations…" (58) (p. 55), then it seems rather simple. Just follow the code of conduct for research integrity (arguably this is one of the clearest examples where the "professional responsibilities as defined by researchers' professional organizations" can be found) and assure that you follow the code that applies to you (this depends, first and foremost, on your institution's geographical location — for instance all Dutch research institutions subscribe the Netherlands Code of Conduct for Research Integrity (59)).

But it is not that easy, for these codes may contain value pluralism about research integrity. Following the responsibilities, norms, standards, or principles (many of which ultimately relate back to particular values) does not give a straightforward answers or clear instructions on how to act. As described in a publication I co-authored (60), the norms in these codes may be in conflict, with for instance one emphasizing *full* openness and the other emphasizing *due* confidentiality.

At other times, norms can be incommensurable, they simply don't map on the same scale[3]. This last part, the incommensurability, matters because it makes it more difficult

3 Peels and colleagues (60) give an example from the Canadian research integrity policy (66) that contains one norm related to, say, *innovation*, namely "researchers should be innovative and that managed risk-taking should be encouraged" (p. 4) but also emphasises the value of *efficiency* by stating that there should be clearly defined deliverables (p. 4). There is no higher order value under which both innovation and

to systematically and transparently compare breaches of research integrity, which then complicates what sanction is fit given a particular breach. This risks an unsystematic way of 'punishing' researchers that breach integrity which is could be interpreted as unfair. And this unfairness, in turn, is exactly what we wish to prevent, as organizational justice theory has it that researchers are more likely to abide by, and even defend, a system that they believe is just, or fair.

We are back to the situation where some practical guidance for RCR may be welcome, with the side-note that conducting research with due integrity is never 'simply' obeying the rules. So what can the findings discussed in this thesis teach us about a responsible research climate? Summing up, the studies described in this dissertation suggest that to create a responsible research climate, the research process should be considered *in its totality.* It takes effort from stakeholders on different levels, as a supervisor may want to improve his or her supervision skills, but if the group leader does not see the added value of better supervision, this may be hard. Based on our findings, I make some concrete recommendations for different stakeholders, that each have to play their part in fostering a responsible research climate. Per stakeholder, I call for one suggested action to take and then I elaborate on how that act could be shaped. To challenge the stakeholders to act, I write in the active tense, and I trust readers will not be upset by my mildly commanding tone of voice.

Institutions

Expand and develop the current assessment criteria at your institution. Think about ways of rewarding publications that adhere to different Open Science criteria (more guidance for this can be found in the Hong Kong Principles, see Moher et al., *2020 (61)*. or highlight work that has had an impact on society. Reward collaborative works, value team science and put less emphasis on individual recognition— sometimes the effort put into making a collaborative endeavor a success goes beyond an individual researcher.

Heads of departments

Bring conversations about research integrity to the work floor. The fact that researchers' perception of research integrity differ between academic ranks and disciplinary fields may itself be a starting point for debate. To prevent this from becoming coffee-machine chat on what's wrong in science, consider asking your staff to reflect on questions such as: What research integrity dilemmas do we face? What sort of behaviors do we observe

efficiency can be systematically compared and based on which we would say, *regardless the circumstances,* that one is more important than the other. Or take the research integrity violations as mentioned in ALLEA (67), there is no higher order value based on which we can say that one violation (e.g. "Manipulating authorship or denigrating the role of other researchers in publications.", a moral wrong (p.8)) deserves heavier punishment than another (e.g. "Establishing or supporting journals that undermine the quality control of research ('predatory journals')., an epistemic wrong (p.9)).

to be rewarded that they may not be fully comfortable with?

Some dilemmas that we came across (in an intervention we are currently evaluating) regarded publishing. Take this dilemma from a senior researcher: Should I attempt to combine different results into one publication, bridging different fields and possibly having major impact (but getting it published may take a long time)? Or should I split up these results into different papers, possibly publishing them quicker and thereby increasing the chances that the PhD student I supervise finishes on time? The point of this example is that there is no right and wrong, but that the decision how to move forward may benefit from careful reflection.

Ensure these conversations are conducted respectfully, ideally following a method that helps to get more insight into why researchers face these dilemmas and which values and norms are at the core of the dilemma. There are different ways to structure debates about these sensitive matters, one of which is Moral Case Deliberation (62). To ensure these conversations are done attentively and professionally, bring in a skilled moderator. A moderator can help your team explore these dilemmas beyond surface-level. In addition, a moderator can make sure everyone participates in the conversation and assure it is confidential, structured and not (solely) based on existing hierarchical relationships.

Be willing to share your perspective, without leaning on your position of power. It might be very interesting for your staff to learn your perspective, as it may improve their understanding of why decisions were made a certain way. Sending around an email enquiring whether there is interest in having a conversation about research integrity dilemmas is good, but suboptimal. Instead, attend one of these conversations to demonstrate the importance of the topic, after all, there are many ways in which you can be responsible role model.

PhD supervisors

Take your tasks as a role model seriously: How do you convey responsible research? What sort of responsible practices do you adhere to and stimulate in your PhD student(s)? Do your PhD students feel sufficiently supported in their publication process? Have you taken their socialization into responsible research practices seriously?

Keep improving your supervisor skills. This regards research skills (e.g. how to maintain a good data management plan, how to write a publication plan or how to preregister your hypotheses) and soft skills (e.g. how to listen, how to give good feedback, how to be clear on mutual expectations). We are currently evaluating the results of a pilot training that combined these two skillsets and hope to share the results shortly.

PhD students

Communicate your expectations and invite your supervisor to explicitly communicate theirs. Supervision will remain insufficient or imperfect if you do not voice your needs,

concerns and connecting expectations (63).

Last but not least, for all stakeholders, be pro-active. Changing a research climate is hard and involves challenging existing practices on various levels. Against the pull of the system of science, it is easy to give in before even starting ("science is doomed, anyway"). Even the best navigators sometimes feel blown away, yet it is my impression that researchers in a responsible research climate are more resilient to the tides and in a better position to navigate the stormy seas of science. They are aware of the limitations of academic research and its inherent problems (and how these problems become especially visible during pandemic times (64), but it does render them pessimistic to the core.

Concluding remark

One focus group participant described the atmosphere in a responsible research climate rather succinctly. In his words, loosely translated, a responsible research climate is characterized by:

"…an atmosphere where the professor gives the good example. But where the good example is *really* listening, to people, to the data, where nothing but the truth matters and where nothing else is important…" [emphasis added]

Whether we agree with his exact wording (or connotations about the 'the' truth — philosophers beware) is of less importance. To me this quote symbolizes what is at the heart of organizational justice theory; if you want people (or in our case: academic researchers) to feel treated justly, you might do well to listen to them.

References

1. Fanelli D. How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. PLoS One. 2009;4(5):e5738.
2. Martinson BC, Anderson MS, de Vries R. Scientists behaving badly. Nature. 2005;435(7043):737–8.
3. Kornfeld DS. Perspective: Research misconduct: The search for a remedy. Acad Med. 2012;87(7):877–82.
4. Sovacool BK. Exploring scientific misconduct: Isolated individuals, impure institutions, or an inevitable idiom of modern science? J Bioeth Inq. 2008;5(4):271–82.
5. Steneck NH. Institutional and individual responsibilities for integrity in research. Am J Bioeth. 2002;2(4):51–3.
6. Casadevall A, Fang FC. Reforming science: Methodological and cultural reforms. Infect Immun. 2012;80(3):891–6.
7. Alberts B, Kirschner MW, Tilghman S, Varmus H. Rescuing US biomedical research from its systemic flaws. Proc Natl Acad Sci. 2014;111(16):5773–7.
8. George SL. Research misconduct and data fraud in clinical trials: Prevalence and causal factors. Int J Clin Oncol. 2016;21:15–21.
9. Antes AL, Brown RP, Murphy ST, Waples EP, Mumford MD, Connelly S, et al. Personality and ethical decision-making in research: The role of perceptions of self and others. J Empir Res Hum Res Ethics. 2007;2(4):15–34.
10. Tijdink JK, Bouter LM, Veldkamp CLS, Van De Ven PM, Wicherts JM, Smulders YM. Personality traits are associated with research misbehavior in Dutch scientists: A cross-sectional study. PLoS One. 2016;11(9):1–12.
11. Anderson MS, Shaw MA, Steneck NH, Konkle E, Kamata T. Research Integrity and Misconduct in the Academic Profession. In: Paulsen MB, editor. Higher Education: Handbook of Theory and Research. Springer Netherlands; 2013. p. 217–61.
12. Wible JR. Fraud in Science: An economic approach. Philos Soc Sci. 1992;22(1):5–27.
13. Broeckelman-Post MA. Building a culture of academic integrity: The role of communication in creating and changing understandings and enactments of academic integrity. ProQuest Dissertations and Theses. 2009.
14. Fang FC, Steen RG, Casadevall A. Misconduct accounts for the majority of retracted scientific publications. Proc Natl Acad Sci. 2012;109(42):17028–33.
15. Resnik DB, Dinse GE. Scientific retractions and corrections related to misconduct findings. J Med Ethics. 2013;39(1):46–50.
16. Fang FC, Casadevall A. Retracted science and the retraction index. Infect Immun. 2011;79(10):3855–9.
17. Steen RG. Retractions in the scientific literature: Do authors deliberately commit research fraud? J Med Ethics. 2011;37(2):113–7.
18. Pennock RT. Fostering a culture of scientific integrity: Legalistic vs. scientific virtue-based approaches. Prof ethics Rep. 2015;8(2):1–3.

19. Pennock RT, O'Rourke M. Developing a scientific virtue-based approach to science ethics training. Sci Eng Ethics. 2017;23(1):243–62.
20. Weed DL. Preventing scientific misconduct. Am J Public Health. 1998;88(1):125–9.
21. Anderson MS, Horn AS, Risbey KR, Ronning EA, De Vries R, Martinson BC. What do mentoring and training in the responsible conduct of research have to do with scientists' misbehavior? Findings from a national survey of NIH-funded scientists. Acad Med. 2007;82(9):853–60.
22. Titus SL, Wells JA, Rhoades LJ. Repairing research integrity. Nature. 2008;453(7198):980–2.
23. Gino F, Ayal S, Ariely D. Self-serving altruism? The lure of unethical actions that benefit others. J Econ Behav Organ. 2013;93:1–14.
24. Umphress EE, Bingham JB, Mitchell MS. Unethical behavior in the name of the company: The moderating effect of organizational identification and positive reciprocity beliefs on unethical pro-organizational behavior. J Appl Psychol. 2010;95(4):769–80.
25. Lombarts MJMH, Arah OA, Busch ORC, Heineman MJ. Meten en verbeteren van opleiderskwaliteiten van klinisch opleiders met het SETQ-systeem. Ned Tijdschr Geneeskd. 2010;154(7):1–8.
26. Fisher CB, Fried AL, Goodman SJ, Germano KK. Measures of mentoring, department climate, and graduate student preparedness in the responsible conduct of psychological research. Ethics Behav. 2009;19(3):227–52.
27. Simha A, Cullen JB. Ethical climates and their effects on organizational outcomes: Implications from the past and prophecies for the future. Acad Manag Perspect. 2012;26(4):20–34.
28. Edwards MA, Roy S. Academic research in the 21st century: Maintaining scientific integrity in a climate of perverse Incentives and hypercompetition. Environ Eng Sci. 2017;34(1):51–61.
29. Walker RL, Sykes L, Hemmelgarn BR, Quan H. Authors' opinions on publication in relation to annual performance assessment. BMC Med Educ. 2010;10(1):2–6.
30. Tytherleigh MY, Webb C, Cooper CL, Ricketts C. Occupational stress in UK higher education institutions: A comparative study of all staff categories. High Educ Res Dev. 2005;24(1):41–61.
31. VSNU, NFU, KNAW, NWO, ZonMw. Room for everyone's talent. The Hague; 2019.
32. NWO. NWO introduces narrative CV format in the 2020 Vici round [Internet]. News. 2019 [cited 2020 Jun 15]. Available from: https://www.nwo.nl/en/news-and-events/news/2019/12/nwo-introduces-narrative-cv-format-in-the-2020-vici-round.html
33. Phillips AW, Reddy S, Durning SJ. Improving response rates and evaluating nonresponse bias in surveys: AMEE Guide No. 102. Med Teach. 2016;38(3):217–28.
34. Cook C, Heath F, Thompson R. A meta-analysis of response rates in web-or internet-based surveys. Educ Psychol Meas. 2000;60(6):821–36.
35. Fulton BR. Organizations and survey research: Implementing response enhancing strategies and conducting nonresponse analyses. Sociol Methods Res. 2018;47(2):240–76.
36. Halbesleben JRB, Whitman M V. Evaluating survey quality in health services research: A decision framework for assessing nonresponse bias. Health Serv Res. 2013;48(3):913–30.
37. Liu M, Wronski L. Examining completion rates in web surveys via over 25,000 real-world surveys. Soc Sci Comput Rev. 2018;36(1):116–24.

38. Little TD, Rhemtulla M. Planned missing data designs for developmental researchers. Child Dev Perspect. 2013;7(4):199–204.
39. Baruch Y. Response rate in academic studies - A comparative analysis. Hum Relations. 1999;52(4):421–38.
40. Czajka JL, Beyler A. Declining Response Rates in Federal Surveys: Trends and Implications. Vol. I, Mathematica Policy Research. 2016.
41. Van Mol C. Improving web survey efficiency: the impact of an extra reminder and reminder content on web survey response. Int J Soc Res Methodol. 2017;20(4):317–27.
42. Bista K. Examining factors impacting online survey response rates in educational research: perceptions of graduate students. J Multidiscip Eval. 2017;13(29):63–74.
43. Martinson BC, Thrush CR, Crain AL. Development and validation of the Survey of Organizational Research Climate (SORC). Sci Eng Ethics. 2013;19(3):813–34.
44. Bouter LM, Tijdink J, Axelsen N, Martinson BC, ter Riet G. Ranking major and minor research misbehaviors: results from a survey among participants of four World Conferences on Research Integrity. Res Integr Peer Rev. 2016;1(17):1–8.
45. Gopalakrishna G. Disciplinary field specific detrimental research practices in the Dutch academic setting - A focus group study [Internet]. 2019 [cited 2020 Jun 19]. p. 1–15. Available from: https://osf.io/8vpwt/
46. Kraay A, Murrell P. Misunderestimating Corruption. 2013.
47. Gutmann J, Padovano F, Voigt S. Perception vs. Experience: Explaining differences in corruption measures using microdata. CESifo Work Pap No 8027. 2019;1–40.
48. Donchev D, Ujhelyi G. What do corruption indices measure? Econ Polit. 2014;26(2):309–31.
49. Elffers H, Van Der Heijden P, Hezemans M. Explaining regulatory non-compliance: A survey study of rule transgression for two Dutch instrumental laws, applying the randomized response method. J Quant Criminol. 2003;19(4):409–39.
50. Horbach SPJM, Halffman W. Innovating the peer review process : A publisher's ethnography. 2019.
51. Haven TL, van Grootel L. Preregistering Qualitative Research. Account Res. 2019;6(3):1–16.
52. Haven TL, Errington TM, Gleditsch K, van Grootel L, Jacobs AM, Kern F, et al. Preregistering Qualitative Research: A Delphi Study. Forthcoming in International Journal of Qualitative Methods.
53. Fiedler K, Schwarz N. Questionable research practices revisited. Soc Psychol Personal Sci. 2016;7(1):45–52.
54. Ruggeri K, van der Linden S, Wang YC, Papa F, Riesch J, Green J. Standards for evidence in policy decision-making. Nat Res Soc Behav Sci. 2020;399005.
55. Plemmons DK, Baranski EN, Harp K, Lo DD, Soderberg CK, Errington TM, et al. A randomized trial of a lab-embedded discourse intervention to improve research ethics. Proc Natl Acad Sci. 2020;117(3):1389–94.
56. Resnik DB, Rasmussen LM, Kissling GE. An International Study of Research Misconduct Policies. Account Res. 2015;22(5):249–66.

57. Resnik DB. Is it time to revise the definition of research misconduct? Account Res. 2019;26(2):123–37.
58. Steneck N. Fostering integrity in research: Definition, current knowlege, and future directions. Sci Eng Ethics. 2006;12(1):53–74.
59. Netherlands Code of Conduct for Research Integrity. 2018.
60. Peels R, de Ridder J, Haven T, Bouter L. Value pluralism in research integrity. Res Integr Peer Rev. 2019;4(18):1–13.
61. Moher D, Bouter L, Kleinert S, Glasziou P, Sham MH, et al. (2020) The Hong Kong Principles for assessing researchers: Fostering research integrity. PLOS Biology 18(7): e3000737. https://doi.org/10.1371/journal.pbio.3000737
62. Dam S Van Der, Abma TA, Kardol MJM, Widdershoven GAM. '"Here's my dilemma"'. Moral case deliberation as a platform for discussing everyday ethics in elderly care. Heal Care Anal. 2012;20(3):250–67.
63. Woolderink M, Van Der Boom H, Putnik K, Klabbers G. The voice of PhD candidates and PhD supervisors. A qualitative exploratory study amongst PhD candidates and supervisors to evaluate the relational aspects of PhD supervision in the Netherlands. Int J Dr Stud. 2015;10:217–35.
64. Gopalakrishna G, Bouter L, Mayer T, Steneck NH. Assuring research integrity during a pandemic [Internet]. BMJ. 2020. Available from: https://blogs.bmj.com/bmj/2020/06/08/assuring-research-integrity-during-a-pandemic/
65. Iskandarsyah A, Mutakin, Abdulah R, Tijdink JK. Publication Pressure in Indonesia [Internet]. 2020 [cited 2020 Jun 22]. Available from: https://osf.io/5dqyt/
66. NRC (National Research Council Canada). NRC Research Integrity Policy. 2013.
67. ALLEA (All European Academies). The European code of conduct for research integrity. 2017.

10

Chapter 10

Summary

What is the influence of the academic research climate on research integrity? How is this research climate perceived across academic ranks and disciplinary fields? Is it a climate wherein researchers perceive high publication pressure? Do publication pressure and the research climate play a role in explaining research misbehavior? And what is a responsible research climate?

In chapter 1, I discuss how the case of Diederik Stapel, who was found guilty for fabricating data, led to intense discussions in the Netherlands and elsewhere about research integrity. From these and other discussions on breaches of research integrity, two themes emerged that paved the way for this dissertation. First, although falsification, fabrication and plagiarism (FFP) are bad, they may not be the most pressing problem. It became more evident that there are a variety of questionable research practices (QRPs) that are much more frequent and therefore may be more harmful to academic research in the aggregate. Second, we should look beyond the individual researcher ('bad apple') and investigate the research climate ('barrel') given that what we perceive around us likely influences us profoundly and will influence what sort of research behaviors we engage in (and which not, as we judge them to be in opposition to 'what is expected').

I conclude with a brief description of the theory that most heavily influenced this dissertation, namely organizational justice theory. Organizational justice theory reasons that the fairer people feel treated, the more likely they are to trust their organization, accept its decisions and not engage in questionable behavior or worse. But the reverse is also true, and when applied to academic research one would expect that in an organizational research climate where the perceived injustice is high, researchers would be more likely to engage in research misbehavior or QRPs.

Before assessing the research climate, I first discuss in chapter 2 theories about why humans go astray that can and have been applied to cases of researchers that falsify and fabricate data. Can these theories help us to better understand cases of research misconduct? My answer is that they might, though all such explanations presuppose certain details about the case that are often unknown…

In chapter 3, I describe the perceptions of academic research climate for integrity per academic rank and disciplinary field, as it seems likely that these perceptions differ depending on the academic rank and disciplinary field. The results indicate that the perceptions of the research climate differ substantially; the humanities perceive their departments' expectations more negatively compared to other fields, whereas the natural sciences' perceptions of the research climate are more positive. Senior researchers' (by which we mean: full and associate professors) perceptions are most optimistic about the research climate, and significantly more positive than assistant professors, postdocs and PhD students.

In chapter 4, I describe how my colleagues and I revised the Publication Pressure Questionnaire (PPQ) in light of previous research showing that publication pressure could be one of the salient aspects that may hamper research integrity. The revised PPQ-r consists of three subscales: Attitude, Stress, and Resources. The previous PPQ measured Attitudes only. Yet, to validly conclude someone experiences publication pressure, we need to know whether they perceived high demands to publish *and* whether they had too little resources to cope with these demands. The latter element is crucial since we can all benefit from a little stress, as long as we have, say, supportive colleagues, to help us cope.

In chapter 5, I describe the degree of perceived publication pressure among academic researchers in Amsterdam. We find that especially the postdocs and assistant professors perceive high publication stress. The PhD students perceive the largest shortage of resources, like help with challenging journal editors. This shortage was less vivid among PhD students in the natural sciences, a field that reported less publication pressure in general.

In chapter 6, I describe researchers' perceptions of research misbehaviors using both survey data and focus groups data. We asked our survey participants to indicate how often they perceived some misbehaviors and how much impact they thought it would have on the validity of the study at issue. To get a sense of the most detrimental research misbehaviors on the aggregate level, we combined the frequency and impact scores and ranked the misbehaviors, stratified per disciplinary field. All top 5's contained one item about insufficient supervision, and the remainder regarded different forms of sloppy science. To ensure we got the relevant misbehaviors in focus, we presented the top 5 items from this list to our focus group participants and asked them to add other misbehaviors they actually perceived in their own work. Their discussions helped us to understand what insufficient supervision really meant. Researchers from the natural sciences and the humanities also came up with research misbehaviors that were not yet on our radar, such as the stealing of ideas or destroying evidence (before publication).

In chapter 7, I relate the perceptions of the research climate and the perceived degree of publication pressure to the perceived research misbehaviors. In other words, how much of the variance in research misbehavior can be traced back to a poor-quality research climate or a high degree of publication pressure? Together someone's academic rank, the research climate and publication pressure explain 32% of variance in perceived research misbehavior. The research climate notably explained 23% of variance. If we correct that for impact (after all, if the impact of the frequently perceived trespasses is judged to be negligible, why bother?), the explained variance due to someone's academic rank, the research climate and publication pressure is 18%.

In chapter 8, I present what characteristics our focus group participants associated with a responsible research climate, what barriers they perceived for a responsible research climate and which interventions they considered fruitful to alleviate those barriers. According to our participants, a responsible research climate would be characterized by fair evaluation, openness, sufficient time, integrity, trust, and freedom. The most noted perceived barriers were the unfair evaluation policies, the lack of support, normalization of overwork and insufficient supervision of early career researchers. Interventions that our participants suggested included training modules for PhD supervisors focused on responsible research, openly discussing expectations and dilemmas, creating formal research time, and sound evaluation policies.

In the chapter 9, I summarize the main findings and connect them to existing trends in research on research integrity. I conclude that these studies align with two trends, namely that QRPs may be a more pressing problem than FFP and that it may be useful to focus on the 'barrel' instead of solely on the 'bad apples'. I review some methodological limitations of our studies and end with a brief list of recommendations.


IO

II

Chapter II

Contributions of each author

Chapter 1. Introduction
T. Haven

TH wrote the draft manuscript and all subsequent versions. JT, RvW and LB commented on drafts and allowed the final version to be part of the dissertation.

Chapter 2. Explanations of Research Misconduct, and How They Hang Together
T. Haven, R. van Woudenberg

TH came up with the research idea, RvW wrote the outline. Both authors wrote parts of the manuscript and contributed to and approved the final version.

Chapter 3. Perceptions of research integrity climate differ between academic ranks and disciplinary fields
T. Haven, J. Tijdink, B.Martinson, L. Bouter

TH wrote the draft manuscript and conducted the analyses. BM helped with formal analyses. JT, LB and BM designed the study and contributed significantly to multiple versions of the manuscript. All authors approved the final version.

Chapter 4. Personally perceived publication pressure: revising the Publication Pressure Questionnaire (PPQ) by using work stress models
T. Haven, M. de Goede, J. Tijdink, F. J. Oort

TH wrote the draft manuscript. FO and TH conducted the analyses. TH, MdG, JT and FJO designed the questionnaire, study and collected the data. All authors contributed significantly to multiple versions of the manuscript. All authors approved the final version.

Chapter 5. Perceived publication pressure in Amsterdam: Survey of all disciplinary fields and academic ranks
T. Haven, L. Bouter, Y.Smulders, J. Tijdink

TH wrote draft manuscript and analysed the data. LB, YS, JT designed the study and contributed significantly to multiple versions of the manuscript. All authors approved the final version.

Chapter 6. Researchers' perceptions of research misbehaviors: a mixed methods study among academic researchers in Amsterdam
T. Haven, J. Tijdink. R. Pasman, G. Widdershoven, G. ter Riet, L. Bouter

LB, GW, JT, GtR, RP and TH designed the study. TH and JT collected and analysed the data. TH wrote the draft. LB, GW, JT, GtR and RP commented several times and the author team met to discuss the core issues. All authors approved the final version.

Chapter 7. Explaining variance in perceived research misbehavior
T. Haven, J. Tijdink, B. Martinson, L. Bouter, F. Oort

TH wrote the draft manuscript. FO and TH conducted the analyses. FO, BM, JT and LB designed the study. FO, BM, JT and LB contributed significantly to multiple versions of the manuscript. All authors approved the final version.

Chapter 8. Researchers' perceptions of a responsible research climate – a multi focus group study
T. Haven, R. Pasman, G. Widdershoven, L. Bouter, J. Tijdink

TH wrote the draft manuscript. All authors contributed to the study design. TH, JT, GW and RP analysed the data. JT, RP, GW and LB contributed significantly to later versions of the manuscript. All authors approved the final version.

Chapter 9. Discussion
T. Haven

TH wrote the draft manuscript and all subsequent versions. JT, RvW and LB commented on drafts and allowed the final version to be part of the dissertation.


S

Nederlandse samenvatting

Wat is de invloed van het onderzoeksklimaat op wetenschappelijke integriteit? Hoe ervaren onderzoekers vanuit verschillende disciplinegebieden en academische rangen in Amsterdam dit klimaat? Is de publicatiedruk die men ervaart te hoog? En spelen publicatiedruk en andere aspecten van het onderzoeksklimaat een rol in het verklaren van bedenkelijke onderzoekspraktijken of schendingen van de wetenschappelijke integriteit? En wat ìs een verantwoord onderzoeksklimaat?

In hoofdstuk 1 beschrijf ik hoe de casus van de Nederlandse hoogleraar sociale psychologie, Diederik Stapel, die in 2011 schuldig werd bevonden aan het verzinnen van onderzoeksgegevens leidde tot hevige discussies over de wetenschappelijke integriteit, zowel in Nederland als daarbuiten. Uit deze discussies kwamen twee inzichten naar voren die belangrijk zijn geweest in de totstandkoming van dit proefschrift. Ten eerste dat falsificatie, fabricatie en plagiaat (FFP) weliswaar een serieus probleem vormen, maar misschien niet de meest dringende kwestie. Er bleken namelijk vele bedenkelijke onderzoekspraktijken (aangeduid als QRPs, "questionable research practices) te bestaan die minder ernstig zijn dan FFP maar die, omdat ze zoveel vaker voorkomen, gezamenlijk uiteindelijk schadelijker zijn voor de wetenschap dan FFP. Ten tweede kwam naar voren dat het zinvol is om naast de individuele onderzoeker (de 'rotte appel') ook naar diens omgeving, het onderzoeksklimaat (de 'fruitschaal'), te kijken. Het onderzoeksklimaat speelt namelijk een rol in wat onderzoekers zien en beleven als gewenst gedrag, het bepaalt "hoe en wat we hier doen".

Tot slot beschrijf ik de belangrijkste theorie achter dit proefschrift, namelijk de *organisational justice theory.* Kortgezegd veronderstelt deze theorie dat hoe eerlijker mensen zich behandeld voelen door hun organisatie, eerlijker ze zich binnen de organisatie zullen gedragen, en hoe minder eerlijk ze zich behandeld voelen, hoe minder. Toegepast op academisch onderzoek zou dit betekenen dat in een onderzoeksklimaat dat als oneerlijk wordt ervaren, onderzoekers meer geneigd zouden zijn om bedenkelijke onderzoekspraktijken of erger te begaan.

Voordat ik het onderzoeksklimaat in Amsterdam ga evalueren beschrijf ik in hoofdstuk 2 welke theorieën er gebruikt worden ter verklaring van fraude in onderzoek. Helpen deze theorieën om bestaande gevallen van fraude bij onderzoekers echt beter te begrijpen? Het korte maar licht onbevredigende antwoord is: misschien, maar alle theorieën en verklaringen nemen bepaalde feiten en omstandigheden aan die vaak niet goed bekend zijn in het specifieke geval…

In hoofdstuk 3 beschrijf ik de percepties die Amsterdamse academici hebben van het onderzoeksklimaat betreffende wetenschappelijke integriteit. Hierbij maak ik onderscheid tussen de verschillende academische rangen en disciplinegebieden, want het lijkt plausibel dat de percepties van dit klimaat verschillen afhankelijk van het

vakgebied waarin iemand werkt of de positie die diegene heeft. De resultaten geven aan dat de percepties sterk verschillen tussen de disciplinegebieden. Onderzoekers in de geesteswetenschappen hebben in vergelijking met andere disciplinegebieden een negatievere perceptie van de verwachtingen van hun afdeling op het gebied van publiceren en subsidies binnen halen. Onderzoekers uit de exacte wetenschappen zijn over het algemeen juist meer positiever over het onderzoeksklimaat. Daarnaast verschillen de percepties afhankelijk van iemands academische rang: senior onderzoekers (waarmee ik bedoel: universitair hoofddocenten en hoogleraren) zijn het meest optimistisch over het onderzoeksklimaat en optimistischer dan universitair docenten, postdocs of promovendi.

In hoofdstuk 4 beschrijf ik hoe we de vragenlijst om publicatiedruk te meten, verbeterden. Eerder onderzoek toonde een verband aan tussen schendingen van wetenschappelijke integriteit en publicatiedruk. De verbeterde vragenlijst heeft drie subschalen: Attitudes, Stress en Hulpbronnen. De vorige versie was gericht op attitudes, de houding van respondenten ten opzichte van de publicatiedruk in de wetenschap. Nu kan iemand tegen publicatiedruk zijn, maar dat zegt nog niets over of die persoon daadwerkelijk publicatiedruk ervaart. De verbeterde vragenlijst vraagt, in de Stress-schaal, of ondervraagden ook zelf publicatiedruk ervaren, en in de Hulpbronnen-schaal of ze middelen hebben met die druk om te gaan (we kunnen allemaal wel eens profiteren van wat gezonde stress op het werk, zolang we maar, bijvoorbeeld, behulpzame collega's hebben die ons helpen met deze stress om te gaan!)

In hoofdstuk 5 beschrijf ik de mate waarin Amsterdamse onderzoekers publicatiedruk ervoeren. Uit de resultaten blijkt dat vooral postdocs en universitair docenten hoge publicatiestress ervaren. Promovendi rapporteren het grootste tekort aan hulpbronnen, zoals hulp bij lastige editors van wetenschappelijke tijdschriften. Dit tekort was minder nijpend onder promovendi in de exacte wetenschappen, een veld dat ook in het algemeen minder publicatiedruk rapporteerde.

In hoofdstuk 6 beschrijf ik hoe onderzoekers aankijken tegen bepaalde bedenkelijk onderzoekspraktijken, waarbij ik zowel data uit het vragenlijstenonderzoek alsook focusgroepdata gebruik. We vroegen vragenlijst deelnemers aan te geven hoe vaak ze de betreffende gedragingen van een vooraf opgestelde lijst hadden waargenomen en hoeveel impact ze dachten dat deze hadden op de validiteit van de desbetreffende studie. Om een indruk te krijgen van wat de meest schadelijke gedragingen waren op geaggregeerd niveau, vermenigvuldigden we de frequentie en impact scores en ordenden we deze van hoog naar laag (opnieuw één ranking per disciplinegebied). Elke top 5 bevatte wel een item over suboptimale supervisie en begeleiding van junior onderzoekers; de rest behelsde verschillende vormen van slordig onderzoek. Om nu zeker te zijn dat we de

relevante misdragingen in het vizier hadden, presenteerden we deze top 5 aan deelnemers in onze focusgroepen. We vroegen de focusgroep deelnemers om na te gaan of dit nu ook de misdragingen waren die zij in hun dagelijkse praktijk tegen kwamen en zo niet, welke misdragingen volgens hen dan belangrijker waren. Onderzoekers uit de exacte en geesteswetenschappen brachten misdragingen naar voren die wij nog niet in beeld hadden, zoals het stelen van originele ideeën of het vernietigen van bewijs.

In hoofdstuk 7 relateerden we de percepties van het onderzoeksklimaat en ervaren publicatiedruk aan deze waargenomen bedenkelijke onderzoekspraktijken en vormen van wetenschappelijke fraude. Met andere woorden: hoeveel variatie daarin kan toegeschreven worden aan de slechte kwaliteit van het onderzoeksklimaat of hoge publicatiedruk? Samen verklaarden het onderzoeksklimaat en publicatiedruk 32% van de variatie in waargenomen bedenkelijke onderzoekspraktijken en vormen van wetenschappelijke fraude. Het onderzoeksklimaat zelf verklaarde 23%. Wanneer we dat corrigeerden voor de toegeschreven impact van de betreffende gedragingen (immers, als de impact van de vaak waargenomen onderzoekspraktijken nagenoeg 0 is, wat maakt het dan uit?), was de verklaarde variatie in misdragingen die kon worden toegeschreven aan het onderzoeksklimaat en publicatiedruk samen nog steeds 18%.

In hoofdstuk 8 presenteren we welke kenmerken focusgroep deelnemers associeerden met een verantwoord onderzoeksklimaat, welke barrières zij ervoeren en welke interventies zij zinnig achtten om die barrières te slechten. Volgens onze deelnemers wordt een verantwoord onderzoeksklimaat gekenmerkt door eerlijke evaluatie, openheid, voldoende tijd, integriteit, vertrouwen en vrijheid. De voornaamste barrières bestonden uit: oneerlijk evaluatie-beleid, een gebrek aan steun, normalisatie van overwerk en suboptimale begeleiding van junior onderzoekers. Onze deelnemers stelden de volgende interventies voor: begeleiders van promovendi beter trainen in verantwoorde onderzoekspraktijken, het open bespreken van wederzijdse verwachtingen, het identificeren van dilemma's, het oormerken van de tijd die voor het doen van onderzoek beschikbaar is, en een gedegen evaluatiebeleid.

In hoofdstuk 9 vat ik de kernbevindingen samen en verbind ik deze aan trends op het gebied van wetenschappelijke integriteit. Ik concludeer hier dat de beschreven studies passen in de twee eerder besproken trends: dat bedenkelijke onderzoekspraktijken een dringender probleem zijn dan fraude en dat het zinvol is om te focussen op de hele 'mand' en niet enkel op een paar 'rotte appels' zoals gebeurde toen de affaire Stapel aan het licht kwam: hij werd gezien als een rotte appel in een overigens gezonde fruitschaal. Ik bespreek enkele methodologische beperkingen van het onderzoek en verbind deze met mogelijke verbeteringen. Ik sluit af met een beknopte lijst van aanbevelingen voor de praktijk.

D

Dankwoord

Misschien schrijf je een proefschrift wel voornamelijk alleen, maar in zowel de inhoud als in de moed om het te voltooien, heb ik bijzonder veel profijt gehad van een groot aantal mensen.

Lex, ik heb bewondering voor jouw onuitputtelijke arbeidsethos en niet op de laatste plaats je reactiesnelheid via de e-mail. Zit Lex in Taiwan? Geen probleem, binnen 10 minuten heb je mail terug. Je tact heb ik vaak geprezen, je hebt een haarfijn politiek gevoel. Daarnaast ben je kritisch, methodologisch goed onderlegd en nauwkeurig als het op terminologie en daarin consistent zijn aankomt. Dit heeft ervoor gezorgd dat ik werd uitgedaagd duidelijk te schrijven. Dank hiervoor, je hebt me scherp gemaakt en gehouden. Daarnaast waardeer ik de vrijheid en het vertrouwen dat je me hebt gegeven, onder andere om een uitstapje naar het Center for Open Science te maken.

Joeri, ik blijf me verwonderen over jouw onuitputtelijke humor en fontein van ideeën. Je hebt me in deze jaren geleerd dat niet alles haast heeft en bent me altijd blijven wijzen op de mooie zijden van de wetenschap. Ik heb warme herinneringen aan de *Superb Supervision* training en workshops de we samen leidden. Je bent iemand die moeiteloos in staat is mensen te enthousiasmeren of de anderzijds voor zich te winnen. In ons eigen onderzoek ben jij degene die talloze goede punten weet op te sommen, sterktes die welgemeend zijn. Als therapeut van de wetenschapper ga jij nog vele kwalen met humor en gezonde zelfreflectie bestrijden, maar o wee als iemand denkt dat een eigen website een goed idee is — pochen is uit den boze.

René, ik ben je ongelooflijk dankbaar voor jouw rol in het zien en verwezenlijken van mijn behoefte om dit proefschrift van een filosofische noot te voorzien. Je bent kritisch, je hebt een bloemrijk uitdrukkingsvermogen en een rijk arsenaal aan gedichten of andere passages paraat, vaak naadloos afgestemd op de situatie. Je bent een rolmodel, niet in de laatste plaats omdat je zo'n heldere schrijver bent, maar ook om hoe je je nog altijd kunt verwonderen. Je bent geduldig en bemoedigend, ik heb me altijd door jou gezien gevoeld als promovenda, maar ook als mens.

Frans, het is door jou dat dit proefschrift van methodologische kwaliteit is. Je hebt me behoed voor fouten en gestuurd in het begrijpen van instrument-validatie, factor analyse en multilevel analyse. Je hebt de bijzondere gave statistiek toegankelijk en helder uit te leggen. Je bent een ware leraar geweest deze vier jaar en ik prijs me gezegend met iemand zoals jij in mijn team.

Mikkel, thank you for supervising my master's thesis, for planting the seed in my head that doing a PhD might be something worth considering. And thank you for forwarding this very vacancy to me and putting in a good word as my reference.

Lieve Margreet, vanaf dag 1 stond vast dat jij mijn paranimf zou worden. Ik was bij jou toen ik het telefoontje kreeg dat ik op tweede gesprek voor deze positie mocht komen. Het voelt alsof we onze promoties, ondanks dat ze fysiek gescheiden waren, samen hebben doorlopen. Je hebt me altijd gesteund, of dat nu met een halve marathon is, een tweede master, of een wetenschappelijk drama waar ik bij jou ongegeneerd over kon foeteren. Je bent een van de slimste en meest veelzijdige mensen die ik ooit heb ontmoet en er zijn veel manieren waarop ik naar je opkijk. Nu we bijna ons eerste decennium als vriendinnen mogen inleiden, kan ik niet anders dan uitkijken naar wat het leven ons de komende tien jaar gaat brengen.

Lieve Thirza, wat was ik blij toen jij een maand na mij toetrad als promovenda filosofie. Ik ben gesteld geraakt op je open, kritische blik. Je stelt ongelijkheid genuanceerd maar scherp aan de kaak. Ik kijk blij terug op onze eindeloze gesprekken, ook wanneer er eigenlijk gewerkt moest worden. We hebben veel gelachen, maar je bent ook iemand bij wie ik kwetsbaar durf te zijn. Daarnaast bewonder ik de manier waarop je modern ouderschap combineert met de grillen van de academie, want ook waar dat schuurt, ga jij gestaag en geduldig door. Ik ben ook blij met onze gezamenlijke sportsessies en niet op de laatste plaats met ons wasrek, de uitvinding van het jaar!

Lieve Wout, filosofie-Wout, mijn 'andere' Wout, wat ben ik dankbaar dat ik vier jaar jou als collega mocht hebben. Ik ben gesteld geraakt op het moment dat ik vroeg de afdeling binnen kwam en jij al zat te werken. Ik herinner me vele, boeiende gesprekken in de ochtend bij de koffie, rond middaguur in de Botanische tuin of op je Kantiaans vaste wandeling. Jij hebt me het sprinten bij het concertgebouw bijgebracht! Daarbij waren we een goed duo op kraamvisite, waarbij het heerlijk is dat jij zo goed met kinderen bent. Je hebt gelachen om hoe makkelijk ik op de kast te krijgen was, maar was altijd bereid mijn worstelingen aan te horen. Je hebt me uitgedaagd afstand te doen van mijn talloze anglicismen, wat makkelijker gezegd dan gedaan bleek.

Lieve filosofie-collega's, wat een fijne afdeling houden jullie samen in stand. In het onderzoek doen naar de rol van het onderzoeksklimaat, heb ik me vaak gerealiseerd hoezeer ik bofte met het klimaat waarin ik me bevond. Eenieder van jullie is behalve briljant, ook geïnteresseerd in een ander perspectief en staat open voor vragen. In het bijzonder: Naomi, Hans, Lieke, Judith, Jojanneke, Ruben, Irma, Eefje, Linda, Elisa, Nina, Merel, Elias, Nora en Roland, wie ik allen dankbaar ben voor de lunches, nieuwe ideeën en steun. En daarnaast Rik en Jeroen, wiens heldere ideeën en goede humor ik bijzonder heb gewaardeerd. Tot slot: Marije, wat was het fijn om jou als rolmodel te hebben.

Dear research integrity colleagues (a group that has, in part thanks to the Research Integrity lunches, become more and more of a group), thank you for your suggestions, ideas, solutions, and cups of coffee when human support was most needed. Notably, I want to extend my gratitude to Fenneke, Coosje, Mario and Dorien — I can not imagine fulfilling this PhD without you.

Tom Paffen, ontzettend bedankt voor het pontificaal beschermen van het belang van privacy in onderzoek naar onderzoeksintegriteit. Jouw humoristische doch kritische blik hebben dit project behoed voor menig drama.

Lieve meisjes, lieve Marloes, Noraly, Josien, Lynca, Lois en Maartje, wat een bijzonder idee dat we al vanaf TAKC al elkaars hoogte- en dieptepunten hebben meegekregen. Jullie kennen me door en door en sporen me altijd aan mijn eigen keuzes te maken. Zolang al in elkaars leven figureren maakt dat je bepaalde manieren van pretentie achter je laat. Ik geniet van onze weekendjes weg, diners of andere vormen van samenzijn, waar ik altijd opgeladen uit wegga.

Lieve Jackies, lieve Nora, Liesbeth, Eva, Celine, Yi An, Anouk, Veerle, Nynke, Lieke en Emma, nu hebben jullie eindelijk behalve dokters ook een doctor in de club (en nog een op komst!). Als zussen hebben jullie me gesteund in het weggaan bij een eerdere baan en het beginnen van deze wetenschappelijke loopbaan. Jullie helpen me successen te vieren en ongegeneerd mezelf te zijn. En zijn niet te beroerd me erop te wijzen als het licht is uit gegaan.

Lieve Lisette (en Rik), wie had gedacht dat toen we samen psychologie begonnen te studeren, we nu (praktisch) tegelijkertijd zouden promoveren? Ik ben trots op wat je doet en bewonder jouw *practice what you preach.*

Lieve Kirsten, je hebt altijd al mijn verwonderingen, nieuwe vragen, manieren waarop psychologie anders moest, geduldig aangehoord. Ik heb ontzettend veel respect voor de professie die jij nu uitoefent, *you rock*.

Lieve Bruce (en José), je hebt me nog wiskunde bijles gegeven, had jij ooit gedacht dat ik een master in de epidemiologie zou voltooien? Je volgt me zorgzaam en getrouw en ik had een brok in mijn keel toen jij de eerste was die me persoonlijk belde toen mijn proefschrift was goedgekeurd. Ik bof met vrienden zoals jullie!

Lieve Juliette, zo gezellig dat we samen in Zwolle, Groningen en toen ook nog Amsterdam zaten. En zeg nou zelf, we zijn toch leuk opgedroogd?

Lieve Charlotte en Sietse, grappig hè, dat ik terwijl jullie niks hadden opgedrongen, toch psychologie in Groningen ging studeren? Ik houd van de momenten waarop jullie mijn uitspraak verbeteren of woordkeuze corrigeren, maar meer nog houd ik van onze lange wandelingen of fietslingentjes buiten. In mijn eigen intellectuele ontplooiing hebben jullie altijd pal achter me gestaan. Ik ben gesteund in het verwezenlijken van een eigen school waar kinderen zelf bepalen wat ze leren (zijn toch de nodige paralellen met een promotie te trekken?). Maar ik ben ook aangemoedigd uit te blinken in het reguliere onderwijs, ook als dat wiskunde studeren op de achterbank van een auto in Portugal of Spanje betekende, want leren voor een toetsweek kon uitmuntend worden gecombineerd met een vakantie. Dat die intellectuele ontplooiing gepaard ging met de nodige sprongen over de zee of oceaan, namen jullie als mooi excuus om de Rocky Mountains of de Highlands te bewonderen. Ook nu nog, ijken jullie geregeld mijn morele kompas, of luisteren jullie geduldig naar mijn tirades. Dan volgen er wijze adviezen zoals *choose your battles* of onze signature *sandwich hugs*.

Lieve Wout, wat een meesterzet van jou om enkele weken voor mijn sollicitatie bij de VU gehoor te geven aan mijn gedram om een konijn in huis te halen. Met Dotty wist je immers zeker dat ik voorlopig in dit kikkerlandje zou blijven. Je bent ongelooflijk solidair maar zult nooit je eigen waarden opgeven. Dus à la het Nederlandse poldermodel ging je wel naar Amsterdam maar niet naar een plek waar je je door eindeloze stromen toeristen moest wurmen om naar huis te komen (lees: mijn vorige woning). Je bent me altijd achterna gekomen, of het nu betekende dat je na een luttele 5 maanden verkering je scooter en je iPad moest verkopen voor een ticket naar Canada, of dat je je lening fors moest ophogen om me om de maand in Edinburgh te bezoeken. In de periode dat ik dit proefschrift voltooide, tolereerde je dat ik soms avonden door zat te werken. Vaker was je echter degene die me overtuigde mijn laptop dicht te klappen of mijn mail van mijn telefoon te verwijderen. Je houdt me met beide benen op de grond. Je doorziet sociaal spel en hebt het ongeëvenaarde vermogen om me emotioneel te verstaan. Je hebt een muur om me heen geduldig neergehaald. Je geeft me alle ruimte en bent degene die me met de dikste berenknuffels weer doet inzien dat ik alle reden heb om zielsgelukkig te zijn wanneer dat besef me even ontglipt. De kers op de taart is dat je me afgelopen februari in de Verenigde Staten kwam opzoeken en aldaar op je knieën ging. Ons staat een geweldige toekomst te wachten, *better together*.

What is the influence of the academic research climate on research integrity? How is this research climate perceived across academic ranks and disciplinary fields? Is it a climate wherein researchers perceive high publication pressure? Do publication pressure and the research climate play a role in explaining research misbehavior? And what is a responsible research climate?