university of groningen

Charting the Constellation of Science Reform

PhD Thesis

to obtain the degree of PhD at the
University of Groningen
on the authority of the
Rector Magnificus Prof. C. Wijmenga
and in accordance with
the decision by the College of Deans.

This thesis will be defended in public on

Thursday 25 August 2022 at 12.45 hours

by

Sarahanne Miranda Field

born on 25 January 1989
in Broken Hill, Australia

Supervisor
Prof. H.A.L. Kiers

Co-supervisor
Dr. M. Derksen

Assessment Committee
Prof. M.A.L.M. van Assen, Tilburg University
Prof. P. de Jonge, University of Groningen
Prof. A.J. Pols, University of Amsterdam

Charting the Constellation of Science Reform

S. M. Field

Printed by RidderPrint
Cover: S. M. Field

Contents

Prologue

This PhD project began as something entirely different to what it is now. My background is in quantitative psychology, having completed a bachelor and honours degree in what was known colloquially as the 'rats and stats' department, at the University of Newcastle in Australia. I trained in a neuroanatomy lab for years in my home country, and was quite content with the idea of a future in brain science until a brush with metascience in 2014 at the end of my undergraduate. I began my first paper that year, a pair of replication studies which eventually got published in *The Journal of Experimental Psychology: General* (Field, Wagenmakers, Newell, Zeelenberg, & van Ravenzwaaij, 2016). I moved to the Netherlands in 2015 and promptly began a Research Master specializing in statistics and psychometrics.

Every assignment in which I could choose a topic myself, revolved around metascience in one way or another. Three such assignments turned into my second, third and fourth scientific articles (Field, Hoekstra, Bringmann, & van Ravenzwaaij, 2019; Field et al., 2020; Field & Derksen, 2021). At this point, I had the vision of combining my deep love for metascience with methodology in future study endeavours. After all, I've come to believe that a good methodologist is never obsolete, no matter the job climate. My sights have always been trained on obtaining a PhD: the idea of going deep into a topic appeals to me on a fundamental level, not to mention that I love learning for learning's sake. Based on my training, I don't think I can be blamed for thinking my interest in that PhD study should or would lie purely in the realm of quantitative empirical research.

After a year of reading, writing the start of my thesis, drafting and reading some more, I began to experience a shift in my interest. From the quantitative, methods-focused proposal I had conceived (guided first by E.-J. Wagenmakers, and later by Henk), to a primarily qualitative sociological focus. I can't pretend that I noticed it myself – Henk and Maarten were the ones who first brought it to my attention. Totally true to my character, I initially denied it. I am headstrong and assertive, by nature. You can lead me to the stream, but unless I have decided

for myself that I definitely want to drink, I won't drink a drop. So, naturally, the first meeting where the shift was mentioned, I was adamant that continuing on the original path was best. *Although my growing interest couldn't be denied, I truly should be following the quantitative path I'd so painstakingly laid out for myself.* So went my internal dialogue.

True to the characters of both the supervisors, they allowed me to lead the way. They've both got wisdom gained from years of supervision of others, and some time previously working with me. They knew I'd come around, and trusted that that timing would be the best time to make adjustments to the project. They were right. I did come around to the idea, a month or two (or six) later. Thankfully my supervisors are supportive, sensitive people who thought it in the best interests of both myself and the project to allow me follow this wholly accidental change in direction.

If you keep reading, you'll learn about a subculture within broader academia. This subculture has emerged in response to a recent huge shift in perceptions of the trustworthiness and validity of the field, resulting in a relatively sudden and fast-growing increase in interest in metascientific and reform topics. I have studied and documented how this community arose, how it defines itself, what its members discuss with one another, what they do in terms of their research and research-related practices, and how and why they do it. I have conducted ethnography, physically and virtually, combining observations, interviews and surveys. I have taken a special interest in the online platform which hosts much of this discussion, Twitter, and add tweets – bite-sized texts – to the mix of information that I call my research material. I have added a few reflections to the text too, which give the reader insight into myself as the research instrument.

1

Introduction

1.1 The Rise of Science Reform

The 21st century ushered in a period of change for science; a shift in perspectives and priorities. According to the literature, it began in psychology and related behavioral sciences, rippling outward to affect most of the scientific community (Spellman, Gilbert, & Corker, 2018). Much of this change is positive, it would seem. Pashler and Wagenmakers discuss the establishment of a clearer distinction between good and bad research practice, an emphasis on transparency, integrity, inclusion and diversity, and a new-found appreciation for reproducibility (2012). It is also marked by the catalyst of that progress; a 'crisis of confidence' (Pashler & Wagenmakers, 2012). The crisis of confidence has many potential causes (and, certainly, whether or not there is even a crisis to discuss is a point of contention; Stroebe & Strack, 2014), but most salient in the literature are reports of fraud and widespread questionable research practice (see Stroebe, Postmes and Spears for a discussion and analysis of these issues in relation to the myth that science is self-correcting: 2012), and failures to replicate many of the field's cornerstone findings (the Open Science Collaboration discusses replication concerns in the light of the crisis, and argues how wide-scale replication attempts will lead to improvement of confidence in science: Open Science Collaboration, 2012). Concerning evidence of these problems has come to light over time, undermining the faith of researchers in their own disciplines' findings, and of lay-people's trust in the enterprise of science (Field et al., 2020), which, among other goals, aims to advise them and keep them safe.

Naturally, the idea that the foundational research of one's field is somehow contaminated or rotten, has spurred many scientists into action, and a 'social movement' was born (Peterson & Panofsky, 2020). The movement's joint enterprise concerns reforming science – to better the ways in which research is planned, conducted and reported – through driving transparency and openness in the scientific process up, and misconduct and poor research practices down.

Most of the people behind the movement, at least in its earlier days, were researchers who took on the burden of contributing to reform and metascientific literature alongside existing content research (Spellman et al., 2018). Others, as the movement progressed, began to shape their entire academic careers around an empirical and 'activism-minded' kind of reform science, or metascience.

This motley group has, over the period of more than a decade,[1] become a community of sorts. They have developed formal societies and sub-communities, a large and growing body of literature, as well as scientific journals in which to publish this research, and software and platforms. They have established a culture, with rules and norms, and punishments for breaking them. There is joint practice, shared repertoire, and knowledge, skills and competencies. All of it has come into being in the service of scientific reform.

I embarked on a research project to understand this community and their goal of scientific reform. *Who are the individuals within the group? What sets them apart from the rest of academic science, and how do they define themselves as a collective? How do they practice reform research and activism, and what are important topics the group explores together during this practice? How have the boundaries setting the group apart from the greater academic community come to exist, and how do they engage with one another within these boundaries? What happens when someone violates norms and expectations? What is the social structure of the community, and what is the nature of the ties between different actors within the community?* This dissertation contains my exploration of these questions and others related to them.

1.2 Chapter Outline

In what follows, **Chapter 2** and **Chapter 3** contain a review of selected literature and incidents which give a background for the most important elements underpinning the research in this dissertation. First, I discuss the crisis of confidence, and describe some of the potential catalysts of the crisis. I present some of the events and topics that have been central to fueling the reform movement's momentum. I then provide an overview of the main practices and initiatives that the community have generated throughout the course of the movement, to date.

In **Chapter 4**, I break down what community might mean, and how it applies to online groups. I provide a general introduction to Twitter, the 'microblogging' site and relate the challenges of defining online community to conducting research using Twitter. As I will show, Twitter hosts much of the activity of the reform movement, and it has been a rich source of data for this project.

[1] Some researchers, such as Wagenmakers (2012), pin the beginning of the current crisis of confidence as having reached some kind of momentum around 2011, though Spellman, Gilbert & Corker (2018) refer to 2012 as the year in which concerns about the state of psychology began to surface.

In **Chapter 5** I describe my ethnographic practice. I begin by reflecting on what the concept of 'the field' means to me, in the context of my research which takes place in online spaces as well as in person. I move on to describe methods like observation and interview, and present how perspectives on different aspects of these methods have contributed to my own practice of them in the context of this study. I share the methodological frameworks from which I have drawn during my study, and reveal how feminist influences have made their mark on how I have conducted my research. Finally, I describe the ethical principles which have guided my handling of the data and materials I have amassed throughout the course of my doctoral research.

Chapter 6 contains an outline of Etienne Wenger's Community of Practice framework (1998). The outline introduces key concepts in the framework, to provide the reader with the theoretical background knowledge and terminology needed to understand how I frame the interpretations of the ethnographic research material I offer in **Chapter 7**.

In **Chapter 7** and **Chapter 8** I present my exploration of the research material. In **Chapter 7**, I provide my analysis of the qualitative material I collected during my virtual and in-person ethnographic fieldwork. This chapter investigates what certain labels and classifications mean to different sub-communities within the greater structure. It investigates the culture, structure and institutions of the community, and outlines how it engages with challenges and controversy within its own walls, as well as with scientific obstacles affecting the greater scientific community. I also consider how the group engages with the greater scientific community, and how it is perceived by the outside. As I mentioned above, the interpretations I share in this chapter find meaning and legitimacy in their relation to concepts Wenger explores in his 1998 book on Communities of Practice. I consider the negotiation of identity and shared meaning, consider the duality of participation and reification and explore issues of conflict and power in the group, in the light of the idea that these disrupt participation and lead to peripheralization of some of the group's membership. This chapter emphasizes one central consideration, that is, whether or not the group can be reasonably considered a *community*.

I argue that although the open and reform science group *can* be considered a community, the discontinuities, fractures, lack of cohesion, and modularity that characterize it make such a categorization problematic. I discuss how treating the group as though it is a community glosses over important heterogeneity and diversity – elements of the group which drive much of its practice. In the latter part of **Chapter 7**, I advance an argument to re-consider the structure of the reform and open community – as a *constellation*, or *group* of communities. I contend that when the group is considered a constellation, the contours of different practices emerge and their competencies can be harnessed to push the joint enterprise – to improve science – forward.

Alongside my qualitative findings, I offer analyses reinforced by quantitative

Twitter-derived data. In **Chapter 8**, I use methods from a quantitative social network analysis toolbox to describe the possible structure of the community's social network online. I take three different approaches to classifying Twitter users as members of open science and reform communities. I collected data using the words people use to describe themselves in their Twitter biography texts, using a common community hashtag (a kind of special keyword), and using a conference hashtag, and describe the qualities of the three network data sets these approaches produced. I analyze metrics like transitivity to explore the nature of the ties between actors within these networks, as quantified by the connections they have made on Twitter (i.e., who they follow and who follows them within the bounds of the corpus defined by the classification approach). With the results of a modularity algorithm (used to detect sub-communities in social networks), I explore the idea of a constellation of practices further. I break down the largest of the three networks, and further analyze the users contained in these subgroups. I consider the results in the light of the constellation structure that I argued for in **Chapter 7**, and provide some corroboration for the constellation idea.

Ultimately, through a marriage of qualitative and quantitative approaches, and using the lens of Wenger's community of practice framework to interpret my findings, I arrive at the conclusion that this group of open science, metascience and science reform people are not a single, homogeneous community, but that *they represent a vast constellation of overlapping yet distinct communities of practice*, which share the joint enterprise of science reform. I argue that despite sharing a common enterprise, each different community of practice approaches the enterprise from a different angle, and that contributorship looks different depending on which group you talk to. I contend that failing to encourage diverse competencies and contributions may cause the movement to stall, while facilitating the different approaches of each community of practice in the constellation can culminate in a fruitful collective push to irrevocably change scientific philosophies and practice.

In line with feminist anthropology, I emphasize that I, in my capacity as ethnographer, am a person with my own distinct biography (a fact which is surprisingly largely under-explored even in today's literature; Hastrup, 1992). My stance as a participant observer (i.e., I am studying a community in which I am a member; I explore this 'condition' in **Chapter 5**), and my personal reflexive research style and proclivity for scientific transparency have strongly influenced how I have conducted my research, and how I have reported on it in this doctoral thesis. I especially value exploring *how* I have conducted this work, which means that the description of the methodology is longer and more detailed than what seems to be the norm for a written ethnographic report.

My open personality and love for reflexivity is evident in this dissertation too. I take seriously Malinowski's assertion that the essence of ethnography is to confront myself (2020), in the pursuit of this research. In the words of Hastrup, I have "exploited the paradox of fieldwork as an intersubjective mode of objectivi-

sation", and in doing so, have transformed myself from "spectator to seer", and the knowledge I have gained during the ethnography into insight (p. 118, *ibid*). To recognize some of this process as it has unfolded, I provide occasional excerpts of reflections and experiences where appropriate. These autoethnographic snippets, derived from my field notes, appear in italics throughout the text.

2

Psychology's Crisis of Confidence

2.1 Crises: Present and Past

Psychology, like any other discipline of science, is dynamic, and, like many scientific disciplines, has undergone changes over time. It has experienced periods of prosperity and periods of unrest, with internal discourse shifting to match the field's focus. The current crisis of confidence is only psychology's latest. This section reviews the current crisis and the crisis of the 1970's in psychology, comparing and contrasting them to provide a context for some of the later content in this dissertation.

The Current Crisis

Arguably, first signs of the current crisis can be traced back to 2005, when Stanford professor and epidemiologist John Ioannidis published the controversial essay: 'Why Most Published Research Findings are False'. It sketches out a theoretical analysis, arguing that more than half of published research results may be false positives, and should not be expected to replicate. Though it has been the target of heavy criticism, the article has made quite an impact since it was published (it has been cited by more than 10,000 other articles at the time of this writing, and is the most downloaded paper in the Public Library of Science).[1] Moreover, his assertions have been quantitatively reinforced since the essay was published (see, for instance, the OSC report which demonstrated that 61 out of 97 published significant results could not be successfully replicated: Open Science Collaboration, 2015).

By some accounts, the concerning state of affairs in psychology had reached a kind of critical mass in around 2011, in the years since Ioannidis's paper was

[1]For instance, see two articles from 2007 – Goodman & Greenland's (2007) analysis which attempts to moderate Ioannidis's claims, and Moonesinghe, Khoury & Janssens (2007), who argue for the benefits of replication in helping address false findings.

published (e.g., Pashler & Wagenmakers, 2012; Wagenmakers, Wetzels, Borsboom, van der Maas, & Kievit, 2012). Wagenmakers (2012) chronicled a set of distinct events, which bookmark this dark chapter in psychology's history. As Wagenmakers tells it, psychology's "annus horribilis" (p. 12) began with a case in which one of psychology's flagship journals published an article showing evidence for seeing into the future, then refused to accept articles showing compelling evidence to the contrary. Other key psychological phenomena such as social priming were also in the spotlight. In September of 2012, Nobel laureate Daniel Kahneman published an open letter with the subject line "A proposal to deal with questions about priming effects". He wrote about doubts concerning priming studies, stating that priming research is the "poster child for doubts about the integrity of psychological research". He closed the letter by encouraging replication and transparency, and emphasizing the importance of collective action. Next in Wagenmakers's laundry list of horrors was a wide-scale fraud case which shocked psychology, and broke into the scientific and lay-media. It involved a high-profile academic admitting to having fabricated data in at least 30 publications, then relinquishing his doctorate title in the wake of public backlash.

Other, more subtle scientific misconduct had also become the subject of scholarly scrutiny. In 2009, methodologist Daniele Fanelli published a meta-analysis of self-report surveys on research practices which gained much traction in the behavioral sciences. The article revealed that while only 2% of researchers admit to data fabrication, up to 72% of researchers report to have engaged in some kind of questionable research practice (QRP). In 2012, a study whose message echoed that of Fanelli's 2009 paper, reported concerning statistics on the prevalence of QRPs in psychological research fields (John, Loewenstein, & Prelec, 2012). The authors surveyed more than 2000 researchers, and reported that almost 10% of researchers in psychology had knowingly "introduced false data into the scientific record" (p. 526). They also estimate that approximately 94% of psychologists, by their own admission, had engaged in at least one practice that is considered questionable. These high rates are not the preserve of psychological research. For instance, a similar report on QRPs has recently emerged for ecology and evolution biology research based on data from 807 participants (Fraser, Parker, Nakagawa, Barnett, & Fidler, 2018). Fanelli (2018) discusses similar concerns relating to clinical data for cancer trials, among other examples.

Over the years, more signs of disquiet in the field have appeared. More cases of fraud, both proven and alleged have come to light. The website RetractionWatch, started in 2010, which aims to track "retractions as a window into the scientific process" gives a rolling account of cases as they are made public knowledge. In 2021, they reported that 25,000 retractions have been logged in their database (Oransky, 2021). Problems with tone in scientific critique are also rife (for an analysis, see: Derksen & Field, 2021), and bitter spats between researchers (e.g., between John Bargh, a prominent social psychologist, and a team of researchers that attempted to replicate some of his earlier work; see Yong, 2012a)

are not uncommon.

The '70s Crisis of Confidence

The current crisis is reminiscent of another crisis in psychology that took place in the 1970s, which, in fact, has been called the 'original' crisis (Faye, 2012; Hogg & Williams, 2000; Hornsey, 2008; Strickland, Aboud, Gergen, Jahoda, & Tajfel, 1976).[2] Originally spotlighted by Elms (1975), the crisis of the 1970s came on the heels of rapid growth in social psychology after World War II. The field was troubled by 'faddism', said to be "directionless", and was reportedly preoccupied by the pursuit of research topics that were deemed frivolous and socially irrelevant (p. 514; Faye, 2012). Its scholars were increasingly concerned about the field's "ontological, epistemological, and axiological limitations" (p. 274; Ellis et al., 2011). Strickland and colleagues (1976) spoke of the crisis in terms of a sickness in the social sciences, referring to "symptoms of an underlying malaise" (p. 3).

The literature leading up to this 'breaking-point' of sorts was full of questions and criticisms of the current methods, topics, theory, relevance of the findings, as well as future directions in the field. Kuhn (1962) warned that the then-current model of scientific inquiry in psychology lacked a strong foundation, and the "most revered premises of scientific truth and knowledge" were the subject of "demoralizing critiques" (Bochner & Ellis, 2016). Some accounts describe how the field arose from its own ashes, to flourish once more. The 1970s crisis was taken seriously by many in the field as an opportunity to "reform social science and re-conceive the objectives and forms of social science inquiry" (p. 273; Ellis et al., 2011), and psychology made progress in the wake of this reform, imply Ellis, Adams and Bochner. As Ellis and her colleagues tell it, interest shifted from the finding of 'truth' and 'fact', to attempting to recognize the complexity of psychological research, reconsidering it as a way to explore meaningful phenomena, laden with morals and ethics. According to Ellis and colleagues, during the throes of the 1970s crisis, researchers felt an increasing "need to resist colonialist, sterile research impulses of authoritatively entering a culture, exploiting cultural members, and then recklessly leaving to write about the culture for monetary and/or professional gain, while disregarding relational ties to cultural members" (p. 274). This new-found sensitivity saw many scholars post-crisis seeking autoethographic methods, to concentrate on the production of meaningful and accessible science, which could sensitize audiences to issues of identity politics, underrepresented voices, and other forms of representation which deepen a reader's capacity to relate to others perceived as different.

[2] According to Faye (2012), scholars were referring to a 'post-crisis stage' as early as 1976, though it has also been attributed to having occurred at some point in the 1980s (see: Ellis, Adams, & Bochner, 2011). Given that concerns over the field were mounting by the mid 1960's, and "crisis of confidence" was coined in 1975 by Elms, I will refer to it as being of the decade of the 1970s.

The past: A lens through which to consider the present

The current crisis has no shortage of similarities to the one in the '70s. Controversial and trendy findings (like Bem's PSI and Carney, Cuddy and Yap's power posing effect), discussions about poorly developed and questionable theory (e.g., theory behind the ego-depletion effect: Pappas, 2016) and questions about the future of psychology abound (Resnick, 2016).[3] Horton (2015), in an online comment piece for *The Lancet* hands science a chilling judgment: "something has gone fundamentally wrong with one of our greatest human creations". In Horton's bleak view, science has "taken a turn towards darkness", and is plagued by "flagrant conflicts of interest" (no page numbers given). The words of Chambers are similarly grave. He cautions that a serious course correction for psychological science is necessary to avoid a fate like that of phrenology: "... like so many other 'soft' sciences, we found ourselves trapped within a culture where the *appearance* of science was seen as an appropriate replacement for the *practice* of science" (p. VIIII: Chambers, 2019).

Ferguson and Heene (2012) labeled the psychological literature as a "vast graveyard of undead theories" – a striking contrast to the image of a healthy flourishing literature the lay-public might envision when they think of science. Luckily, just like in the 1970s, the current crisis has been and continues to be a powerful catalyst for change in the social sciences. In the wake of the crisis, argues Chambers, the focus is largely constructive, revolving around rebuilding the "castle of science", rather than on its demolition and re-construction (2019). Concerns over the impact of unreproducible findings and poor scientific practices have led to a sharp increase in the number of incentives and practices generated by different groups in the field to improve the state of science.

Despite the similarities, the crises differ in several key aspects. First, while the 1970s crisis involved what seemed to be an extended and thoughtful debate about issues and potential remedies, the current crisis has prioritized a quantitative estimation of the extent of the problems. Studies like Fanelli's (2009), mentioned earlier, are used to communicate the severity of the issue, and aim, at least partly, to bolster conceptual elements of the debate.

Another difference regards too little versus too much: In the 1970s, psychology seemed to be at a loss for "important problems to investigate and models to employ [social psychology's] research and theory" (Elms, 1975). In comparison, the current crisis refers to a field that features a wealth of research problems

[3] I should mention a minor sticking point here: While the crisis of the 1970's was a crisis in psychology (in social psychology in particular) it was not reported to have affected science more broadly. This is a similarity too, however, because the present crisis centers on social psychology findings (or at least, largely did in the beginning of the saga), just as the previous crisis did. From now on, I will broaden the remit of the discussion to include science at large, rather than only focusing on crisis discourse in psychology. This reflects that my coverage of the literature widens, but also that, as we progress through the crisis chronology, the 'radius' of the concerns and discussion surrounding them increases.

as well as models to investigate them. Psychological research seems to have a quality deficiency, as I have already discussed. Others have pointed out an insufficient understanding of the mechanisms behind the countless complex effects and phenomena reported (a problem Cesario described as being symptomatic of the relative "infancy" of the field, p. 45; 2014). For instance, some have argued that science reform has prioritised methodological and statistical issues of sample size, *p*-hacking, replication and transparency, while largely neglecting issues of theory development, which may be crucial to reaching reform goals (see e.g., Devezer, Navarro, Vandekerckhove, & Buzbas, 2020; Gervais, 2021).

A third difference lies in the general tenor of the debate. In the 1970s, psychologists seemed greatly concerned with reflecting on their own role and responsibility in the crisis. The problems of a loss of momentum and purpose of the field largely concerned an issue with *identity*. Iconic and controversial arguments such as those advanced by Kenneth Gergen (who considers that perhaps psychology ought not to be thought a science at all) sparked debate about how psychology should see itself going forward (see, for instance: Gergen, 1973, 1985). Researchers discussed and debated how to better their field, in an era where psychology was 'brimming with self-reflection' (p. 48; Bochner & Ellis, 2016). By contrast, the current crisis prominently features finger-pointing and ad-hominem attacks between researchers, where direct accusations of incompetence and data fabrication are hurled around between people and fields, and as I mentioned before, tone problems are numerous (see e.g., Engber, 2017; Derksen & Field, 2021; Bastian, 2016).[4]

Examples include allegations of witch hunting, bullying and 'methodological terrorism' (typically directed at parties conducting replications of original research; Gelman, 2016; M. Meyer & Chabris, 2014), the coining of a new term 'research parasites', which refers to researchers who use or build on existing data in their publications (Lowe, 2016), and the branding of 'replication Nazis' (Chambers, 2014). While some such responses are natural and human (as Nobel laureate Kahneman points out, a lot of passion and ego are involved in science: 2014), they can and have caused a lot of damage to the relations within the field of psychology. So profoundly damaging that replications should be *prohibited* unless the replicators consult with the original authors before conducting the replication, argued Kahneman.

One focus especially central to the current crisis is replications and research reproducibility. While the crisis of half a century ago discussed frequent replication failure as *one* concern in a multitude of problems, reproducibility is at the heart of the most recent crisis of confidence. To be sure, the term 'replication crisis' is used interchangeably with 'crisis of confidence' (Gelman, 2016). Ebersole, Axt and Nosek (2016) highlight the current importance of reproducibility

[4]The discourse surrounding the crisis of confidence certainly involves issues of identity, however they do not seem to be so central as in the 1970s crisis.

in the perceptions of those in the field. They asked participants about how they evaluated scientists based on attributes described in profiles. These attributes related to research and reproducibility, and the nature of the scientists' findings, and included loaded descriptive phrases such as 'exciting but uncertain', and 'very reproducible'. While the participants judged the certain but boring scientists more favorably than their exciting but uncertain counterparts overall, they perceived the 'reproducible' scientists as "smarter, more ethical", "better" and "more likely to get and keep a job" (p. 2).

2.2 Impacts of the Current Crisis

The crisis of confidence has no doubt shaken many fields, affecting their practices and communication on a fundamental level. It has had an indelible role in precipitating change on different levels and through different channels for these disciplines. It has influenced peoples' ways of thinking, and has led to the still continuing reform of entire fields' ways of 'doing'. Researchers have shifted their research operations to devote time and energy to overhauling fundamental aspects of practice, including prioritizing reproducibility and transparency in general, structuring, reporting and execution of research studies, and overhauling publication and incentive systems. Long-standing debates have reignited over topics in philosophy, theory, statistics and methodology.

Threats to Scientific Credibility

Earlier, I alluded to cases of fraud in psychological science fields, and areas of research in the literature of these disciplines that have suffered hits to their credibility in recent years. I now describe some of the literature on scientific misconduct to explore the issue in more detail, as it forms the basis of much of how researchers think about scientific studies and practices in this 'post-crisis' time, and motivates many incentives and practices that have been generated, as well as being fuel for some of the debates about tone I have referred to. Moreover, discussions about misconduct motivate good-faith researchers to practice research in ways they might not have done *pre*-crisis.

People may be more cautious about letting manuscripts be submitted without seeing the data first, for instance. They may be more likely to think more carefully about their own protocols when it comes to writing and sharing code and data, and it may prompt them to consider using preregistration or registered report formats. First, I will talk about a well-known fraud case that shocked and disillusioned psychology at the beginning of the crisis, and drove much of the early crisis narrative.[5] I will then discuss some focal points in the literature in

[5]I emphasize that this case is one of many which involve alleged and proven fraud. Please also note that these cases are at the extreme end of what is best conceptualized as a spectrum of QRPs,

psychology that have been at the center of problems with credibility. [6]

Scientific Misconduct

Questionable Research Practice As I have already mentioned, the literature indicates that QRPs appear to be widespread in the scientific community. According to some, QRPs tend to arise as the result of ambiguity as to which exact methodological and statistical procedures to follow, in conjunction with a desire to report interesting and eye-catching findings (Simmons, Nelson, & Simonsohn, 2011). Simmons and colleagues state that QRPs are usually committed in good faith. Regardless of intentions, QRPs disrupt the 'hypothetico-deductive research method' in several different ways, according to Simmons, Nelson and Simonsohn, and are still a form of scientific misconduct (Simmons et al., 2011).

Hypothesizing after the Results are Known (HARKing; Kerr, 1998) is a common QRP. Though there are different forms of HARKing (discussed in detail by Rubin: 2017), the most well-known form involves presenting an hypothesis which has been developed based on the data (i.e., post hoc), as if it were predicted in advance (i.e., a priori), by a theory development process. HARKing is problematic as it heavily influences the validity of the research it involves. Kerr, and much later, Rubin elaborated on the costs HARKing can have to science, which include translating type-1 errors (i.e., false positives, or results which are significant due to a fluke, rather than being the result of having measured a true effect) into theory, and undermining the value of the conclusions drawn from affected research (1998; 2020).

The File Drawer problem (Rosenthal, 1979) refers to another common QRP, in which researchers fail to publish their null findings. This bias leads to the construction of a literature that does not faithfully represent all the research findings which have been produced. It also leads to the waste of scientific resources like time and labor, as researchers build on a flawed and incomplete foundation of literature.

P-hacking, cherry picking and salami slicing are further examples of QRPs. On their own they are not especially egregious, but can be detrimental to the fidelity of the scientific record when committed on a large scale (Field et al., 2019). With p-hacking, the data are 'massaged' (i.e., the analyses are run and re-run) until a significant p-value appears. In the case of cherry-picking, one can selectively report findings depending on what 'fits' the existing theoretical expectations. Lastly, salami-slicing refers to the process by which a researcher 'cuts up' a single large set of studies (on a common data set, produced from a single study sample, with the same hypotheses and study aims) into single studies, and publishes them, representing them as being independent studies. These QRPs undermine the reliability of any research finding influenced by them,

which were likely the norm in psychology, at least until recently (John et al., 2012).

[6] And as with the cases of misconduct, these examples form part of a long list.

because they produce results that misrepresent the true nature of a study (and its corresponding hypotheses, materials, findings), typically making them more compelling than they are in reality. Salami-slicing is unethical in an additional way: The researcher can artificially boost their research output by publishing many small studies rather than one large study.

Fraud I have just described different ways in which scientists can interfere with the research process to engineer their results to be more compelling (or less disappointing), and outlined why QRPs undermine research quality, especially when a large proportion of the scientific community engages in them. As I discussed, research indicates that ambiguity, naivety and the publish or perish culture in academia tend to lead to their commission. They are different to outright fraud, where the intent is on intentional falsification of the scientific record. It is unclear how common scientific fraud is, as it is difficult to empirically prove (though Stricker and Günther report that researchers estimate that between 9.3% and 18.7% of other researchers fabricate data; 2019).

Perhaps the most infamous case of proven scientific fraud in psychological science, and the key case which set the ball rolling on the crisis of confidence in psychology (Stroebe et al., 2012) is that of Dutch social psychologist, Diederik Stapel. His autobiography *Ontsporing* ('Derailment' in Dutch, published in 2012 after the fraud case had unfolded), details Stapel's history and the case, and provides an interesting look into the mind of one of psychology's biggest fraudsters. Stapel was considered by many to be an academic superstar. In the fifteen years before he was investigated for academic fraud, he had published more than 100 journal articles, supervised more than 20 PhD students, won several prestigious awards and grants and became dean of the social and behavioral sciences at Tilburg University (TiU).

Seemingly overnight, Stapel went from academic giant to a pariah. In September 2011, TiU suspended Stapel pending an investigation into his research activity upon receiving reports alleging widespread and serious academic misconduct. Three committees were convened to take part in the inquiry, and their investigations concluded with a statement that at least 30 of Stapel's studies were 'fake' – he had tampered with or outright fabricated entire data sets, including some data sets his PhD students had used for dissertation research. Additionally, one of the committees found 'strong indications of fraud' in two chapters of Stapel's doctoral dissertation (Levelt Committee, Noort Committee, & Drenth Committee, 2012). The final investigation report states that failures of the scientific community contributed to the fraud continuing for so long undetected. It refers to "... a more general failure of scientific criticism in the peer community and a research culture that was excessively oriented to uncritical confirmation of one's own ideas and to finding appealing but theoretically superficial ad hoc results" (no page number; Enserink, 2012). As of 2019, a total of 58 articles coauthored by Stapel had been

retracted.

This case of serious academic fraud resulted in career-ending fallout for Stapel, and other serious grievances for his many coauthors and students. An entire cohort of researchers have been left seriously traumatized, the effects still felt years later. These people have had to repair themselves and their reputations, and attempt to move on from their associations with Stapel as best they can.

Other effects of the Stapel case can be seen in practice. First, TiU now requires PhD supervision by at least 2 supervisors. This measure was put into place to ensure accountability and curtail future misconduct by supervising academics. PhD dissertation data must now be collected by and analysed by the student, and the doctoral thesis examination board is required to establish the provenance and quality of the data that were collected for the thesis. It is now also required that Master students at TiU learn to conduct replications, and be trained to understand and apply principles of scientific ethics, integrity, and master methods and statistics that enable the responsible practice of research.

2.3 Replication and Reproducibility

Fraud and misconduct discussions are dominant in the discourse surrounding the question of what led to the crisis of confidence, though other issues have been thoroughly explored, including replication and the reproducibility problems of some of psychology's foundational phenomena. Reproducibility problems in psychology were a major focus, due, at least in part, to the Open Science Collaboration project (Open Science Collaboration, 2015), which revealed a seemingly unexpected high rate of replication failure.

This now-iconic article, 'Reproducibility Project: Psychology' was initiated in 2011, led by the Center for Open Science, which is still spearheaded by Brian Nosek. Two-hundred and seventy contributing authors attempted to replicate 100 original studies from three of psychology's flagship journals. Their aim was to assess the reproducibility of psychological research on a larger scale than had been ever attempted. While 97 percent of the replication target studies reported significant results, only 35 percent were replicated in the attempts of the OSC. The OSC expected replication success for a minimum of 89 percent, if all original studies reflected true effects. Some of the replication attempts revealed effects going the opposite direction from what was reported in the original studies. Type-I error and publication bias (the inclination for journals to only publish significant and novel findings) are two possible causes of the OSC's troubling findings.

Replication Background

Replication, the repetition of empirical studies, was initially discussed in literature from the 17th century and is now promoted as the cornerstone of science

(Bauernfeind, 1968; Shapin & Schaffer, 2011); a "basic tenet of scientific advancement" (Smith, 1975). Dunlap highlighted science's need for replications:

> The proof established by the test must have a specific form, namely, repeatability. The issue of the experiment must be a statement of the hypothesis, the conditions of test, and the results, in such form that another experimenter, from the description alone, may be able to repeat the experiment. Nothing is accepted as proof, in psychology or in any other science, which does not conform to this requirement. (p. 503; 1925)

Replication is said to be at the heart of scientific inquiry (Moonesinghe et al., 2007; Simons, 2014). According to the American Psychological Association (2010), replication is the "the essence of the scientific method" (p. 4), one of the "most obvious ingredients of science" (p. 91; Schmidt, 2009), and an essential activity for scientific research (Cesario, 2014). Robert Boyle made this argument: Repetition of the same experiment will result in certainty of fact (Shapin & Schaffer, 2011). Karl Popper's Logic of Scientific Discovery lays out the importance of repeatability similarly:

> Only when certain events recur in accordance with rules or regularities, as is the case with repeatable experiments, can our observations be tested – in principle – by anyone. We do not take even our own observations quite seriously, or accept them as scientific observations, until we have repeated and tested them. Only by such repetitions can we convince ourselves that we are not dealing with a mere isolated 'coincidence' but with events which, on account of their regularity and reproducibility, are in principle inter-subjectively testable. (p. 23; 2005)

All of this is to point out that scientific information moves from being a single (potentially) chance observation to having evidential value *only* when it can be demonstrated again.

This topic is enjoying renewed popularity as of the last decade, due, in part, to the findings of the Open Science Collaboration (2015), which I mentioned before, and the implications of its findings for science. But, low reproducibility seems to be widespread in science, not only affecting psychological fields. For instance, concerns over replication rates in biomedical and pre-clinical research are evident (Ioannidis et al., 2009). A 2011 drug-development program reported a replication rate of 20-25% (Prinz, Schlange, & Asadullah, 2011), while Amgen Incorporated announced an 11% pre-clinical research reproducibility rate (Begley, 2012). Although quantifying reproducibility in science is possible due to the evidence provided by such reports, improving the validity of our research rests in establishing the underlying causes of low reproducibility.

Numerous contributions have been made to improve the replicability of research findings in the form of changes to institutional policies (Burgelman et al., 2019), mass replication attempts (Klein et al., 2014; Nosek et al., 2015), articles aiming to raise awareness of the issue and explore its causes (Pashler & Harris, 2012; Yong, 2012a) and funding (e.g., the Dutch Research Council – NWO – ran a program between 2016 and 2019 during which time it funded three million euro worth of replication studies). The topic of replication and reproducibility has reached many scientific domains, and continues to spark debate and controversy within the scientific community, as well as causing division within and between disciplines and other sub-communities (Pashler & Wagenmakers, 2012; Spellman et al., 2018). It is of particular interest to reform and open science advocates, either because they are conducting replication studies themselves, or because they disagree with the emphasis that open science activists put on them.

There are a great many definitions (Clemens describes a 'mess' of definitions; 2017), descriptions and discussions of the concept of replication in the literature. Although replication as a concept may appear to be quite straightforward – repeat an existing experiment and see if the same finding is obtained again – it is far from easy to define. This is partly because the topic of replication has been circulating throughout the scientific literature for decades, and because it comes in many forms, depending on the motivation behind it and the way in which it is conducted, and its results can be interpreted in different ways.

Schmidt states that replication, in its most basic form, is a tool of scientific methodology based on repetition, that aims to establish "fact, truth or piece of knowledge" (p. 5; 2009). Shapin and Schaffer's definition is very similar: "Replication is the set of technologies which transforms what counts as belief into what counts as knowledge" (p. 225; 2011). In their textbook on experimental and quasi-experimental research design, Campbell and Stanley wrote "... the experiments we do today, if successful, will need replication and cross-validation at other times under other conditions before they can become an established part of science, before they can be theoretically interpreted with confidence" (p. 3; 1963). Schmidt (2009) concludes that most of the definitions in the literature describe the repetition of a study procedure, and suggests that it is important to differentiate between different notions of replication in terms of their scope. Since Schmidt's article, replication terminology has been debated at length and the (sometimes heated) discussion has zoomed in to scrutinize the finer points of replication definitions. An example of this is the direct versus conceptual replication debate, in which the typology of replication features prominently as I will describe in the next section.

The Dichotomy of Replication

For practical purposes, the different replication types are most often grouped together in discussions in the literature into 'direct' and 'conceptual' varieties. In

early articles on the topic (in the 2012 special issue on replicability and the crisis of confidence in *Perspectives on Psychological Science*, for instance) only these two forms were discussed.[7] In very basic terms, direct and conceptual replication can be distinguished from one another by aim or functionality. For instance, Gross (1997) describes two specific roles. The first is that replications check the probability of error in hypothesis testing; that is, the likelihood of the commission of Type-I (so-called 'false positive') and Type-II (so-called 'false negative') errors (in direct replications). The second role holds that replications control for confounding variables that may have influenced the original findings (in conceptual replications). As discussed by Schmidt (2009), the aim of the replicating author may be to copy as precisely as possible the experimental protocol of an original experiment, and see whether the resulting data reflects that of the original experiment. It may be to control for a chance result (Type-I error), to control for artifacts, to control for fraudulent results, or questionable research practice. Direct replication is known by other names too, that reflect this aim, such as 'pure', 'literal' or 'exact'.

Zwaan and colleagues (2018) state that conceptual replication doesn't serve the same purposes as direct replication: conceptual replication has the broader goal of establishing whether a certain way of testing a theoretical idea can be quantitatively reproduced. Conceptual replication has the role of evaluating a theory's robustness to varied study designs, variable operationalizations and experimental samples (p. 10). As with direct replication, conceptual replications also go by other names reflecting the aim, such as 'varied', 'external' or 'partial'.

Certainly, being able to classify replications is useful in many ways. It allows researchers to refine the language they use to accurately communicate and describe their replication results, and allows them to set realistic expectations as to what the replication's results should reveal. Regardless of how a replication study is categorized, however, two points should be noted. First is the issue of choosing a study to replicate. Do researchers just replicate random experiments that catch the eye, or that are caught up in the most recent fraud case, or can they approach replication more systematically? Another issue revolves around how replicators are supposed to quantify their results. What does a failure mean, and how can and should success be measured? Both of these questions have been explored in depth by open science and reform advocates, and I briefly touch upon them below.

What Studies Should We Replicate?

There are a few points to keep in mind when considering the issue of what study to choose when you are in the position to conduct a replication study. First, there

[7] Though in the years since this special issue, nuance has emerged, as have other explorations of how to categorize replications, such as that of Peterson and Panofsky, who talk about diagnostic versus integrative 'motives' for replication (2021).

is the problem of a massive body of literature. In psychology, as in many fields, the body of literature is very large, meaning that the pool of potential replication candidates is too. Even when one's interests in a replication target are paradigm-specific, most literature bodies focusing on a specific psychological effect are sizeable. If one had unlimited resources, this wouldn't be so problematic – one could just conduct a group of replications to thoroughly explore the boundary conditions of a phenomenon of interest. Most entities, however, have finite resources in terms of time, funds and participants.[8]

The issue of resources is a second problem to consider. Often when one has the capacity to plan and execute a replication study, one has to keep limited resources in mind, and must be conservative with one's choice of replication target (Field et al., 2019). So far, studies have been selected for replication somewhat haphazardly, often with an emphasis on suspicions about the original studies' authors, or suspicions about the reliability of the findings. I published a study with colleagues from the University of Groningen which gave a recommendation to address these problems. We aimed to provide researchers with a methodological guide as to how to select a replication target (Field et al., 2019). We showed how a Bayesian reanalysis of potential replication targets, combined with a follow-up qualitative exploration of studies with the least compelling evidence (which the Bayesian reanalysis can indicate) can be a methodical and justifiable strategy to select a study to replicate. We showed how one can reduce a large pool of candidate replication targets to a select few, relatively quickly and efficiently, and avoid research waste.

When Replication Fails

Once a replication candidate has been chosen, and the study conducted and results analysed, the outcomes must be interpreted. In essence, we want to know whether or not the replication was successful. Replication *success* means that a replication study showed results adequately comparable to the original study. Unsuccessful replication means that a replication study does *not* yield a result that is adequately similar to the original it is based on.[9] Unfortunately, poor reproducibility is typically considered to reflect a negative scientific event, but this bad press fails to take into account the nuances of 'failure'. The responses to the 2015 Open Science Collaboration which appeared in the media, and in other scientific journals illustrate this. They tended to use words to incite dismay and concern (and presumably action) among lay-people and scientists alike. An online *Nature*

[8]Individual researchers are not the only ones who conduct replications – increasingly, replication is conducted by groups of people ranging from a small group of collaborators to a large international group of researchers (a 'distributed laboratory network') such as those who conduct the replication studies for the Psychological Science Accelerator (`https://psysciacc.org`).

[9]Again, the judgment of 'adequate' is rather much at the discretion of she who conducts the replication.

news story, for instance, published a piece that suggested most scientific results shouldn't be trusted (Baker, 2016). Gigerenzer (2018) asks rhetorically whether science is "on its last legs" (p. 199). Engber's article in *Slate* is titled "Everything is Crumbling" in reference to failures to replication (2017). "Is psychology about to come undone?" asks Bartlett in his incendiary article for the Chronicle of Higher Education (2018).

The scientific and lay-media are not the only concerned parties. Often, when failures to replicate occur, researchers conclude that the original study was the result of type-I error, or that the effect it reported is too brittle to be meaningful (see Cesario, 2014). Cesario is skeptical of failures to replicate, warning that once the original methodology has been deviated from, the result of the failure is ambiguous. In some instances, this suggestion is used to imply that the replication does not have any information value (see e.g., Stroebe & Strack, 2014). The opinion of Clemens seems more nuanced. He argues that a replication failure can provide different information, depending on the context. It can signal "a legitimate disagreement over the best methods" or "incompetence and fraud" (p. 326; 2017). Others imply that failures to replicate may indicate that a consistent effect across a wide sample or under different circumstances is unlikely to exist (Diener & Biswas-Diener, 2018). Maxwell, Lau and Howard (2015) offer logical explanations for poor reproducibility, stating that failures to replicate are not failures – they are the result of underpowered replications, and a failure to understand the importance of series of replications (as opposed to single, isolated replication attempts).

Priming research and the case of power posing

One notable byproduct of replication increasing in popularity as a topic is that many well-known psychological phenomena have been discredited. For priming research in particular, issues of replicability have been well-documented. Naturally, every area of research has some degree of contradictory literature, however an especially high number of article retractions and failed replications in behavioural priming research have sparked serious doubts about whether many 'carry-over' effects are robust enough to contribute to our understanding of unconscious influences on behaviour. The state of affairs in priming research is arguably what originally prompted people to identify a 'crisis of confidence' in research findings (Pashler & Wagenmakers, 2012), as I mentioned earlier. The effect of power posing is a very salient example of priming effects which have failed to replicate (though several other well-known effects have also shared the spotlight in recent years: Jarrett, 2016).

These problems of replicability in priming research have captured the attention of the broader field of psychology, with some prominent researchers now calling for a collective effort to further replicate these priming effects, and reestablish the field's integrity. As I discussed earlier, Kahneman issued a direct statement to

priming researchers in the form of an open letter, imploring them to decisively address the problem of inconsistency and non-replicability marking the literature on priming effects (Yong, 2012a): "I believe that you should collectively do something about this mess. To deal effectively with the doubts you should acknowledge their existence and confront them straight on, because a posture of defiant denial is self-defeating." (no page numbers given; Kahneman, 2012) An attempt to help return integrity and merit to the field, this confronting but constructive letter is a reflection of the issue's magnitude, reach and impact.

Strong evidence exists that some kinds of priming are valid and reproducible (for a review, see e.g., Pickering & Ferreira, 2008), despite concerns. Initial work on priming in the 1970s involves the activation of concepts in memory; facilitating and speeding up the processing of subsequently presented related stimuli. In one experiment demonstrating this, Meyer and Schvaneveldt (1971) measured the time taken by participants to respond to 48 pairs of words (associated words, non-associated words and non-words). One of the first of its kind to measure unconscious memory, this study presented evidence for semantic priming, demonstrating that participants respond to associated concepts in memory more quickly than they do to unrelated concepts. A research focus that emerged some ten years after initial work on semantic priming refers to a method of priming in which the procedure in itself acts as the prime. That is, in typical procedural priming experiments, participants perform a certain activity or task which itself primes for a desired processing style. The prime or carry-over effect is then measured through a change in performance or processing in a second, unrelated task.

Priming is a phenomenon with a broad definition, and many different kinds of priming paradigms are subsumed under its umbrella. Power posing, an effect in the embodied cognition area of research, is more narrowly defined. The effect and the authors of the studies in which it originated (Carney, Cuddy and Yap) have been the focus of heavy critique in both the lay media and in the scientific literature after multiple failures to replicate their findings put the phenomenon in the spotlight. The idea of power posing is simple, and involves what one of the study's authors calls a 'mind-body nudge', where a primitive mind-body link can be exploited to use physical positions to evoke a mindset of confidence in oneself (Cuddy describes the 'mind-body nudge' in her 2015 book, Presence).

Carney, Cuddy and Yap's *Psychological Science* study (2010) has been cited 1400 times to date, and its findings (and dramatic and unreasonable extrapolations of them) appear in numerous popular media articles. It demonstrated that participants who assumed certain positions with their bodies – hands on hips, chest out, feet wide apart, for instance – reported feeling stronger and more confident than before assuming the positions. Along with the subjective measures, objective measures were taken, which revealed physiological changes: levels of the stress hormone, cortisol, decreased, while testosterone (responsible for 'male-typical' behaviors such as aggression) increased. Finally, risk-taking behaviors were tested. Participants who had posed expansively took greater financial risks

than those in the other condition. The authors of the 2010 study concluded that "... a simple 2-min power-pose manipulation was enough to significantly alter the physiological, mental, and feeling states of [...] participants. The implications of these results for everyday life are substantial." (p. 1366; Carney et al., 2010) The popular media would seem to agree. Numerous articles and posts celebrated the effect and implored readers to try the wonder-woman pose to boost their confidence and skills in contexts like salary negotiation and to reduce stress (for just a sample of these, see: Perlis, 2012; Clear, 2013; Cuddy, 2012a). In 2012 Oprah magazine picked up the story, and Cuddy was invited to deliver what has become an influential TED talk. It is the second most popular talk in the series, to date, with over 64 million views since it was released in June of 2012 (Cuddy, 2012b).

The effect of power posing is seemingly not robust, despite its high profile. The results of at least 11 studies show that the effect is unreliable and malleable at best. Perhaps the most high-profile failure to replicate appeared in *Psychological Science* in 2015. The replication conducted by Ranehill and colleagues (2015), failed to find complete support for the 2010 effect despite testing with a high-powered design, featuring an *N* almost five times larger than the original. They reported that although they did not replicate the physiological or risk-taking behavioral findings, participants did report *feeling* more powerful in the expansive posture condition. Although Ranehill and her coauthors did slightly deviate from the methodology, the deviations were intentional, and aimed to give the effect the best 'shot' at working. First, they extended the time during which participants were required to hold the poses they had adopted from one minute to three. Another deviation involved the extent to which the experimenter interfered with the participants. While in the Carney et al. study, the participants were posed by the experimenter, a computer instructed Ranehill's participants (who were also monitored on camera to ensure that the positions were correct and being maintained for the required period).

Reproducibility and Predictors

'Reproducibility' as a concept is general in its scope, but at its core refers to the outcome of a replication study. It is a way to explain how well an effect can be replicated. One might say, for instance, that an effect is not reproducible when replications continually fail to validate it. Similarly, one might suggest to conduct a replication study to test the reproducibility of a claim or finding. A broad meaning for reproducibility comes from de Ruiter. In the abstract on his 2018 article, he writes: "A scientific claim is a generalization based on a reported statistically significant effect. The reproducibility of that claim is its scientific meaning." (quotation from abstract; 2018) The question that springs from this statement, which allows us to consider how to operationalize it, is: "What *is* that meaning?" The problem with using this single word: 'reproducibility', is that the outcome of a replication study and what it means is heavily dependent on *what*

is being compared between an original study and its replication. Indeed, what is the *basis* of the assessment we deliver when determining whether something is reproducible or not?

The nebulous and general definition of 'reproducible' does not suffice to describe the distance between an original study and a replication of it, nor to clearly communicate the comparison of interest (that is, are we comparing the original and replication effect in direction only, or are we comparing both direction and magnitude of the effect?). Compounding this issue is the fact that the use of the word reproducibility is used very loosely in research reports. One only has to read a handful of the literature on the topic to see that terms like 'replication' and 'reproducibility' are used practically interchangeably (Schmidt, 2009).

Another issue with defining reproducibility is due to the range of tools that are currently used to quantify the success or failure of a replication attempt. After all, when do we actually know whether a replication *p*-value is significantly different from the *p*-value of the original study? How similar in magnitude does an effect need to be in an original-replication pair? How different do these values need to be before they are meaningfully different and we declare 'poor reproducibility' of an original study? Determining how to quantify a replication outcome is difficult, and is another point of debate between some groups involved in reform and open science.

Although replication success can depend on many discipline-dependent variables (Open Science Collaboration, 2015), the pressure on academics to publish novel findings, and many of them, is relevant to all fields of scientific research (Steele, Butler, & Kingsley, 2006). These pressures can engender a culture in academia in which questionable research practices (QRP) and fraud may flourish (Barbour, 2015). Further undermined by publication bias toward novel findings, the filing away of null results and poor statistical analyses, an unstable body of literature has emerged (Steckler, 2015; Rosenthal, 1979), characterized by low reproducibility rates. Other predictors of replication failure are domain-specific, such as the lack of specificity in published research reports. Studies suggest that the large proportion of articles lacking sufficient detail regarding methodology and materials has limited the reproducibility of findings. This is especially applicable to biomedical fields (Carp, 2013; Vasilevsky et al., 2013; Ioannidis et al., 2009), though in a previous study which reanalyzed numerous findings reported in the journal *Psychological Science*, I experienced problems with poor statistical and methodological reporting practices, and dwell on related concerns (Field et al., 2019).

Ultimately, though many would argue that the renewed interest in replication and reproducibility brought on by the crisis of confidence is positive, this topic has also brought much disruption and anguish to many scientific disciplines. For instance, news of failures to replicate garner strong responses from the community. Many authors of original experimental effects feel as though they are being witch-hunted or bullied (Cuddy and Schnall are two examples – see, respectively:

Dominus, 2017; Rhodes, 2015), and replication has been at the core of many 'tone debates' in recent years. Those who have authored papers that have been the subject of failed replications sometimes lay blame at the feet of the replicators. Arguments as to why they are the cause of the failures range in their reasonableness and reasoning (see Yong's article reporting on one example: 2012b), but they often point to the carelessness or lack of expertise of the replicating authors as the problem.

2.4 The Emergence of a Movement and a Community

Early discussions on the topic of replication have also been a catalyst for broader debates in reform circles. For example, the questions about replication I have considered in this section – *What effects do and do not replicate? In what fields do replication failures occur most frequently? How do we conduct the best replication studies? Why and how do we select studies to replicate?* – have fed into questions about tone in scientific critique – *How should researchers engage with one another about issues like replication failures?* – and assessment of the reliability and validity of qualitative research – *Can and should it be assessed in interpretivist (as opposed to positivist) traditions?*

At the end of his piece on science's poor state of being, Horton writes that "...nobody is ready to take the first step to clean up the system." (no page numbers given; 2015) Not everyone, it appears, would agree. Hilgard and Jamieson (2017) argue that science is self-correcting (and they're far from the only people to advance such a claim – take Merton, for example: 1973). Francis Collins, the director of the National Institutes of Health, is also optimistic, says: "science will find the truth. It may get it wrong the first time and maybe the second time, but ultimately it will find the truth." (Achenbach, 2015) Nosek and Munafò are also on the record with positive attitudes toward the progress of science in the wake of the crisis (see Baker, 2016; Yong, 2015).

The problem with the "science is self-correcting" narratives, is that they suggest that the process is passive. That science's wheels will continue to turn, and the creases will just iron themselves out. The truth is that if there is hope for a reliable and valid social science in the future, it will not be born of a passive process. Science does not just exist; it is constructed by humans who have developed philosophies and tools to do the job of producing research and communicating it.

3

Open Science and Reform Practice

3.1 Open Science and Reform Practices

> Like shadows, openness is the result of nuanced encounters between light and darkness, whose visible results reflect both the obstructions and specificities of each setting. Again like shadows, openness is inherently positional and relational and is subject to dramatic qualitative shifts depending on the characteristics of the locations involved or the personal relationships between the individuals and groups involved. Whether this is explicitly acknowledged or not, openness entails judgments about what counts as a valuable research output or practice, such that particular enactments of openness lead to the endorsement of some things as valuable, and others as not. It is not just a question of what should be made open but also about how particular instantiations of openness value some forms of care and labor over others.
>
> Levin and Leonelli (p. 295; 2017)

Definitions, Schools and Philosophy

As the word 'crisis' came into use to describe the turmoil of the field, the "contours of a community of critics began to emerge" (Derksen, 2019; p. 321). Derksen calls the community's reaction to the crisis the 'Open Science initiative'. This initiative describes the collective strategies scientists have contributed toward solving the field's problems. These strategies generally emphasize transparency in scientific practice (and reward researchers for transparency, says Derksen), which should result in "tightening standards and limiting unwanted flexibility" (p. 324). So far, a considerable effort by individual researchers, academic institutes and lay-community funding bodies alike has started to improve our collective understanding of the mechanisms behind the crisis, and of how we might modify our

methods going forward to improve the reliability and certainty of our findings (Pashler & Harris, 2012; Chambers, Feredoes, Muthukumaraswamy, & Etchells, 2014). The recognition of problems in science and a drive to resolve them has led to a movement (Robson et al., 2021; Vazire, 2018). These people are united by a common interest in open science, and focus in particular on the practices which underpin that joint enterprise.

Open science is not a category of science in and of itself; it is not different from science in the traditional sense, say open science advocates. They argue that it is simply research which has been carried out with transparency and integrity and with an emphasis on collaboration. Robson and colleagues describe it as a "range of behaviours that aim to improve the transparency of scientific research." (p. 1; 2021) FOSTER, an e-learning platform dedicated to educating researchers about open science principles and practices defines it as a movement whose goal is to "make scientific research, data and dissemination accessible to all levels of an inquiring society." (`www.fosteropenscience.eu/about`) As this definition says, open science involves ensuring that components of the research itself such as materials, computer code and data, are made available to others. It involves researchers being transparent, reflexive, and honest about the decision processes involved in their research in general, and, crucially, when things in the research process do not go as expected. The open science and reform movement describe a focus on conducting 'good faith' science, that is, they strive to execute science with integrity and honesty of purpose. It is not science free of error or uncertainty – "mistakes can and will happen" assure Allen and Mehler (2019). Rather, they clarify, it is science that is free of assigning reprimand or blame, and which fosters a culture around transparency in the scientific process, such that error and uncertainty are clearly signposted and discussed.[1]

Fecher and Friesike (2014) recommend against using a precise definition for open science, instead proposing five schools of thought to shape the discourse surrounding open science. These schools, or 'threads', provide a framework for understanding the goals of open science as a movement:

- The infrastructure school. This thread is concerned with technological architecture, holding that research requires readily available tools, services and platforms to make dissemination and collaboration efficient.

[1]This mindset is what drove the founders of *Journal of Trial and Error* to establish their journal. They explain the emphasis on discussing error as a part of the scientific process in their manifesto (www.jtrialerror.com/the-manifesto-for-trial-and-error-in-science/): "We want to publish (1) methodological errors which have productive conclusions for the scientific community at large, and (2) conceptual errors in the form of negative results. In addition, our initiative aims to create a platform to openly talk about failure. That does not mean that we want to publish sloppy science. Rather, we believe that in talking about errors, scientists can learn about the do's and don'ts of their methods and concepts." (see their editorial piece for a more detailed explanation of the goals of the journal: Devine et al., 2020)

- The Public school is concerned with accessibility of knowledge creation. It holds that truly impactful research requires societal engagement (citizen science), and the provision of understandable, accessible research outputs.
- The measurement school seeks alternative scholarly metrics, holding that traditional metrics for evaluating research impact are problematic and insufficient for their purpose.
- The democratic school is concerned with access to knowledge, and holds that information gleaned through research effort should be available to all.
- The pragmatic school holds that the (co)production of knowledge is more efficient when it involves collaboration, and is strengthened by critique. The goal of this school is to connect researchers and to make the research process transparent.

Fecher and Friesike's approach to categorizing practices and different priorities in the open and reform space provides an interesting alternative way of viewing the group of people advocating and practicing. Until now, I have largely presented this group and their focuses as a single community of people who take on a large and diverse roster of activities in order to pursue their joint enterprise. However, as I will advance in the analysis chapters (**7 and 8**), this is too simplistic a view of the space.

In this chapter, I describe the major kinds of practices that have been generated by the group so far, and their corresponding goals and potential benefits. Through an exploration of a selection of the literature, I lay out the guiding principle behind the group's joint goal – scientific improvement: Opening up science can improve the validity and reliability of research findings, can boost their reproducibility, and has the potential to reduce academic misconduct (Vazire, 2018; Banks et al., 2019).

Practice

The Four Pillars of Open Science Practice

The principles of open science translate into a wide variety of initiatives or 'behaviors' and a great deal of information on best practices and guidelines. Open science principles are applicable to every aspect of the research process, from the start of the process, to the end. Of course, the research pipeline varies with different academic traditions, but many share core components which different practices target. For instance, for psychology and other, similarly positivist disciplines an ideal 'open' research process might happen roughly as follows: Starting with theory exploration, hypothesis generation, preparation of a preregistration document or registered report proposal, through to data collection and storage, sharing of the code and materials used in the experiment and analyses of the

data, open peer review of the written research report, and free access to the published article (Banks et al., 2019; Munafò et al., 2017). Often, the different kinds of practices are categorized, or described in terms of 'pillars'. For instance, Utrecht University, the open, interdisciplinary journal *4open* and FOSTER reference a four-pillar system of open science practices, while the University College London, the European Commission and the University of Groningen refer to eight pillars. I will explore the simpler four-pillar system, and then describe other practices which do not fall under any pillar categorizations.

Pillar One: Open Data Open data is an important pillar of the open science movement, according to FOSTER. There are different levels of data openness. The Open Data Institute (`https://theodi.org/about-the-odi/our-vision-and-manifesto/our-mission/`) describes open data as data that anyone can access, use and share. Shared data are described as widely accessible, but have conditions as to their use (e.g., relating to reuse and attribution). Closed data is self-explanatory – these data are often sensitive on some level, and may be impossible to share.

Data should be shared as soon as possible, even if they are not necessarily ready for wider use. It should also be noted that while the researchers who collected the data are major stakeholders in the use of the data, they do not own it. This implies that they must take other entities into account when they make choices about data use, storage and sharing. FOSTER and other organizations which focus on open data recommend drawing up a data management plan (DMP; a document which describes plans for how data are to be handled during data collection and after it has taken place) before data collection occurs. Such a document is 'living', and should evolve along with the project. The goal of creating a DMP is to manage data optimally both during the project's lifetime, and afterward. Ideally, a DMP ensures that waste and mismanagement of valuable data can be avoided.

The FAIR data guidelines (see: Mons et al., 2017) describe characteristics of data which help them be discovered online and reused by others. FAIR stands for Findable, Accessible, Interoperable and Reusable:

- *Findable*: Data should be located on repositories or websites such that they are easy to find should anyone look for them. They should be accompanied by detailed metadata and be identified using a persistent identifier online.
- *Accessible*: The data themselves and their metadata should be understandable by humans, machine readable, and stored in a reliable repository. They should be non-proprietary.
- *Interoperable*: Data should be able to be integrated with other data, and be able to be utilized by applications and systems for processing, storage and analysis purposes.

- *Reusable*: Data should be optimally reusable. This means that the data and metadata should be sufficiently and accurately described, such that they can be used to their maximum potential by others. Their provenance and context should be clear.

A useful tool has been provided by the Australian Research Data Commons which helps researchers assess their own data with the FAIR guidelines (`https://ardc.edu.au/resources/aboutdata/fair-data/fair-self-assessment-tool/`). The tool provides guidance for making sure that one's data meet FAIR guidelines where applicable. Other tools have been developed to help facilitate access to open data sources. For instance, rOpenSci (`https://ropensci.org`) is a technical infrastructure which facilitates the development of R packages through community driven learning, review and maintenance of contributed software in R.

The State of Open Data survey of 2019 (Fane et al., 2019) provides an interesting look at academics' views on open data. Although concerns about sharing data (such as their misuse by others, lack of clarity about licensing and copyright, and not receiving appropriate credit) are serious among the study's 8423 participants, 67% of them think that funders should withdraw funding or penalize researchers in some other way if they do not share their data as mandated by the funder. An even higher proportion of participants (69%) think that funders should mandate data sharing as part of requirements for granting funds to researchers. Interestingly, about half of the study's respondents who share their own data regularly have not heard of the FAIR principles.[2] Support for a widespread mandate for making research openly available has grown more than 25% since 2016, with 79% of respondents in 2019 voicing strong support for such a mandate. Despite this, figures from 2016-2019 vary with regard to what proportion of respondents reporting that they had curated data for sharing: 2019's figure of 65% is lower than figures for 2017 and 2018 (both reported 74%), and 2016 (67%).

Researchers were motivated to share data by a number of things, say Fane and colleagues (2019). Ranked highest were impact and visibility of their research (62%) and public benefit (60%). Other motivations involved the researcher themselves benefiting by, for instance: full citation (61%), co-authorship (42%), consideration in job reviews (45%), and financial reward (38%). Transparency and reuse motivated people to a lesser degree (48%), but only 2% of respondents reported that they had no intention of sharing their data.

Open data policies are not only important for academic research – scientific research for governmental bodies is heavily influenced by open data rules and most countries have laws which ensure that data conform to sufficient standards of openness. Open governmental data policies are thought to confer benefits to the public, according to Data Europa. These include enhanced performance of governmental research (i.e., sharing leads to efficiency of data use), and benefits

[2]This is interesting because it reflects the possibility that this group are not as cohesive or 'close' as one might expect. I will explore this issue in much depth in **Chapters 7** and **8**.

to local communities (which can benefit from easy access to valuable information).

Although numerous official open data initiatives have been established, academic researchers commonly use repositories such as the OSF, Dryad, and Figshare. It is worth noting that FOSTER suggests that researchers use domain or discipline-specific data repositories. This enhances ease of access and visibility within a given discipline. FOSTER shares some tips for best practice in data sharing, which include giving your data set a DOI, provide a recommendation for a citation for the data set, and share the data set's metadata and other important documents along with the data set itself.

Open data, according to proponents, has the potential to benefit society on a number of levels. In the context of the Coronavirus pandemic, open data has played a vital role in speeding up scientific discovery and innovation, leading to societal benefits like vaccination against the virus. The Organisation for Economic Co-operation and Development (OECD) website lists multiple benefits of open data. For instance, open data can reduce waste and redundancy in collecting, creating, transferring and re-using scientific material, and can increase productivity which is especially valuable where public spending is concerned. Huston, Edge and Bernier (2019) discuss how open data can promote greater citizen engagement, which can lead to more active participation in scientific experiments, initiatives and data collection. Open data increases opportunities for scientific collaboration, enriches society's capacity for research and analysis, and enables transparency and improves accountability of scientists. In the more specific context of public health, benefits include the faster detection of threats to the environment and health and the increase in capacity for evaluation performance indicators. Additionally, open data can inform interventions and policy decisions.

Fane and colleagues Fane et al. are optimistic about the future of OD. They write:

> We need to make open data matter, and one way to ensure that is the reuse of data that adequately credits the author, gatherers, creators, and even participants of those datasets. Another is to ensure that we build, support, and promote credit mechanisms that pay attention to the social mechanisms embedded in citation and authorship; not something that publishers, or funders, or researchers can do on their own. (p. 17)

Pillar Two: Open Code Sharing code and sharing data are related, because code tends to be written for a particular data set, or to be used on data. The benefits to sharing code are very similar to those of sharing data. Arfon Smith (in Mitchum, 2015) emphasizes that open code facilitates collaboration and learning from the work of others, as well as being able to build on what others have already done. He links code sharing to reproducibility, and inclusivity as well (be-

cause no 'pay to play' system means researchers with relatively little funding can participate). Smith talks about a culture of reuse, where "absorbing and reproducing methods can be freed from the shackles of traditional journal articles". He discusses the need for substantial cultural shifts in how academics work, including accreditation, trust amongst researchers, and software tool discoverability. Kathawalla, Silverstein and Syed (2021) add a few more pros of open code to the list – open code helps one modify one's existing analyses and visualisations, and signals one's commitment to transparency. They make code sharing more attractive to those who are not currently using software like R, noting that even sharing syntax used in programs like SPSS or the analysis files generated by JASP is a form of code sharing and useful in the same way for open science purposes.

The UK Reproducibility Network (UKRN), a national peer-led consortium, provides a useful primer on open code and software (`https://www.ukrn.org/primers/`). The primer, one of many aimed at educating on open science practices, outlines the benefits to sharing code. Though most benefits are already outlined above, the authors of the primer highlight that without source code, it is hard to test software and make sure that it does what it purports to. The primer also shares tips on how to open software and code, including a clear simple checklist. They suggest giving your code a descriptive name, licensing it, and pushing a versioned release of it.

The primer lists several resources people can use for learning about how to open their code, as well as places they can store or release it. For instance, The Carpentries is a global support community which provides training tools and support for people who want to develop their data and computational skills. The Turing Way is a handbook which was written to support students, their PIs, funders and editors in showing that reproducible data science is 'too easy not to do'. Stack Exchange and Stack Overflow are online resources which help support the data science community and others interested in writing and using code. Systems also exist such as Github, Gitlab, OSF, Binder and Sourceforge which can be used to store and release code and software.

Easterbrook (2009) highlights why open code is necessary with this pithy statement: "Poor code is endemic" (no page number). Technical debt, says Easterbrook, is behavior whereby one defers issues like code readability and maintainability until 'later'. These chores usually don't get finished, however, resulting in a debt which must be repaid by someone in the future – either oneself or another – in their effort to rerun or modify the code. Journals which require open code help remedy this issue, because authors are strongly motivated to clean up their code each time they produce an article.

Although Easterbrook lists a number of barriers to code sharing, Walters' 2020 article, 'Code Sharing in the Open Science Era' declares that sharing source code alongside papers has become the norm in many disciplines, and that science has reached the point where meaningful technical barriers to reproducibility no longer exist. Indeed, many journals require that authors share their code at the

time of publication, or even at the time of submission. For instance, *Science*, *Frontiers*, *PLOS* and *Royal Society Open Science* all require open code; *Nature* does not require it, but strongly recommends it.

Walters (2020) sees resistance to code sharing as a societal issue, rather than a feasibility or scientific issue. This is interesting, because Walters highlights potential arguments against code sharing which reflect certain elements of academic culture. Arguments include code going out of date, that code can be made available on request (so why bother front-loading it at the time of article submission or publication?), fears that one might lose the competitive edge, especially if grant funding is on the line, and finally, fears that people will be critical of code which is messy, contains errors, or is otherwise unprofessional. Walters counters these issues in turn in his article, but the main thrust of his piece is that open code is powerful for sharing the advancements already made by others, learning from them, and building upon their work. He recommends three key changes to help facilitate wider code sharing in the academic community. First, he recommends the institution of journal policy to endorse articles when their authors share their code (which is interesting in a sociological sense). He recommends that authors be required to include explicit code availability statements in their articles. Finally, he recommends the development of clear policy for reproducibility and inclusion of source code with articles.

Pillar Three: Open Peer Review Many open science initiatives relate to the peer review process. Historically, peer review has been a closed system, however a review by Ross-Hellauer (2017) reveals an almost exponential growth in the interest in open peer review (OPR) since it first appeared in the literature as a concept in the 1980's. OPR is just like regular peer review, except that the process of peer review (including the reviews themselves, and often the authors of the reviews) are made publicly available, alongside the paper they refer to. Ross-Hellauer finds that OPR began to gain momentum in the "early-mid 2000s onward" (p. 6), and demonstrates that the social sciences are leading the adoption of OPR, followed by medicine and health sciences, and the natural sciences. Apparently, big funding entities also see the value in backing OPR. For example, the platform F1000Research gained substantial funding not long after its launch from both the Wellcome Foundation and the Bill & Melinda Gates Foundation (Butler, 2017). Despite blossoming interest in OPR, no standard or agreed upon definition exists for it, and no schemes of features or characteristics are to be found. Part of this, reasons Ross-Hellauer, is the fact that OPR is an evolving and complex concept with many facets, and lacks a coherent vision (in comparison with open access, for instance).

The literature documents debate about the value of OPR, as Ross-Hellauer's review article shows. Older articles such as Ware (2008), Godlee and colleagues (1998) and van Rooyen et al. (1999) report perspectives that OPR has limited

feasibility and benefit to the quality of the reviews, and find that the suggestion of newer initiatives is met with only lukewarm interest from researchers. The Nature publishing group conducted a trial in 2006 to test a platform to combine open participation with peer review. It failed to show feasibility and sufficient interest from potential participants to be put into motion as an initiative. These findings all predate the crisis of confidence, however, and newer articles which survey researchers' opinions on OPR reveal increasing interest in OPR becoming mainstream in science, and evidence that opening the review process leads to higher quality articles with improved tone (see e.g., Walsh, Rooney, Appleby, & Wilkinson, 2000; Ross-Hellauer, Deppe, & Schmidt, 2017). Since the crisis of confidence has dominated much of the narrative surrounding research practices in the social sciences, clear agreement has been reached among open and reform activists that the peer review process is fraught with problems and is in dire need of reform (see e.g., Tennant et al., 2017; Csiszar, 2016). Ross-Hellauer (2017) cites a series of complaints against the current process, including a lack of reliability and consistency with the reviews, a lack of accountability with reviewers, lack of incentives for reviewers, and risks of subversion and damaging biases (both social and publication). In general, OPR is aimed at bringing peer review more in line with the emergent goals of the open science movement.

Ross-Hellauer (2017) identifies seven of what he calls 'traits' of OPR, some of which provide an insight into why some open and reform activists prioritize OPR. The most salient ones I have listed below, with my insights and comments in italics:

1. Open Identities – This trait involves unblinding, in which reviewers identify themselves in their reviews. The idea with blinding is to tackle potential bias (for instance, a reviewer might be inclined to submit a biased review for an article from a competitor research group), however true blinding often does not succeed and can lead to problems. *Open identities in peer review can enhance accountability of reviewers, and ensure that credit is given to reviewers for the work they do. It is possible that transparent reviews are of higher quality than blinded ones, as accountability to their words makes people more motivated to do a good job. Another benefit of open identities is that people might be more cautious in their tone and more inclined to provide reasoning for critique.*

2. Open Reports – Reviews are published alongside the paper. This trait ensures that possibly useful information contained in reviews is kept as a resource, potentially for reuse in other contexts. *Open reports increase accountability and provide incentives for peer reviewers, by making their reviews part of their documented scholarly activity. Open reports also provide article readers with additional context and information about articles that reviews sometimes contain, which can help for when researchers attempt replication studies.*

3. Open Pre-review – This trait involves preprints, a bottom-up incentive according to Armeni and colleagues (2021). Practically, this entails the posting of research on a repository before/in parallel with submission to a journal, and allows others to provide commentary on the preprint article. Leaning on existing repositories and services/software heavily, open pre-review is very popular in the social sciences. Preprint servers (such as psyArXiv on the Open Science Framework, Figshare and PeerJ preprints), and other services make Open Pre-review easy for researchers to use. *It allows researchers to assert priority when it comes to reporting results, and increases the visibility of research, enables open participation and can increase the quality of manuscripts at final submission. Preprints are popular especially in psychological disciplines, and have become a prominent priority for many methods-focused reform and open advocates.*

4. Final-version Commenting – This trait exists in many forms already. *Social media platforms like Facebook and Twitter, for instance, make final-version commenting easy and quick. The drawback here is that this kind of critique rarely improves the research it is aimed at (as the research is already persistent in the public record). It is unmoderated, tends toward tone problems (especially if the article in question is controversial) and can be of low quality. Many of the tone issues of the reform and open group stem from this trait (Derksen & Field, 2021).*

Wolfram and colleagues (2020) describe different levels of OPR implementation, with concrete examples. *Frontiers* and many other journals only implement OPR in that they require open identification of reviewers. Journals such as *PeerJ* provide open identification and open reports, but these are optional rather than required. *BMC* OPR involves mandatory open identification and publishes open reports alongside published outputs. Finally, *F1000* has totally opened the peer review process in an unprecedented manner. They do not use traditional accept/reject thresholds, but instead use a kind of endorsement system, whereby articles are only indexed in bibliographical databases once they are 'approved' by at least two reviewers. The entire peer review procedure is transparent from the get-go, and developments in the process are visible as they unfold.

Evidently, the OPR process is multi-faceted, complex and dynamic. Ross-Hellauer (2017) argues that this is a feature, not a bug: "The large number of possible configurations of options presents a tool-kit for differing communities to construct open peer review systems that reflect their own needs, preferences and goals." (p. 15). Although adoption of OPR is growing, there is still much room for improvement, with only a minority of OPR journals having adopted a completely transparent peer review process (Wolfram et al., 2020).

Pillar Four: Open Access Open access (OA) is a short name for what is a free online scholarship movement (Mering & Hoeve, 2020). It's a global initiative which seeks to grant access to knowledge sources (defined as scholarly articles, raw data and metadata, source materials, digital representations of pictorial or graphical materials, and scholarly multimedia material, according to the Berlin Declaration on Open Access, 2003) for free to everyone. It describes the scenario where no barriers (technical, legal or financial) exist to access, meaning that anyone can download, copy, read, search for/in, print or share the content, or use it for educational purposes.

The Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (BD; `https://openaccess.mpg.de/Berlin-Declaration`) is considered a great milestone in the OA movement's history. The BD was drafted to "promote the internet as a functional instrument for a global scientific knowledge base". At its core is the philosophy that the dissemination of knowledge is only complete when everyone has free access to it.

OA takes different forms. With Gold OA, an article is fully accessible online via a publisher's website, and APCs (article processing charges) are paid by the author (or their institution). An article is Green OA when it is available on an online repository (such as PsyArXiv, for instance). Gold OA is growing very fast in popularity, writes Björk (2017), due to large successful journals offering rapid publishing with innovative peer review approaches, however interest in Green OA seems to be languishing somewhat. Progress to 100% OA is slow, according to Björk (2017). Larivière and Sugimoto (2018) report that two-thirds of articles are available to read via at least one of these methods, and that half of all articles are classified as both Green and Gold OA. The purest form of OA is diamond OA, and describes cases in which an article has been published, but for which no APCs have been required of the authors.

Björk's (2017) article describes another form of OA, which he calls Black OA. Black OA is symbolically named, and refers to articles which have been made OA illegally. With Black OA, subscriptions, payments of any kind, and bureaucracy have been circumvented. Björk says that the blame for fast-growing interest in Black OA can be laid at the doorstep of academic social media/network sites. Björk refers primarily to Research Gate and Sci-Hub in his article. Research Gate is an academic social networking site where people can post PDFs of their publications, and others can request full texts of articles which aren't otherwise available.

Sci-Hub, a so-called "shadow library" website, takes Black OA one step further into fully illegal territory. Founded by Alexandra Elbakyan in Kazakhstan as a reaction to high paywalls, the site holds over 85 million research articles (as of January 10, 2022) making them freely available to anyone to search using article DOIs. Its domain name cycles frequently due to court injunctions, and has been championed by much of the academic community for its approach to providing OA. While some criticize it as a pirate site, the sense of indignation at big journals earning money off the backs of the scientific community is relatable. Black OA

has become popular because it is convenient for accessing paywalled articles, is not morally objectionable to most researchers, carries limited legal risks, and, of course, saves money. Björk (2017) estimates that as many as 50% of articles are available via Black OA channels. Black OA is a symptom of a greatly flawed traditional publishing model, contends Björk. He predicts that the popularity of Black OA will decrease as 100% Gold OA is reached.

One important predictor for OA adoption appears to involve research funders. Larivière and Sugimoto report on whether work supported by funding agencies with mandates for OA actually achieve compliance (2018). Some funders require authors to either publish the funded research output OA (i.e., via the Gold OA channel), or deposit the articles in repositories (the Green OA route). They find that figures vary widely, and depend greatly on funding entity. For example, they report that while 90% of National Institutes of Health (NIH) and Wellcome Trust funded research is OA, only 23% of research funded by Canada's Social Sciences and Humanities Research Council achieves OA compliance. Enforcement and the provision of infrastructure by the funder seems to be key: while the NIH and Wellcome threaten withdrawal of funding if recipients don't adhere to OA mandates, and require output to be published OA, many funders allow OA after publication. NIH and Wellcome also provide repositories and other support to help researchers more easily achieve OA. Discipline also seems to influence compliance: biomedicine, math, clinical and health science fields lead OA compliance, compared to social sciences and humanities fields.

The BD states that the success of OA is contingent upon active commitment from every scientific contributor to the knowledge base; Larivière and Sugimoto (2018) use the words 'cooperation and foresight'. They emphasize that OA needs a sustainable financing model (i.e., who will pay the costs that article production entail?). Suber, one well-known theorist and advocate of OA is optimistic. There's a lot of room for creativity when it comes to who pays the bills for OA, he says. The BD recommends practical approaches to encouraging the adoption of OA: advocating for OA publications to be formally recognised in promotion and tenure evaluations, and highlighting the merit of OA contributions to science are two such recommendations. As I mentioned just above, it is clear from Larivière and Sugimoto's work that funding bodies should support researchers with OA by providing platforms and repositories, as well as enforce compliance to OA mandates. Mering and Hoeve (2020) argue that while many OA-related goals may be successful, some are currently only aspirational, and it will be a matter of time before we see if they are feasible. Ultimately, sources tend to highlight OA's flexibility to change and develop, and are optimistic about its potential for the future of science. Evidently, support from institutions and funding agencies is crucial for the realization of many of OA's long-term goals.

Other Open and Reform Initiatives

Preregistration and Registered reports

Preregistration (PR) and registered reports (RR) are open science-related practices which come into play early on in the research: usually at the time of theory development and hypothesis generation, and almost always before the data are collected. PR refers to when a researcher chooses to write up his or her plans for a project – including study rationale, hypotheses, design and analysis plan – and makes them publicly available in some form. A growing number of researchers choose to upload preregistration documents onto sites such as the Open Science Framework (OSF), or onto their own personal websites (Chambers, 2014; Munafò et al., 2017).

RR takes the PR process further, involving the peer-review of the preregistration document through a publishing outlet, just as in the review process of a fully complete research manuscript. Once the preregistration plan has been accepted, the study has been accepted by the publisher in principle, irrespective of its outcomes, providing the authors have stuck to their plan or describe any deviations in detail. Dubbed a "critical part of urgent wider reform" (Chambers, 2013), both PR and RR are thought to aid the production of more trustworthy research findings. The key to their value is largely in the timing of when the study is registered – the proposed study aims, methodology and intended analysis are recorded and disseminated at the point of study design, but before data collection has taken place. RR holds a value over and above this, combining PR with peer-review and outcome-independent publication. The popularity of PR and RR appears to be gaining momentum in the scientific community as part of the regular research process: PR and RR are now widely accepted by at least 288 major scientific journals (for a full listing, see `https://www.cos.io/initiatives/registered-reports`).

The protocols of PR and RR have several qualities that have the potential to improve the credibility of the research produced through them. Their increasing use in the scientific community is accompanied by a growing body of literature that argues for their benefits (Alvarez, 2014). One key benefit described in this literature is that PR and RR provide researchers with a means to document the distinction between confirmatory and exploratory research findings (also known as "hypothesis-testing versus hypothesis-generating research", p. 4; Nosek et al., 2015). Clarifying this distinction can boost the reproducibility of research findings (Chambers, 2013; Wagenmakers et al., 2012).

Another major benefit of PR and RR is that they can help counter many QRPs researchers can engage in. As most of these QRPs can have a detrimental impact on the validity and interpretability of the statistical conclusions drawn from research findings, decreasing their prevalence can, in turn, help safeguard the overall quality and veracity of the literature. As already discussed in the previous

chapter, QRP refers to the many different choices that researchers make, and behaviors in which they engage, which push the boundaries of acceptable research practice (John et al., 2012).

Further benefits to adopting PR and RR are directly relevant to the researcher's career and academic reputation (Wagenmakers & Dutilh, 2016). It is thought that researchers will produce higher-quality, more reproducible research. This will ultimately benefit them as scientists, for different reasons. It is possible that researchers' academic work will be trusted more by other academics when PR or RR have been part of the research process. By adopting PR and RR, proponents argue that researchers will increase their chances of getting articles accepted by journals, regardless of whether or not the results obtained favor their hypotheses.[3] Finally, they may be more confident in trusting the work of colleagues in their own fields if they know that others' work has been preregistered, or is a registered report. In the specific case of RR, authors will benefit from extra review and input on their methodology before they conduct the study. This allows them to save their time and resources for the highest-quality studies.

Myself and colleagues attempted to assess whether PR/RR studies were perceived as more trustworthy compared with their non-registered counterparts. The study features a promising pilot which demonstrated a strong, predicted effect of registration on trust. The full study, however, conducted in 2019 and based on the data from over 200 academic participants, yielded ambiguous evidence. With such a small sample (an N of 209 spread unevenly over 6 experimental conditions constitutes very low power to detect a difference), this is unsurprising.

Incidentally, our study, an RR itself, demonstrated a further benefit of RR. That is that when they use the RR format, researchers have the freedom to report on the study truthfully and in detail, without fear that the study will be rejected by the journal for the lack of a compelling 'story' (providing the plans were adhered to, or sufficiently rationalized if deviations occurred). In the specific case of the study in question, my coauthors and I could: "...be transparent about the trouble we had with our sample, our findings, and that we cannot conclude anything from the study. This may then serve as a warning to other researchers who may attempt to study a similar phenomenon." (no page numbers given; Field, 2020) These reflections underscore the importance of new initiatives like PR and RR in overhauling a publishing system riddled with limitations of reporting and publication bias. They underscore the value of transparency, and the fact that transparency is made possible, and even attractive, by changes to article report formats and protocols.

[3] Whether this benefit is actually reaped by those who preregister as well as by those who submit RRs has yet to be established – it is possible that acceptance of null results remains challenging for publications even when they have been subject to PR.

Journal Clubs, Project Workflow, and Open Education

Open science and the science reform movement has affected the way people practice research at every level. Kathawalla, Silverstein and Syed (2021) describe a number of open science practices that I have not yet covered, such as journal clubs, project workflow, and transparent manuscript writing. Although their article is geared toward guiding graduate students through open science, the material is relevant to anyone interested in adopting transparent, credible, reproducible and accessible research, they say (p. 2). I will briefly walk you through these practices too, starting with journal clubs.

Kathawalla and colleagues describe journal clubbing as an easy way to do open science. Journal clubs are meetings that occur regularly, and involve a group of people discussing an article that has been selected and read by the group before the meeting. They vary in how formal they are and are aimed at increasing familiarity of the group with a particular topic and literature body. In this case, of course, that topic is open science and its related topics (metascience, philosophy of science, etc). They are straightforward to start up (Kathawalla and colleagues suggest that all you need to initiate a club is one other person who is interested in participating with you), or attend, when one in your sphere already exists. They can also provide a support network and open up opportunities for collaboration with like-minded others.

On the bureaucratic, infrastructural side of open science, are practices relating to reforming and opening up one's project workflow – another open science practice judged to be easy by Kathawalla, Silverstein and Syed (2021). Project workflow, they write, describes how one organizes projects and moves through the stages of the research process, and involves one's approach to file management and storage and version control. Committing to a clear and dedicated project workflow system facilitates reproducibility and collaborations, can help minimize mistakes and will help 'future you' with organization, they say. Opening your project to others with whom you are working can be a good motivator to work in a more organized fashion, and to be completely transparent about your research process, even if it is not publicly open. The authors share several ideas for platforms and systems to assist with project workflow, including setting up a project on OSF.

Finally, Kathawalla, Silverstein and Syed (2021) provide guidance on transparent writing, a practice which is quite self-explanatory if not necessarily easy to execute well. Transparent writing involves being open about all aspects of the research process, to as much a degre as possible given the constraints of journal requirements. Kathawalla and colleagues emphasize the need to justify decisions you have made in the execution of the research, including, for instance, your stopping rule, if you are using a frequentist statistical framework, and why you have made certain analysis choices.[4] Kathawalla et al., state that transparent writing

[4]A stopping rule is an arbitrarily chosen point at which one stops data collection. Not selecting

can provide others with a good sense of what was done, and helps them "calibrate the implications of the findings". (p. 7) They suggest it will help hasten the peer review process, because it will decrease the likelihood that reviewers will have to ask for additional information. An additional benefit to transparent writing is that if can facilitate replication attempts (in the case of empirical and computational research studies) because the method is easier to follow, in comparison with articles written in a more opaque manner. As with project workflow, the authors include helpful recommendations on how to improve your writing, including an article from Gernsbacher (2018), which guides readers through a writing process current and consistent with open science practice.

The Psychological Science Accelerator is an example of a kind of initiative which employs large numbers of individuals and lab groups to achieve impressive and otherwise impossible open science goals. Its website describes it as "... a globally distributed network of psychological science laboratories (currently over 500), with over 1400 members representing 71 countries on all six populated continents, that coordinates data collection for democratically selected studies." It is a global network of laboratories, which work together to generate reliable and reproducible psychological research.

Other initiatives tackle open science from the perspective of educating researchers, and facilitating their adoption of open science practices (and see also Kathawalla et al., 2021). A key example of this is FORRT – the Framework for Open and Reproducible Research Training. FORRT's website (`https://forrt.org`) states that the organization's goal is to provide an educational infrastructure to support the teaching and mentoring of open science philosophies and principles. It approaches this goal with an evolving, and community-driven organization, which works to raise awareness of the pedagogical implications of open science and the associated challenges. The organization actively supports sharing of teaching and mentoring materials, which in turn can facilitate access, discovery, and learning to those with reduced access to such resources. FORRT was established at the 2018 SIPS conference at a hackathon focused on teaching reproducible and open science. It is supported by the UKRN, the Center for Open Science, the Psychological Science Accelerator and the journal *Meta-Psychology*.

The Declaration on Research Assessment (DORA) was established in 2013, and is aimed at abolishing the use of the journal Impact Factor (IF) as a means of evaluating research and researchers' output. The vision, broadly, is to advance practical and robust approaches to research assessment globally and across all scholarly disciplines. It aims to raise awareness about new tools and systems

a stopping rule until you have seen the data strays into the territory of a QRP, because the p-value is highly sensitive to sample size, and can be easily manipulated to significance simply by collecting 'just a bit more data'. Berger and Wolpert (1988) describe the Stopping Rule Principle, which holds that statistical conclusions should be independent of when we decide to stop collecting data. Wagenmakers, Gronau and Vandekerckhove (2019) provide an illuminating discussion about stopping rules in a preprint

in research assessment, and encourage the use of metrics that promote transparent and consistent decision-making. It seeks to help develop new policies and practices for funding, promotion and hiring decisions by institutions, and to push for research assessment reform. Finally, DORA's signatories aim to improve equity by calling for representation of researchers in the setup of research evaluation approaches which address the structural inequalities which exist in academia. According to the website (`https://sfdora.org`), DORA approaches these goals practically: Through community engagement, resource curation and development, partnership with other organisations, provision of advice to institutions and funders, and convention with a variety of stakeholders. DORA's website reports that, as of the beginning of 2021, more than 2,200 institutions and over 17,000 individuals worldwide had signed the declaration. *Springer Nature* signed the DORA 2021, and is the largest publisher to have done so, to date.

Adopting Open Science

Despite numerous potential benefits to adopting the open science practices outlined in this chapter, most articles report that the adoption of open science on a wide scale has been somewhat slow. On their website (`https://www.startyourosc.com`), the International Network of Open Science and Scholarship Communities (INOSC) states that open science is "not yet the norm". Armeni and colleagues (2021) "actual adoption lags behind on the widely shared vision" of open science (p. 610). Other literature suggests a variety of potential barriers for open science adoption. For instance, Allen and Mehler (2019) discuss three challenges: restrictions on flexibility in the research process, time cost and incentives not being in place to facilitate open science practice uptake. The argument that open science places some restrictions on flexibility in the research process is not a new one. When preregistration and the registered report format began gaining momentum, for instance, Sophie Scott writing for the *Times Higher Education* announced that "preregistration would put science in chains". (2013)

With PR and RR 'fixing' parts of the research process like timelines and hypotheses,[5] they put limits on the continuous learning process that the scientific process represents for the researcher and may discourage some researchers from adopting some open science practices (Scott, 2013). Allen and Mehler (2019) echo these concerns, and discuss that the benefits of PR and RR are difficult to weigh against the time costs, especially for early career researchers (ECRs). Publishing mechanisms like RR and the work and organization involved with making aspects of the research process open swallow up time that, for example, graduate students might not have, they say. Finally, Allen and Mehler express the concern that the benefits of using open science incentives don't outweigh the loss

[5] I.e., by making these elements such that they cannot later be altered, except by noting them as deviations to what was preregistered and setting them apart from the confirmatory hypotheses and findings in the written report.

of productivity and flexibility, because there are not many incentive structures in place to reward the use of open science practices, but ultimately argue that ECRs are likely to benefit in the long run from adopting open science practices. The standards of open science in academia are still developing, and it will take some time before open science practitioners have the support needed to fully commit to open science without a potential loss of competitive edge against peers.

Gagliardi, Cox and Li's (2015) article on institutional inertia and open science describes operational and institutional barriers to peoples' engagement with open science practices. They mention four operational barriers, which concern 1) the difficulty involved with assessing the quality and impact of research results when disseminated as preprints (or other forms than peer-reviewed articles), 2) the time taken by researchers in contributing to open science initiatives, 3) the difficulty faced by individuals when they engage the public in research in terms of the resources and platforms such a venture would require, and 4) that researchers may lack the skills to engage in some open science tools which are necessary to contribute to some open science practices.

Gagliardi and colleagues' findings indicate that the main operational barrier is the first – assessing the quality and impact of articles which don't conform to the traditional peer-reviewed article (2015). They found that concerns over lacking skills to contribute to open science was second most important to participants, followed by the concern over lack of time to participate in open science. The institutional barriers they identified are 1) lack of funding to engage in open science initiatives, and 2) inertia (which they associate with the traditional approach to scientific practice) which hinders the uptake of open science by not providing incentives to researchers to share their findings). According to Gagliardi and colleagues, the latter institutional barrier is the primary one faced by researchers. They also report that both discipline and age/career stage affect potential adoption of open science. Older researchers were more reluctant to adopt certain open science practices, like sharing aspects of the research process before publication, compared with younger counterparts. They report that researchers in the natural sciences and engineering and technology are more likely to share findings compared with researchers in other disciplines (though the sample sizes vary quite a bit across the disciplines that were surveyed; e.g., $N = 34$ for the natural sciences, compared with $N = 4$ for the agricultural sciences).

Armeni and colleagues (2021) share a different potential set of challenges to the wide-scale adoption of open science. They discuss the issue of reaching a critical mass of researchers to achieve broad cultural change. They contend that bottom-up open science communities (thanks to their peer-based nature and format) have the greatest potential leverage to reach a critical mass, as they provide researchers themselves with a voice. They also talk about perceived costs of change for the academic community. They note different sources of resistance among researchers. For instance, ECRs may experience hesitation from supervisors and collaborators who prefer older systems, and better-established closed

workflows. Researchers might be afraid of others scrutinising their research, and fear getting 'scooped'. They suggest that this kind of resistance can be especially difficult to tackle for researchers existing in local environments where open science is not the norm. Like Gagliardi, Cox and Li (2015), INOSC share that approaches to adoption of open science can vary as a function of scientific discipline. Part of this may be because the (perceived) usefulness of open science practice is not constant across fields. They give the example that data and materials sharing is more common in biomedicine than psychology, while OA is more popular in psychology. Although this diversity has the potential to open up "fruitful avenues to share best practices across disciplines" (p. 15), it does provide a challenge when seeking to understand and affect behavior change toward uptake of open science.

Another possible deterrent to the adoption to open science practices is that people might find the sheer number of open science practices alarming. Most academics are already time-poor, and are unlikely to make things even harder for themselves by wading into the open science morass. One key take-home message in the conclusion to Kathawalla, Silverstein and Syed's (2021) article is to ease yourself into open science practice by engaging in simple and easy changes to your approach to science to begin with. Robson and colleagues describe "nudging" – the adoption of "small, easy-to-avoid changes to a person's decision-making environment that alter behaviour in a predictable way without forbidding any options or using economic incentives." (p. 5; 2021) Their recommendation reflects the 'open science buffet' analogy, introduced by cognitive scientist Christina Bergmann (Bergmann, C. [@chbergma], 2019) and elaborated on in a piece by Whitaker and Guest (2020) which has caught on in the open science community. Bergmann cautioned new open science adopters against attempting too much open science too soon, as it can leave them overwhelmed. In other discussions about the buffet, on Twitter for instance, people explore the idea further. Bergmann posted: "So glad the buffet metaphor is catching on, there are so many solutions out there. Don't try to stuff yourself on everything, select what works for this study and let's steadily improve our fields... #openscience". This was a tweet in response to another's post summing up a talk by Nuijten at the International Convention of Psychological Science, where Nuijten discussed a 'buffet' of practices including those relating to transparency,preregistration, multi-lab collaborations, and attending the SIPS conference, among other things.

Although the barriers discussed here should be taken seriously by those interested in motivating others to join the open science movement, one final barrier to the uptake of open science concerns inclusivity. To illustrate, consider again the article of Whitaker and Guest (2020). They extend the buffet analogy, bringing in the idea that diversity and inclusion are foundational to open science. They don't mince words: "There is no open science if science is not open to all." The authors urge their readers to re-imagine the open science buffet as a pot-luck, a meal where different open science practitioners bring different skills, approaches,

questions – dishes – to the table of open science. Members can contribute in the way in which they're able, and open science as a community is free from the limitations of homogeneity, and benefit from greater diversity. They advance the point that the essence of open science is the same as the motivation for conducting research in the first place, that is, to learn from and educate others. Emphasizing that the open science movement can benefit from a variety of different contributions may be one route to sparking interest in the movement for many people. Shifting the culture of the movement such that people genuinely feel welcome must be the first thing to change, they argue. If people feel as though the doors of open science are not open to them and their unique contribution, they are unlikely to be swayed by other more objective attempts at facilitation.

Levin, Leonelli, Weckowska, Castle and Dupré (2016) emphasize the complexity of adopting open and reform practices. They write:

> . . . decisions about what to make open, and how and when, can vary widely depending on a number of factors: the ethos and hierarchical structure of the research field and community, the varying degrees of technical difficulty and labor involved in disseminating resources and results, the existence of useable infrastructures, and the degree of competitiveness and commercial stakes around the given research activity. Research methods, processes, settings, and goals are highly contextual, such that Open Science policies need to remain sensitive to the diversity of research contexts to which they might, or might not, apply. (p. 137)

I also consider the words of Robson and colleagues, as they describe the role that each researcher plays in putting pressure on institutions and big stakeholders to promote and facilitate open science practices: ". . . Significant improvements in infrastructure, norms and reward structures are needed before policy change is even possible or seen as necessary. The behaviours of the various agents in the scientific community ultimately determine the quality of the research that is generated and disseminated." (p. 31; 2021) The onus is on the individuals in the open science community to drive change, to shift its culture such that it actively and effectively welcomes newcomers, and demand reform from elements in the greater research ecosystem such as publishers, funders and universities.

The Rise of a Community

While I aimed for this section on practices to be largely descriptive, it reveals interesting things about the community or communities involved. First, open science and reform is not just a movement generated by the community, which affects only them and their close periphery. That governments, independent funders and policy makers get involved in imposing and enforcing requirements (about,

for instance, sharing code and data) tells us that open science practices have gone beyond their origins within the 'reform' or 'open science' sub-community of academia.[6] They have come full-circle, and have had an impact on a much larger scale of science. This is significant, because this pervasiveness complicates the picture for the open and reform group in terms of how they derive individual as well as a collective identity from these practices, and how they go about negotiating their joint enterprise.

Second, this section has revealed what open and reform science really *is* in practice. All of the tools, primers, tutorials, fora, institutes, agencies and frameworks that exist (again, I have only started to scratch the surface of what this already-prolific group has achieved) are artifacts of this community, of these communities. Even the kind of language surrounding many of these artifacts (*You can do open science too! It's easy, all you have to do is. . .* ; Kathawalla et al., 2021) tells us something about this movement. These things all reify this community's joint enterprise, making concrete examples of otherwise abstract values and beliefs about what is best for a new, reformed science. In giving researchers tools and language and platforms with which to practice, what open and reform science *is* is made tangible and tractable.

[6]It also indicates that governments and other similar governing bodies have a stake in open science which opens up the possibility that their interests are not the same as those of researchers.

4

Open Science and Reform Online

4.1 The Reform and Open Science Community

One profound outcome of the crisis of confidence is the emergence of a reform community which began to form as people collaborated to develop ways to resolve the crisis.[1] This 'community of critics' (Derksen, 2019) is subsumed under the umbrella of the broader scientific community, and represents hope for science which has emerged off the back of one of its darkest periods. It is not immediately clear when this community truly began its formation, however its profile has become somewhat distinct and more prominent in the years following 2011.

Possibly, the main catalyst of the community's formation is due to the generation of groups of individuals, which came together to spearhead and then maintain the initiatives mentioned before. Unsurprisingly, such clusters of like-minded individuals would eventually establish ties with one another and join up to form greater groupings with wider boundaries. This collective is supplemented by what I call 'content researchers', who contribute to the movement by choosing to practice more transparent science, and improve their general research practices. These are people like social psychologists who watched their discipline fall under attack in the wake of cases like Stapel's have risen to the challenge of changing their modus operandi. They are a varied group of researchers who see the possible benefits that reforming research can bring to their fields.

This group has formed its own borders and cultures, and established protocols and rules. Some of its members are somewhat aggressive in their attempts to improve science. They have come to act as the 'foot-soldiers' of open science (Bartlett, 2018), and often take the role of the watchdogs of the scientific community's integrity. Some members even go so far as to call themselves 'data

[1] Though, I should flag here that not all the open science communities find their origin in the crisis of confidence. As I will discuss in **Chapter 7**, different sub-groups within the broader group have varied and unique trajectories.

thugs' (Marcus & Oransky, 2018). They set out to educate the scientific and lay-communities about topics relating to scientific transparency and integrity. Other members of the group simply discuss good science practices and share resources and tips with one another on how to achieve more transparency and better quality research output.

The Location of a Community

The central question for me has been how do I go about identifying the members of this potential community in order to study their culture and structure? Who *is* it that I am observing and interviewing during my fieldwork? Community is a difficult concept to grapple with, as this chapter shows.

When I think of community, I think back to my childhood in semi-rural Australia. I think of going to the neighbour's house to 'borrow' a cup of sugar, or to use their swimming pool on days when the dry, brown grass and packed, scorched dirt were too hot to walk on with bare feet. I think of how we never locked the front door at night, or when we left the house to go to the non-denominational congregational church down the road. I remember sitting on the front gate with my best friend who lived a few streets away, eating huge wedges of watermelon that tasted like the sun. My dad knew the owner of the local chip shop, and the publicans of the many pubs that dotted our small semi-rural coal-mining town, and the random person walking down the street at any given time. We seemed to know everyone else's business, maybe even before they did. Family-owned businesses were actually still owned by the original families; my favourite meat-pie shop had been using the same secret recipe for their steak pies as they had for the last hundred years. I left that small town 16 years ago, to study as an undergraduate in the 'big smoke' as my mum used to call it, and nowhere else I have lived since has had that sense of community. [This excerpt is a reflection taken from my field notes.]

But community doesn't have to be tied to a location, as mine was for me more than two decades ago. At least, not anymore. Certainly not since the invention of the internet, and from it, the development of social media platforms (Lovejoy & Saxton, 2012). Hillery (1955) described a total of 94 different definitions of the notion 'community', settling on one that involved social ties and geography as key components. Newer definitions that account for McLuhan's global village drop the geographical element, and instead describe community as groups of people who share a common identity and character, and are located such that communication is possible (Ramsey, Annis, & Everitt, 2002). Indeed, co-location can be a resource (Beaulieu, 2010), but it is neither sufficient nor required for community formation. Certainly, it does not guarantee availability of interaction with others (Goffman, 1978). Clifton (1999) adds the element of trust to what makes community – it cannot easily develop when people don't trust one another. Wenger's (2000) definition emphasizes the 'joint enterprise', the establishment of mutual relationships through interaction, and a shared repertoire of resources.

Community Online

Although it can be useful to define community, especially in newer literature where online communities are more and more frequently the subject of research, Bruckman observes that "much ink has been spilled trying to work out which online communities are really communities" (p. 463; 2005). She suggests that developing definitions for online communities might not be useful, and takes a somewhat nominalist view of online community delineation. She advocates for embracing the concept of community as one with fuzzy boundaries, that is more appropriately defined by its members' characteristics. That is, to her, an online community does not have an essence that defines it per se, rather it is simply a collection of individuals and attributes. Rheingold didn't see defining online community as so problematic, though he does provide a similar description of online communities as "social aggregations" that emerge when enough people engage in public discourse for long enough, and with enough "human feeling" that personal relationships form in cyberspace (p. 6-7; Rheingold, 1993). Akar and Mardikyan (2018) lay out a set of discrete characteristics in their description of online community. They say that online communities need adequate numbers of members actively communicating with other members. Community members should share a mutual interest in interacting with one another, and their discussions should be conducted with respect to a code of conduct; rules and practices for engagement. Akar and Mardikyan write that participation is the "fundamental mechanism of online communities" (p. 1), and assert that members need to feel a sense of belonging in the community. All of this implies to me, that although defining online communities

Virtual platforms, where users can "affiliate with a co-present, impermanent community, by bonding around evolving topics of interest" (p. 791; Zappavigna, 2011), often facilitate flourishing communities, which can be described in terms like those above, and which are bound together by interests shared by people.

Social research sheds light on two fundamental dimensions of society: culture and structure (Gleave, Welser, Lento, & Smith, 2009). Culture refers to the social behavior, norms and customs of a particular group of people; structure refers to patterns in ties between the group members. Twitter provides an excellent opportunity to study both of these elements of the open and reform group as represented on Twitter.

4.2 Twitter – A Location for the Online Reform Community

Twitter, a 'micro-blogging' social networking service, is one of the most visited websites since 2013, having grown greatly in popularity since it was founded in 2006 (The Editors of Encyclopaedia Britannica, 2022). The site handles a huge

amount of user-related data. On average, approximately 6,000 tweets are posted per second; 500 million per day. Now, 330 million people use Twitter monthly, with over half of these people logging in on a daily basis. Over 70% of Twitter users use the platform for daily news, and 85% of small to medium-sized business use it to provide fast-response customer services (Wojcik & Hughes, 2019). In the fourth quarter in 2019, the site was worth just under 25 billion US dollars (Statista Research Department, 2022). On Twitter, users interact and post using 'tweets', tiny snippets of text. These are used by posters to share information, or to respond to the information others share. The site allows the use of up to 280 characters per tweet. It takes less than a week for the site's users to post a billion Tweets (Statista Research Department, 2022). Founder Jack Dorsey captures the concept at the core of Twitter as he explains its name: "... we came across the word 'twitter', and it was just perfect. The definition was 'a short burst of inconsequential information, and 'chirps from birds'. And that's exactly what the product was." (Sarno, 2009)

In addition to being a repository for a huge archive of tweets, Twitter also holds demographic and geographical data on its active users. Other data Twitter handles, as a core functionality, is information about social ties. While it was not initially intended as a social networking site (the purpose was more aimed at information exchange), Johnson and Yang found that one of its primary functions is a social media service (i.e., an online platform aimed at helping users create relationships with other people who share an interest, cause, or in-person relationship), as it facilitates 'meeting' of users who share common interests (2009), and provides ways of easily maintaining the ties established (e.g., by using the 'direct messaging' service to privately communicate with others, or by 'tagging' other users in Tweets). Each Twitter user can use the site's inbuilt search engine to find like others, or follow suggestions made by Twitter's so-called 'Who to Follow' algorithm-driven recommendation service. Users can 'follow' users which interest them, and be followed in turn, if the interest is reciprocated. Myers and colleagues suggest that such networks are most appropriately named "interest graphs" (as opposed to the perhaps too-simplistic moniker 'social graphs') given that they show patterns of ties which are born of interest-driven following behavior (2014).

Members of the open science and reform community on Twitter follow each other based on interests and shared background, with the addition of the broader academic context. They use it as a social medium and an information/news medium. They share information on initiatives, new articles and news, provide support, to facilitate networking and scientific collaboration, for educational reasons, and to develop social capital. They also use Twitter to gossip, chatting about politics, and to exchange recipes and photos of things they have cooked, and to share pictures and videos of pets, home improvements and holidays.

The 'following behavior' of open science advocates and reformers can be captured in network objects and can be used to look at 'raw' community structure,

rather than to describe content-related links between agents (which tweet/retweet content analyses might be used for). Specifically, the 'follow graph' (Myers et al., 2014), an intuitive way of describing and understanding social capital through Twitter relationships, features heavily in this work. The network data collected as part of this work will be explored and described in **Chapter 8**.

Twitter's Role in Community Building

Twitter communication can take different forms, and so can facilitate building a community in different ways. At its most basic level of functionality, Twitter gives the possibility of *posting* (*tweeting*, *retweeting* and *hashtags*).[2] Tweeting is simple: one composes a short written text (using up to 280 characters), and publishes it. This simple mode of communication is used for different purposes. Java, Song, Finin and Tseng (2007) define four categories of 'user intentions' with respect to posting: daily chatter, conversations (threads), sharing information and reporting news. Naaman, Boase and Lai (2010) split Twitter communications into categories based on tweet content. These categories include self-promotion, opinions/complaints, statements/random thoughts, questions directed at followers, and presence maintenance. Retweeting, considered to be a reliable indicator of both popularity and impact of tweets, is where a user re-posts the content of another user's tweet (along with the tweet and the user's metadata). They make up more than a quarter of all tweets, and have the effect of easily, quickly and substantially amplifying a message. A hashtag is a word or phrase logically associated with a certain event or concept. Hashtags are searchable through both Google and the native Twitter search engine, and help categorize Twitter content.

Although not originally intended as a community-building site, Twitter has come to host different groups of people such as those who advocate for open science philosophies and practices. Communication between open and reform science community members (as with members of the greater academic Twitter community) is like that of any other group on Twitter, and mostly conforms to the categories laid out in the articles of Java, and Naaman and colleagues (2007; 2010). It can be informative, with members sharing information about things such as new publications, grants and awards received and new initiatives. Notice-board style memos such as those advertising academic positions, and requests for advice on certain topics (like statistics or methodology) are also common. Users also share personal information, such as their progress with baking sourdough bread, when they welcome children and new pets to their families, and frustrations about working hours, sickness, or unfair treatment by colleagues or students. As with a physical community of people, members sharing information about themselves with other members strengthens ties. Twitter also provides

[2]Following, another basic functionality of Twitter, can also be considered a communication tool. Tie formation is related to network structure rather than communication per se, however, and so I will leave discussing it in detail until **Chapter 8**.

a platform for real-time discussions about random minutiae. This often seemingly pointless chatter (which Jakobson called 'phatic communication'; Jakobson & Sebeok, 1960) helps generate a sense of proximity and familiarity, from which community can be built (Stephansen & Couldry, 2014).

Twitter is also used by the open and reform science community to communicate about larger issues in society which affect them personally in some way. During the period of time I was conducting the virtual ethnographic work this thesis presents, several important events happened which affected many of the open and reform science community's members directly. First, the coronavirus outbreak, which began over the transition between the years 2019 and 2020, has irrevocably changed the world for everyone, which of course includes all of the open and reform science community. Since then, many events (many taking place in the United States, but visible to the rest of the world) have impacted much of the open and reform science community, especially as many of them reside in the USA. For instance, a number of highly controversial cases of police killing people of color in America took place between 2019 and 2021. The presidential elections in the USA which saw Donald Trump get elected, for a term, then lose the presidency to Joe Biden in 2020 were also highly impactful on the open and reform science community, as was the war between Ukraine and Russia which began with Russia invading Ukraine in February 2022. Trans and sexism issues played a recurring role in affecting many of the group's members too.

Twitter also indirectly facilitates organization, networking and communication around conferences, which in turn encourages longer-term community development and maintenance. Before conferences, for instance, users can start hashtags to drum up interest in attendance, hype up existing delegates or to organize social events around the conference. They can also be used to set up 'early adopters' (Reinhardt, Ebner, Beham, & Costa, 2009). They can be used to make announcements related to conference programs and registration, help attendees organize travel and accommodation, and set up pre-conference programs. The hashtag #SIPS2018 was used in relation to the 2018 SIPS conference in Grand Rapids, Michigan, USA. A prominent open and reform science community member, Brian Nosek (@BrianNosek), tweeted: "If you care about research integrity, open science, and reproducibility, then follow #SIPS2018 for the next few days. You won't regret it. And, next year, whatever your field, you might decide to go to the conference in Rotterdam in July: http://improvingpsych.org" (2018) on the day before the conference began. This tweet clearly broadcasts the values of the group attending SIPS, draws like-minded others to the content on Twitter, likely expanding the community.

Although it would have been too late for anyone not already registered to still attend SIPS, the tweet stirs interest for the next SIPS conference and creates FOMO (fear of missing out) for future delegates. One attendee of SIPS 2018 describes even feeling FOMO themselves over people posting about already being in Grand Rapids (the conference venue) in advance of the conference, despite the

fact that they themselves would be attending the conference later in the week. During SIPS, the conference dinner was advertised on Twitter using the hashtag, and people used it as a way to meet new people in the open and reform science community.

Other uses for the conference hashtag are to encourage people to post content of talks and sessions in real time, as they happened. Some even used the hashtag to ask questions raised in the talks. The role of hashtags on Twitter plays during a conferences centers around increasing discussion and information exchange. After the conference, the nature of the conference hashtag's use evolves again, to thanking people (organizers thanking delegates or other organizers, and delegates thanking colleagues and organizers), reflecting on the experience, and in the specific case of the organizers of the conference,[3] gathering feedback for future conferences (Fried, 2018).

The use of conference hashtags is useful for open and reform science community members for logistical and practical reasons, but their use also has a strong role in cultivating a feeling of community and belonging. Using hashtags to evoke FOMO in non-attending community members and using them to network with present community members has much to do with creating and reinforcing a sense of in-group in the community. Hashtag use can also reinforce the community's boundaries in a negative way. OS practitioners who have had toxic dealings with members of the in-group are likely to feel even more part of the out-group when they see people they follow using in-group related hashtags and engage in in-group specific discussions.

[3]Eiko Fried, an open and reform science community member, whose first SIPS conference was the one in 2018, extended the reflections he made to a blog post titled *SIPS18 collected resources, and reflections of a SIPS virgin*.

4.3 Guiding Questions and Research Aims

Broadly, the research I present in this thesis seeks to examine the open science and reform community which has emerged from the 21st century crisis of confidence in science. A rich understanding of this community and how it approaches and achieves its goals has yet to be reached, and I aim to contribute to this understanding.

The research I present in this dissertation was guided by the following questions:

- Ethnography
  - What are the characteristics of members of this community?
  - What do they prioritize in the pursuit of scientific reform?
  - How do they see the issue of membership and identity in the context of their shared goals?
  - How are boundaries established and maintained?
  - How are problematic actors handled?
- Network Exploration
  - What is the best way to define the community based on its online presence?
  - What is the overall structure of the community?
  - Are there detectable sub-groups within the structure?
  - How does the reform community use Twitter to achieve its goals?

5

Ethnographic Methods

> *The native researcher chooses not only a project in which she is deeply situated, whether by geography or simply 'inside' experience, but also one in which she is invested in those factors, as they inform the 'act' of research.*
>
> (Kanuha 2000, p. 441)

Ethnography was a natural choice for pursuing my research questions, which I realised in transitioning from investigating the quantitative predictors of reproducibility, to exploring how a community is defined and structured, and how its constituents identify themselves, and engage with one another. This chapter describes my use of observation, conversation, interview and document analysis – ethnographic methods I used during the process of this project. I explain the processes I have followed, and discuss some perspectives on them in the literature which have been especially formative or salient to me as I have studied the craft of ethnography. I also review some of the methodological challenges I faced during this project, and explore my stance and how it has influenced my study. Finally, I discuss reflexivity in some depth (which involves examining one's own judgments, practices, and beliefs during the research process with the aim of identifying biases that can affect one's interpretations; Finlay, 2002), as I have employed it as a means of continually scrutinizing myself in my position as researcher and part of the community I am researching.

5.1 Demarcating 'the Field'

I am studying a group of people. We are now in the 21st century, characterised by Web 2.0 (the 'social' web), and the individuals making up this group meet online as well as in person. This means that I must use a combination of tools to capture their engagements with one another, and their participation in the collective. The

older components in my toolbox are those which ethnographers have used since Herodotus of ancient Greece, and involve methods of physical observation and interviewing. The newer tools are merely the older ones reimagined for a virtual application. In this case, the observation is no longer only physical, but conducted partly in an online environment. This provides a challenge which I explore in this section.

"Where is 'the field'?" is one of the most difficult questions I have faced during the formulation and reformulation of this project's aims and methods. This is difficult, because it is not "self-evident where, how long, or with whom one should conduct the fieldwork" (p. 167; Scheffer, 2007) in ethnographic research. This is doubly difficult in the case of ethnographic research that is multi-sited, including research that is conducted (partly) online because the "ethnographer is confronted with alternative versions of where/what the field is" (p. 168, *ibid*), and must share their attention between different 'facets' of the field, integrating material and experiences from different sources and platforms. I will briefly explore the layers of complexity I see in this challenge.

The Internet as a Medium, and Reform as a Collective

As Beaulieu points out in her article exploring the challenge of internet-based ethnography, the discussion about the appropriateness of ethnography for studies which use the internet as a 'field' is not a novel one (2004). She describes how technologies like the internet have been seen as a barrier to ethnography, with many arguing that "computer-mediated communication" would not provide a sufficiently rich substrate for meaningful social interactions to be used as a a field for ethnographers, in part because of the lack of face-to-face engagement (p. 143, *ibid*). This marks one layer of complexity.

Another layer of complexity concerns the would-be reform group itself. Traditional cultural anthropology tends to deal with a cohesive, and clearly formed collective in a single physical location. In contrast, sociological ethnography involving complex groups of people (which describes the research this dissertation showcases) often deals with the study of 'fuzzy fields' (Nadai & Maeder, 2005). That is, "fields without clear boundaries with regard to many dimensions" (p. 4, *ibid*).

Some characteristics of the community I have studied make the issue of field delineation an especially salient one. For one thing, reform, open science and metascience represent an inherently reflexive research area; one in which a person can simultaneously be the researcher and the researched. For another thing, this group of people is spread across the world, meaning that even if I did not conduct online research at all in the pursuit of studying this group, it would have been hard to know where to focus my time and attention during the fieldwork.

In other words, for me, determining what constitutes 'the field' has not been straightforward. Multi-sited ethnography presents researchers with methodolog-

ical and practical challenges which must be overcome to produce valid findings (Nadai & Maeder, 2005). That being said, I have places at which to begin. As I said earlier, the platform of Twitter provides a wealth of research material for my use in the context of this project.

Twitter: One Platform for Virtual Ethnography?

Twitter allows for possibilities that have not been available to earlier metascientists,[1] or science reformers. For instance, the platform allows for the establishment of a network which transcends grouping by location. Although research groups across the globe have met at conferences and symposia for generations, there has always been a strong sense of co-location which grounds the identity of the groups, and allows them to identify and refer to those in other groups. Often, it is the case that one research group might refer to another by their location (e.g., members of our own metascience group in Groningen, sometimes refer to the metascience group at Tilburg University as 'the Tilburg group'). The distinctions between labs or research groups are temporarily blurred during (inter)national meetings like conferences and workshops, however they fall predictably back into place once the delegates return to their respective homelands after the conclusion of the meeting. The Twitter platform allows the metascience network to transcend the limitations of co-location – actors may engage easily in inter-lab discussion and debate and share ideas and results of projects freely in an easily-accessible public forum.

Twitter also allows for the possibility of a specific kind of ethnography – not the original kind in which the researcher positions themselves physically within a group, and makes observations of in-person interactions – but another, contemporary kind, in which a researcher positions themselves *virtually* within the group. This possibility challenges traditional assumptions about co-location, as Beaulieu (2010) discusses, and allows for the researcher to consider new avenues for the study of a community that do not involve a rigidly defined physical location, such as a lab. Beaulieu calls this concept 'co-presence', and juxtaposes it with co-location. Co-presence in the absence of co-location, that is, being present without being physically among a group of others, is an important concept for those interested in conducting ethnographic research on an online community.

Co-Presence and Co-Location

A focus on co-presence, according to Beaulieu (2010), "...highlights the centrality of shared meaning achieved in and through interaction." (p. 5) Switching

[1] I refer here to anyone who has studied or written about their field itself, rather than about that field's content: those who wrote on the crisis of confidence in the 1970's, for instance. The word 'metascience' was not in use then to describe researchers of this ilk, however.

from a need for co-location requires one to reconsider what other elements of interactions can take the foreground in observation and analysis. One must observe interactions, and learn a new set of linguistic and communication rules. One is required to reestablish what cues are salient for analytic purposes, and which might be ignored. There are more than a few differences between in-person and online interactions, thus this is no easy task. As an example, consider turn-taking. In a physical group setting – a group of people conversing at a birthday party, for instance – there is a certain etiquette to be observed when one communicates with the rest of the group when it comes to who speaks when.

We learn to take turns in conversation with others at our mother's knee – small children frequently interrupt the speech of others, and, with the guidance of adults, usually learn the rules of turn-taking in time. An ethnographer studying a group of adults closely as they converse may consider breach of these rules by the participants salient. In person, if many adults in a group were talking at the same time, the resulting cacophony would render meaningful communication almost impossible[2]. If just one adult individual were talking out of turn, it would also be strange, but for different reasons.

Such turn-taking rules are somewhat moot in the case of Twitter. This is largely because interactions between actors do not need to (and often do and cannot) take place in real time. These kinds of interactions – what Beaulieu (2010) calls 'time-shifted interactions' – are a feature of 'conversations' on Twitter. A person located in Europe often cannot realistically converse with someone in Australia due to different time-zones, unless at least one person is awake at an unusual time *and* online. Others can also weigh in on the conversation without having to wait for other people to post on a thread – in fact, in a busy thread, it is often the case that many people are posting comments and responses to one another simultaneously. If one of the participants has not adequately learned turn-taking etiquette, it is unlikely that this social deficiency would be immediately obvious from a Twitter thread.[3]

Meeting the Challenge of Internet Ethnography

So, the ethnographer intending to use Twitter as 'the field', must learn to become sensitive to a whole new set of features. Beaulieu refers to a 'reinvention of ethnography' in its encounter with the novel object of the internet; of adaptation and the shaping of aspects of knowledge production to accommodate the new contexts presented by internet platforms (p. 142; 2004). Another challenge associated with virtual ethnography concerns how the researcher can achieve co-

[2]Of course, I am referring to a small group with this example. In a very large party, naturally many people are talking at once, and the group to be observed would instead be dyads or triads between members of the larger group.

[3]Such an individual might make *other*, more obvious social faux-pas on Twitter, which may be salient to observers, however.

presence. Beaulieu (2010) discusses her use of a mailing list during fieldwork as a means of achieving co-presence.

She makes her own posts on the mailing list a way to connect with academics in her chosen field of study. She also used the mailing list as a means of data collection, and describes what seems to be an interesting reflexive situation where her data source is also her interface with the field; her way to be there without being there. She also achieved co-presence through her online presence outside of the mailing group, and kept her identity transparent and consistent in all of her online activities, including her academic writing. While an ethnographer researching a physically organized group would achieve physical closeness, and relatability as another academic in the field (rather than being a distant, foreign entity), and maintain a consistent physical appearance, so too can the virtual ethnographer connect actively to their chosen online community, and maintain a consistent online presence.

Pink (2016) discusses a method of ethnographer John Postill, who at one time worked with a community he called 'freedom technologists'. He wrote up his field notes and blogged during the ethnography process, corresponding with his subjects throughout. His approach might be seen as a hyper-involved way of achieving co-presence; a participatory ethnography which sees the ethnographer collaborating with the target community. This approach opens new avenues for engaging participants and others in the research process.

I have used my online presence as a way of 'outing' myself (Finlay, 2002); of revealing my work and my role as an ethnographer online, and to achieve my own kind of co-presence. As I discuss in **Section 5.4**, I used my Twitter biography as a way of signposting my work, and to let would-be participants know that I was using my presence on Twitter as a research tool. I also used tweets and polls in a way to actively, publicly broadcast this information. For instance, in poll I posted in April 2021 concerning my quotation of people's tweets in my thesis, I drew people's attention to my use of Twitter data (see Figure 5.1). In engaging with potential research participants in this way, I was able to maintain my presence online, and in a sense collaborate with the reform group in determining how I could most ethically use their data.

Beaulieu (2010) emphasizes that the virtual ethnographer must be aware of the framing of her research and her online presence in the digital environment, and the effect this framing may have on the community under observation. Pink highlights the need for reflexivity, and maintaining a conscious handle on how we produce knowledge in the digital environment, and what the status and impact of that knowledge might be, academically or otherwise. In virtual ethnography, and especially in the case of using the highly involved and collaborative approach mentioned before, the ethnographer's awareness and reflexivity is not only paramount to the success of the project, but to respecting the community under observation. I discuss reflexivity and my stance shortly.

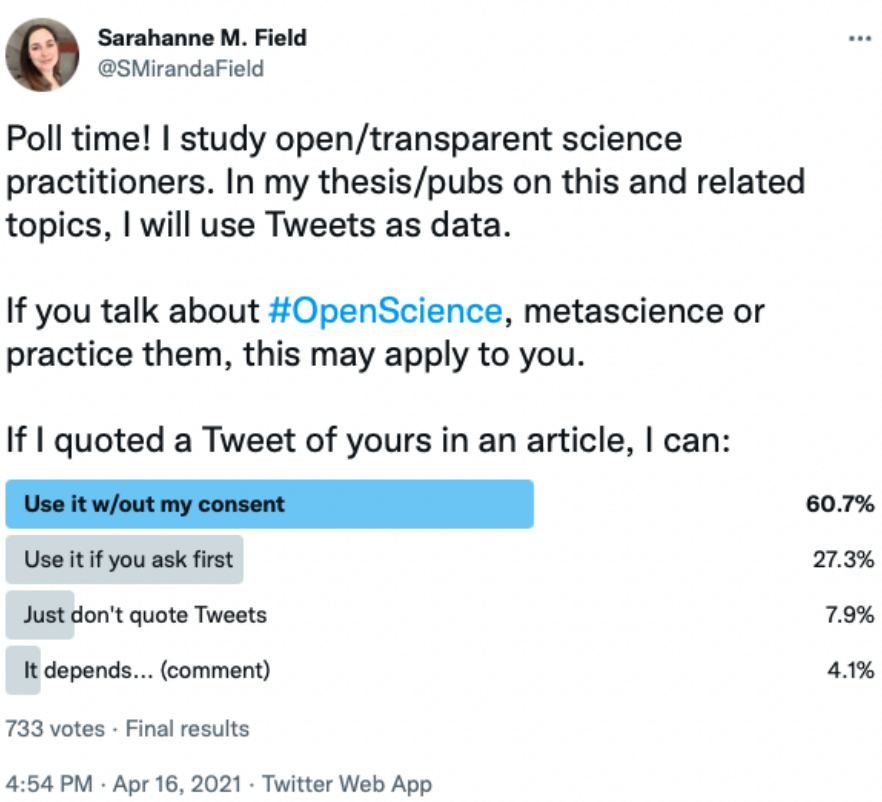


Figure 5.1: Screencapture of a poll I posted on Twitter on April 16, 2021, requesting opinions about how people would like their tweet data to be used.

5.2 Fieldwork

Fieldwork, writes Hastrup (1992), is situated somewhere "between autoethnography and anthropology", and plays the role of connecting an "important social experience with a general field of knowledge" (p. 116). It is an attempt, she says, to systematically collect information about the world of other people, though in my case, I used my fieldwork to acquire knowledge about a world of which I am a part. Though not unique, of course, it is rare enough for the ethnographer to study their own community. It brings with it complications and benefits, and demands much of my methodology. It is for this reason that this section is long and detailed. In my opinion, it is not sufficient for me to simply briefly describe my methods and move directly on to the findings, as I have read in other written reports of ethnographic studies. While I do not wish to burden readers with a stream-of-consciousness or blow-by-blow account of my activities and motivations, in this part of the dissertation, I am transparent and explicit about what I

have done and why, and about challenges that have arisen during my practice.

Observation

Inevitably, entering any setting begins with observation. Even when we board a train, enter a lecture hall, or a shopping center, we observe. We smell the coffee another passenger has brought into the train compartment with them, or hear the buzz of students talking with one another. We can feel as our skin prickles with goosebumps as we leave a warm summer day for the air-conditioned interior of a supermarket. Even if it is only superficially, or unintentional, we observe. Observation is natural, and yet I found that I had to train myself to do it properly. In the context of active observation it can be exhausting. For me that was certainly true. Perhaps, at least partly, because it involves constantly drawing from all the senses, as well as remembering to commit many things to memory. It seems to me that the word 'observation' is something of a misnomer, because observation often implies only the use of the eyes. Erlandson, Harris, Skipper, and Allen (1993) give an apt explanation of observation. Observation is somewhat like a written photograph, they say, whose details are constructed from all of information in a setting available to the senses. This strikes me as being a good description, except for one thing – while a photograph is an image recorded for you with a device, observation requires that you record all aspects of the 'photograph' yourself with your own 'equipment' (though, certainly, other devices such as an iPad can be used in the process). Importantly, for instance, someone conducting participant observation will need to remember environmental details, snippets of discussion during informal interviewing, who speaks to whom (and what verbal and non-verbal communication during that interaction takes place).

This recording process is easier when you are in a setting where you can naturally take notes (for example, during a conference session where other people are taking notes on the content of the talk, or, more likely, catching up on email), as opposed to in the context of networking or socializing with colleagues in a hotel lobby or bar, when sitting there with a laptop, typing furiously will look out of place or be next to impossible. Luckily, in comparison with most ethnographers, I have not had to juggle the practice of observation with building a rapport with the community I am studying: I am already known by many members to be part of the community and actively take part in events and discussions with them. In fact, Schensul, Schensul, and LeCompte (1999) discuss using observation as a means of establishing relationships and 'becoming known'. In my case, I have, in the words of Schensul and colleagues, used observation to: "get the feel for how things are organized and prioritized, how people interrelate, and what are the cultural parameters", and to "to show the researcher what the cultural members deem to be important in manners, leadership, politics, social interaction, and taboos" (p. 11), and ultimately to help me pinpoint the best topics to explore in greater depth with informants.

Field Notes

I kept field notes throughout the course of the ethnography, developing my modus operandi as I learnt. According to Hammersley and Atkinson (p. 141; 1995),"the making of field notes has been part of the invisible oral tradition of craft knowledge, and many who embark on their first project have to find their own way of doing things."

Beginning with the first meeting early in 2018 I attended with other young metascience researchers in Rotterdam, I took notes during periods where that was appropriate (for instance, during conferences where others had laptops out, or during workshops where content-related note-taking by others was taking place). I almost always typed on my laptop, and set the colour of the text in the word processor to light yellow so that it was difficult for others to read what I was typing. This was largely because I was sometimes writing observations about people and their behavior, and felt uncomfortable thinking that some people might take offense at what I had written. Some researchers argue that it is best to focus on active observation and record detail later (Fetterman, 2019), others recommend taking detailed field notes throughout the observation process (Kjerholt, Wagner, Delmar, Clemensen, & Lindhardt, 2014). I took a hybrid approach, because, as I explained, sometimes the context made note-taking easy and natural, while at other times, I did not want to disrupt the organic flow of discussions and socializing to take notes. In the latter kinds of cases, I recalled what I could, and took pictures and made sketches. Once or twice, in cases where I had a good rapport with a source, and an easy, natural mode of communication (such as Twitter's direct messaging app), I revisited conversations I had, and retained those chats to supplement my memory.

Although my in-person observation and discussion opportunities were a rich source of material for later analysis, I conducted a great deal of 'virtual ethnography', using Twitter and other online meetings as my 'field'. Virtual ethnography was especially important during the lockdowns of the coronavirus pandemic. The kind of field notes that I produced during my virtual ethnography were of a different nature to those I collected in in-person contexts. I collected many screenshots of interesting tweets and threads, discussed tweets and discussion threads in direct messages on Twitter, and 'back-channelled' with colleagues and friends during online workshops and conferences.[4] I typed copious notes also, to which the online format was conducive (you can often mute your microphone and turn off your camera if you wish, and type while listening and observing others relatively easily, especially in larger 'audiences' where muting and not sharing video

[4]Back-channelling in this context refers to the process of having a discussion with someone in private in parallel to another discussion or event. For instance, during breaks in between speakers at an online conference, I would instant message with friends discussing the content or other observations using a private messaging service like WhatsApp, Twitter direct-messaging or the Microsoft Teams private chat function.

feed is advantageous and even required at times).

I typed 'field notes' most days during my Twitter feed monitoring sessions, but these were often in the form of short memo-style notes, rather than three or more pages of description followed by reflections and initial analyses that the more intense in-person meetings lent themselves to. I often discussed my observations, notes and personal reflections with my daily supervisor Maarten, who helped me develop my ideas for analysis of the materials. This helped me to maintain reflexivity, which, as I discussed previously, was challenging given my participant observer stance.

Interview and Conversation

During my ethnography, I conducted nine formal interviews. Four of these were in person, three were virtual (i.e., via an online meeting platform like Zoom or Google Meets) and two were via email. Due to restrictions of the coronavirus pandemic and the fact that many interviewees were based in the US (with awkward time-zone differences), I gave interviewees the choice of the interview medium. Each of the in-person and virtual interviews were between forty minutes and an hour and ten minutes each, and I recorded the audio for these. Additionally, I conducted semi-formal discussions with some sources in busy settings, which I recorded with their permission so I could listen back and recall important details. I transcribed all the audio I obtained in the formal interviews, and the transcriptions (along with the notes I had written during the interviews) were kept for later analysis.

These interviews were nearly all conducted after I had analysed my observational material, and I selected the interviewees based on my analyses. I either requested interviews from people who had been central actors in important exchanges on Twitter (or elsewhere), or who had a position that I thought would be advantageous to helping me with my interpretation in some way. I primarily conducted interviews and discussions with people to help me interpret and clarify observations or explain anomalous findings. I asked them about tweets they had posted which had raised questions for me, or about tweets that had been posted by others that they might have seen and reacted to. Some 'incidents' and debates on Twitter were ubiquitous and impactful enough that I was able to ask people about them and get different perspectives on what had transpired, which helped me triangulate and validate, contextualise or even challenge my own interpretations.

Two interviews were with people who I knew to be on the periphery of open science and reform. One such source was a person who had contacted me to discuss the content of a paper they had drafted on the open science community. They said they were concerned about getting the paper through review because of their peripheral position and because the paper was critical of the movement. They said they felt they were an outsider (partly due to the field they were in, and

partly because they had felt unwelcome). The interviews, whether in person or via email were semi-structured, with a set of questions tailored to each different source. I followed up on some things they said too, and varied the order of questions depending on how the flow of the discussion unfolded.

I had countless informal discussions with sources too, particularly in the context of in-person fieldwork. As I know many of the participants personally (and many of those friendships and working relationships predated my ethnographic work), these discussions were easy and occurred organically. Where appropriate, I would ask them questions about their own observations and thoughts about interactions or issues I had noticed myself. As with the formal interviews, the sources played a valuable role in helping me evaluate the soundness of my own interpretations, or challenge some of the preconceptions and assumptions I had made. One such discussion, which took place over the course of three hours during a long walk in Berlin with a colleague and good friend, had a big impact on how I perceived some things in relation to incidents that had occurred in the community during my study. In cases such as that, I had to maintain an open mind and reflexive stance, in order to absorb the source's input and let it influence my thinking and interpretations.

I asked each source to state whether or not they wanted to be anonymous if and when I quoted them in my report. Most requested anonymity though three people were comfortable with attribution. When I quote sources who wished to be anonymous in **Chapter 6**, I refer to them by the letter *S* for Source, and the order in which they appear in the text (e.g., *S1* is the first anonymous source I quoted in the text). The informants comfortable with attribution are referred to by name. All formal interview participants signed a consent form.

Interviewing in Ethnography

Before I began my PhD research, I was conducting research focusing on reflexivity in the academic community and its link to higher quality research. Through that study (during conducting the interviews themselves, and interpreting the findings), I learned much about how important reflexivity in interviews was. I briefly review some important considerations of reflexive interviewing in both formal and informal approaches in this subsection, as they have heavily influenced my methods throughout the ethnographic process I followed, and the writing of this report on the research.

Like observation, participant interviewing is a commonly used tool in qualitative research. The approach can be structured or more loose, depending on the objectives of the researcher. In my work for this thesis, I have employed the semi-structured interview approach, and have conducted unstructured interviews on the fly during observation with community members. The latter was most fitting during my SIPS Rotterdam and Metascience Stanford conference visits. In the case of the latter, the conference schedule was very rigid (meaning that the time

of most participants would be valuable), and the duration of my stay was relatively short (I arrived the evening before the conference and left the same hour the conference ended).

The informal interview approach also allowed me a great deal of freedom to discuss issues with people as they arose during the conference. Of course, questions with a semi-structured approach can be modified as needed. By this, I mean that I am perhaps less likely to miss subtle meanings and interpretation of participant responses if they are outside of the expectations created by pre-specified questions (and the expected answers). This advantage is emphasized by Qu and Dumay (2011), who state that the interviewer can "remain open to new and unforeseen phenomenon rather than imposing ready-made frameworks or categories." Another benefit to informally interviewing participants during the course of the conference was that I could assimilate into the group and remain inconspicuous. Although I do not like the feeling that I am in some way deceiving the group (which I sometimes get), I do think it is important that I blend into the setting such that people act and speak most naturally. A final consideration is that the informal approach gives participants more space to speak freely and give information that is less carefully constructed. In turn, this can give me more insight into what the participant is really thinking and feeling about a certain topic (Bolderston, 2012). These kinds of informal, conversational interviews fall under what Qu and Dumay call the 'localist' approach (Qu & Dumay, 2011). This approach sees the interviewer as "people who are involved in the production of answers through complex interpersonal interaction", and the interviewee as "people who are not reporting external events but producing situated accounts" (p. 241). The accounts produced by the interaction between interviewer and informant are situated, and must be interpreted within the relevant social context.

My approach is in opposition to that of the neopositivist, according to Qu and Dumay (2011) – I am not interested in discovering an objective reality perceived by participants per se. Though I do see the interview as a means to collect research material, I don't see participants as truth-tellers or myself as a conduit through which 'data' are presented to an audience. That said, sometimes, a rigid structure in interviews is beneficial. For instance, some of the interviews I mentioned earlier were conducted with informants over Skype or Zoom during working hours. In such cases, I need to work under time constraints and therefore have to use the time I have wisely and prioritize asking certain questions and topics of discussion. For other interviews which were conducted via email, I had to deliver all clear and structured questions to participants in advance so that they could answer them via a reply email.

The habitus of both interviewer and interviewee should be taken into account, as our lived experiences, personal attributes and skills and knowledge impact how we perceive events and occurrences. Participants accounts and opinions are so embedded in their habitus that objective truth can rarely exist. Interactions are further complicated by the fact that social capital is deeply dependent on the

context of a given social setting, at least in the view of Bourdieu (Bourdieu, 2013). My interpretations of verbal texts are also heavily rooted in my own habitus, such that I am not able to be a clean conduit for any objective truth if it did exist.

I took notes during each interview and after as many informal interactions as my schedule permitted, to record my reactions, impressions and thoughts, and referred back to them when analyzing the findings to ensure that my interpretations were based on all the information I had.

Interviewing and the Limits of Reflexive Sociology There is much debate as to the merit of the interview as a communicative event, regardless of the context in which it is used (Briggs, 1986). Though the interview is fraught with inherent issues which may threaten validity and reliability of results, such as bias, distortion and expectation, it can be nevertheless used as a valuable vehicle through which one might understand a given phenomenon. Bourdieu's piece, Understanding (1996) provides a unique reflection on the use of the interview as a sociological research tool. He highlights the limitations of the analysis of his own interview data, identifying that it is impossible to transcribe interviews into readable dialogue without imposing personal interpretations. Though many influences on Bourdieu may be identified within his writing (Durkheim and Weber are two well-known examples), Understanding plainly showcases the role of Marx in shaping Bourdieu's thinking. Like Marx, Bourdieu saw elements of sociological practise as an opportunity to implement change in society, in the way a 'midwife' assists birth, as the facilitator and aid to a important and difficult process.

It is through Bourdieu's insistence on reflexivity in qualitative research that he offers a betterment of the practice of the social sciences, which can be easily applied to the context of the ethnographic approach I describe here. Maton (2003) provides a positive comment on the role of reflexive practice in sociological research, finding that it lends itself to richer descriptions of social phenomena, and reinforces a more 'practically adequate and... secure social science' (p. 53). Byoung-Kyo (2010) disagrees, arguing instead that reflexivity may threaten the reliability of research data, and the artefacts of self-awareness may somehow contaminate the information interviews can yield. To make a link to **Chapter 3**, this is a key argument of many proponents of replication and reproducibility, that subjectivity and personal bias are the enemy of what should be a 'clean' and objective process (Field & Derksen, 2021).

McRobbie is also critical of Bourdieu's reflexive approach to interviewing, questioning the value of the interview in implementing change though reflexivity. She argues that Bourdieu, in his 'self-congratulatory' piece neglects to discuss the data in the frame of their cultural context. She suggests that Bourdieu is overly reliant on the 'decontextualized voices' of his subjects, and asserts that without cultural and social contextual anchors, these voices are merely personal opinion and '... exist merely as the stated truths of personal experience' (p. 131; 2002).

She pushes her critique, arguing that Bourdieu's 'voices of pain' would be better legitimised and would provide the insight that Bourdieu claims they do, had Bourdieu given "... a brief glance in the direction of cultural studies" (p. 132), and learnt from how those scholars explore peoples within their historical and social contexts.

Despite criticisms like McRobbie's, Bourdieu's reflexive use of sociological practice as explored in Understanding (1996) provides a solid foundation for the use of reflexive methods in social psychological research. One key reason for this, in my opinion, is that reflexivity represents a concrete and active *research practice*, which can be easily ported between scientific disciplines and used without the need to familiarize oneself with cultural studies. For my own methods in this research project, this concreteness and portability is valuable. Additionally, both Bourdieu's Understanding and McRobbie's critique of it (1996; 2002) help to explore both the utility *and* limitations of the interview as a qualitative methodology, and highlight potential pitfalls for the use of reflexivity in practice.

Documents

Documents and websites can be an invaluable source of information about subcommunities and community members, because they are artifacts – reifications of community ideals and ideas for community behavior and activity. Take the example of codes of conduct of SIPS. I downloaded the code of conduct in 2018 and in 2021. They were different documents, containing different content, which reflects important changes in what kinds of conduct are tolerated and how behaviours are defined. These documents reify the values of the SIPS organization. I saved copies of codes of conduct for organisations such as SIPS, as well as copies of the Metascience Symposium's conference website (again, interesting changes in the text and how it is emphasised gave me much material for later interpretation). I saved other relevant websites and documents relating to, among others, the INOSC organisation (International Network of Open Science and Scholarship Communities), the Center for Open Science, the Melbourne Open Research Network, and the Open Science Communities (OSCs). Finally, I saved many blog posts from different members of the group, when they related to topics relevant to the study.

Research Material Analysis

Every piece of research material that I describe in this chapter, I saved into a project in ATLAS.ti (a software suite built to support the documentation and analysis of qualitative research; I used version 22 for Mac OS). I conducted a qualitative thematic analysis on all of the materials. This process was based on Mayring's (2014) steps for inductive category development, which involves the inductive development of themes from the research material and the revision of

those categories. I reviewed all of the research material thrice (which took some time because of the volume of the material – the field notes alone were hundreds of pages long), recording only analytical memos when something struck me as particularly salient.

After I felt I was familiar with the material, I began coding. I assigned codes to small chunks of text, or 'quotations' (the unit of analysis for the textual material was 1-2 sentences usually, which reflects the structure of tweets and small paragraphs of website text). Sometimes the same section of text was assigned to more than one code. I went through the material twice (at two different time points) for the purposes of coding, to ensure that I was as internally consistent as possible, and at the point of thematic saturation (i.e., where the repeated analysis of the findings produce no new codes), I established a thematic framework (Attride-Sterling, 2001). I reviewed each of the transcripts, documents and other materials, as well as the codes twice after initially establishing the thematic framework to maximize internal consistency (a process I found worked well for me during a previous qualitative research project). This process yielded 35 salient codes, which I grouped into 13 major themes. There were several other 'loose' codes which, though they were potentially interesting enough to code, did not fit into major themes.[5]

Each of these major themes had several codes assigned (from the pool of 35 I discussed earlier). Some codes within these major themes were salient enough that I 'promoted' them to sub-themes, and linked them to other codes which related to similar observations. For instance, the sub-theme *BAD BEHAVIOR: Repels potential open science adopters* is one of these. I assigned this code to 12 different quotes within the material, and it was salient in the light of other analyses relating to the negative effects of 'bad behavior', which led me to want to spotlight it as a salient part of the analysis within the *conflict and power* theme. This stage in the analysis – re-visiting the codes and their structure and looking for links, differences, similarities and explanations across codes, and seeing how they fit within the broader 'shape' of the analysis – is all a part of processing qualitative research material. The more the material and analysis structure is revised, the better an overview the analyst has over the material and the patterns contained within. From this, the analyses will be more rich and complex, which is usually good. The qualitative research analysis process has been described as iterative to reflect the repetition of revisiting the material and codes I describe here (Srivastava & Hopwood, 2009).

[5]Some codes did not get grouped because they were too specific and were assigned to one-off examples in the material. For instance, 'survivorship bias' was mentioned twice in the material, but ultimately it did not relate sufficiently strongly to other established codes or come up in enough parts of the material to warrant being grouped as part of a major theme. I coded it in the initial stage of the process nonetheless, because at the outset I did not know whether it would contribute to a pattern of interest or not. This is often the case with inductive coding, and why qualitative researchers tend to go over codes and revise coding structures more than once.

During the analysis of my ethnographic materials, I read Wenger's book, *Communities of Practice: Learning, meaning, and identity* (1998). In doing so, I realized how well many of the themes I took from the material mapped onto Wenger's framework, which I will discuss in the next chapter. From that point on, I used his framework to interpret my findings, and mapped out many of the themes onto the concepts Wenger describes in his framework. The most dominant themes in my analyses are presented in **Chapter 7**, and many of the themes (or groups of them) have been named to coincide with elements of the community of practice framework where appropriate.

After conducting the initial thematic analysis, I grouped themes which focused on similar issues into what I refer to as major, or dominant themes. I re-assessed the material relevant to these themes, checking for interesting patterns or anomalies (things I might not expect, for example, which can lead to interesting follow-up) and verifying that such grouping made sense. In this way, I constructed 13 major themes: *community description and identity, participation and reification, conflict and power, heterogeneity in the community, community boundaries, SIPS (relating to both the society and the conferences), activism, terminology, funding, metascience (relating largely to events and topics arising from the symposium series), Twitter/online presence, Curate Science Leader board, and Code of Conduct Violation Case*.

I have focused on the first five of these themes for two reasons. First, I explored them in my analyses because they were by far the most dominant and rich. For instance, although the latter two groups seem interesting, only one or two sub-themes were assigned to them each, and it felt difficult to adequately interpret them with so little 'data'. Compare this to the community boundaries grouping, which contained 10 sub-themes. Second, I focused on these major themes because they helped explore my research questions and aims most effectively. Funding is an interesting topic related to my study aims and questions, however it is not nearly as central as other topics were. Moreover, with this theme, much of the material I coded for it was based on my own reflections and supposition (written as part of my field notes), which felt uncomfortably more like guesswork than an actual, legitimate, source-validated finding (which the other themes were, in my opinion).

The sections in **Chapter 7** reflect the five major themes I describe above, while sub-sections and paragraph-level headings reflect sub-themes and elements of interpretation relating to Wenger's theoretical CoP framework. I provide a more detailed mapping of the themes and the interpretation early in **Chapter 7**.[6]

[6]Please note that this explanation of my method and the themes derived from it might give the impression that the thematic analysis process was more structured or formal than in it was in reality. I have described the process in somewhat formal terms to explain it to people less familiar with qualitative research methods, but the artistic element involved in my method should not be understated. As many people do, I used a thematic analysis approach as a way of becoming familiar with my research materials, and to be able to impart structure to patterns I interpreted in

5.3 Artifacts, Challenges and Ethics

As with any research process, the ethnographer has to be aware of potential methodological issues. The concepts of validity and reliability are just as important in qualitative as in quantitative research, though they are different in practice (LeCompte & Goetz, 1982). Bias, when it is not brought to the fore and explored through reflexivity and reflections on positionality can undermine the validity of an ethnographic study, for instance. Not recognizing and combating methodological concerns like observer effect (which is observable when a "respondent's sensitivity or responsiveness to a measure is affected by the process of observation"; Singleton & Straits, 2010) and habituation (the process whereby participants become accustomed to the observer's presence, and which, when complete, reduces observer effect) can cause research artifacts that interfere with the reliability of an ethnographer's conclusions. Ethical considerations, though not a direct or obvious threat to the quality and authenticity of research findings, are an equally high priority. People trust that the information they share with an interviewer will be kept safe and treated with respect, and it is of the utmost importance that ethnographers do not break this trust.

Habituation and Observer Effect

Van Maanen (2011) says that in a sense, written reports of ethnography "sit between two worlds or systems of meaning" (p. 4), that is, between the worlds of the ethnographer themselves, and the world of the participants. For some ethnographers, when they study cultures and peoples that are very different from their own, the distance between themselves and the object of their study is great. In contrast, the distance between myself and the reform community does not exist per se, because I *am* an active and known part of that community. I think it is likely that any distance that potentially could open up would be artificial – of my making (though of course it is conceivable that some in the community had distanced themselves from me without my knowledge). I tried to generate such a distance at times throughout the ethnographic activity. Sometimes I would need to step outside of the role of a reform activist and metascience researcher (i.e., where I had to acknowledge my emic view and consciously adopt an etic, or 'outsider' view) to see an issue or interpretation in a different light, or to become more open and receptive to a challenge or contrasting view of another person. Sometimes I did it in order to further explore an inconsistency or contradiction, and it

the findings. I have chosen to highlight certain aspects of the patterns that I found important or meaningful in the context of my experience during my ethnography. Because of this, I don't discuss all sub-themes in **Chapter 7**, nor do all codes in sub-themes feature in the analyses. In some cases too, there are examples or codes which could reasonably apply to more than one angle of analysis. Again, the choices I have made in such cases are subjective and have served the interpretation that I felt best captured my ethnographic material. The inherent subjectivity of this process brings with it a certain artistry or creativity which, for the sake of transparency, I feel compelled to emphasize.

was at these times that some of the most interesting tensions and irregularities of the research material became apparent.

Agar (1982) discusses these inconsistencies or irregularities that the ethnographer picks up on during the course of observation, calling the phenomenon 'ethnographic breakdown' (a word which he says he borrows from Heidegger: p. 783, *ibid*). The 'disjunction' between the world of the ethnographer and the world of the subject (i.e., when the ethnographer does not have a system or framework in place to make sense of what is being observed; when something does not make sense, or expectations are not met; p. 783, *ibid*) is what drives the process of discovery in ethnography, and what also causes understanding or coherence to occur. In order to be sensitive to possible disjunctions of this nature, and, in turn, to achieve coherence, I had to sometimes play the role of the outsider.

That there existed little or no distance between me and the community was complex in other ways, too. It meant that I had to frequently and, sometimes aggressively, confront my own agenda and feelings on certain issues, and attenuate some of my responses to what sources told me during interviews and discussions. Luckily, reflexivity is a focus of mine, and I have been studying and practicing it for some time. I do not mean to say it was always easy to keep abreast of myself and my own influences on the research process, simply that reflexivity is a valuable tool that I had already been familiar with before I required it for the current study.

I have been so far describing the challenges of being a participant observer, but there were great boons to this unique stance that impact upon this study. For one thing, I did not have to habituate (at least, not to the extent that most other ethnographers would have to). I did not have any problems integrating with the community I studied, nor were there any awkward introductions or situations I had to navigate as a novice or newcomer to the setting. I fit in, and I 'looked like' many of the individuals I wished to observe. I was just another academic attending a conference, workshop or networking event, with a lanyard around my neck, and a laptop open in front of me. This made it easy to conduct my fieldwork without needing to navigate the difficulties of a strange context, and meant that when an interesting incident or debate took place, I was included in the gossip. Oftentimes, when I asked a source to share their perspective on something, it was, on the face of it, a chat between friends or peers.

There was also the benefit of decreasing observer effect, when people know they are being watched and act differently, or 'perform' (Monahan & Fisher, 2010) for the observer's benefit. Although I was open and transparent about the research I was doing, most of my colleagues did not know what kind of research I was doing. Some did, however. That I was already 'one of them' largely decreased the possibility of me being intrusive or affecting behavior by involving myself in potentially sensitive discussions. Naturally, those who were aware of my research might have been acting differently around me, but I did not get the sense that this was the case. Moreover, even if it did occur, I judge that it would have

been minimal, given that I observed different sub-groups at varying points and in different contexts.

One final benefit of being a part of the community I was studying, is that the differences in status between myself and my sources is minimized (though hardly diminished, as I discussed above). If anything, due to my junior status, it would have been that my sources held a higher position or greater power than myself (rather than the other way around, which is more problematic in this context). Bourdieu coined the non-physical kind of violence that can arise from power differentials between people and groups 'symbolic violence'. He used this concept to discuss how inequalities in social class are reproduced (Bourdieu, 1990), however the concept can be reasonably ported to highlight how inequalities in status of other kinds in addition to social class (for this factor plays a strong role in some institutes in academia even today) can affect the development of a healthy and fruitful relationship between an observer or interviewer, and their subject (although, in saying this I risk being judged as failing to "appreciate the complexity of Bourdieu's analysis"; p. 16: Connolly & Healy, 2004). Symbolic violence can undermine mutuality and a feeling of rapport and trust between an ethnographer and their informants, for instance. It can prevent the establishment of mutual understanding and co-production of knowledge between the two parties, which in effect prevents successful ethnography, and can be difficult to resolve.

The potential challenges presented by habituation and observer effect in in-person fieldwork were less relevant when it came to the virtual fieldwork I conducted. Although I did need to get a feel for the community, to find my way 'around' it, I could easily scroll Twitter at any time of day or night, watching threads unfold in real time as their authors published posts in reply to one another without needing to engage with anyone, or try to observe without being obtrusive or conspicuous. During virtual meetings, I could focus on the non-verbal cues of attendees when they had their cameras on, and watch their responses to speakers and discussions between panelists. I could watch and analyse chat text during live conferences, and see who was saying what to whom. In that way, I did not have to consider how I was presenting myself as is the case with physical fieldwork.

5.4 Feminist Influences, and Myself as Part of the Process

As I have navigated my way through the research I present in this study, I have had to make many decisions with regard to how I conduct myself and what I think. I have had to learn about myself and how I see the world, taking this into account as I observe my subject, and interpret the data I have co-produced with the community. As an observer of the community in which I am also a member, this has been a crucial part of producing valid and high-quality research. I have

been guided by methodological frameworks and ethics (which I explore in this section and in **Section 4.5**, respectively).

As Pranee Liamputtong prescribes in her extensive chapter on the topic, qualitative inquiry tends to be situated within methodological frameworks which help determine how researchers will produce material to analyse and interpret (2019). The research I report on in this dissertation has been influenced by feminist scholarship, and I discuss this influence in this section.

Feminist thought, with its emphasis on emotion, subjectivity and reflexivity, has heavily influenced qualitative research methods. Clarke and Braun explore this influence thoroughly in their (2019) review of feminist methods in psychology. They discuss how feminist methodologists advocate the legitimacy and value of methods that examine the uniqueness, experiences and subjective views of the individual in the group under study.

Feminist thought has politicized the research process, rejecting the notion of objectivity, and underlining the inherently political and ethical nature of research. In all of this, the *process* of research is heavily emphasized, and held as being just as important as the outcome of the research. A feminist methodology, argues Jo Moran-Ellis (1996), allows researchers to include their own experiences (as women and as scientists), not only in the process of conducting the research, but also in the analysis of the findings, and the written report of the whole inquiry.

Feminists have examined the innately reflexive nature of the research process, and focus on the ways in which the researcher and the researched interact and integrate with one another in the research process. Feminist thinking has underscored the importance of reflexivity in research, holding that the researcher's articulation of self must be public to ensure the veracity and integrity of the research (Liamputtong, 2007). Flexibility in methodology, too, is a hallmark of the feminist influence on research methods. Mixed methodology and the use of photo, video and drawings, for instance, can be combined with more traditional methods like interviews and surveys.

Feminist methodology also emphasizes the plurality of the human experience (Mann & Keller, 2013). What constitutes reality and truth is situated within the meanings construct according to how they experience life, and perceive their existence (Grbich, 2004). Seeking to represent the many voices of the people under study, feminists focus on details and personal stories (Dow & Condit, 2005).

The emergence of alternative ways of conducting qualitative inquiry are examples of how feminism has influenced its methodology. Online, and arts-based research, as well as autoethnography are some such approaches (Reavey, 2012).

Feminism in My Practice

My own research methodology is clearly marked by the influences I describe here. My own process emphasizes reflexivity, and embraces the subjective. I reject the notion that I should (or even could) separate myself from the research process I

am conducting. My findings reflect a mix of qualitative and quantitative methods, and my field notes include photography and sketches. I conduct part of my research online, and incorporate autoethnography throughout my research process. This dissertation foregrounds the research process much more than other ethnographic studies tend to, and I readily incorporate reflections throughout the text of this dissertation. I have not consciously drawn upon these frameworks, or explicitly used them to determine my methods (as Liamputtong would recommend; 2019). Rather, they are formalizations of aspects of my personality and research modus operandi. In discussing the existing frameworks here, I can describe elements of my research process and what has guided it in a way that other researchers can understand.

Reflexivity

Reflexivity is the subject of a 2021 publication in the European Journal of Philosophy of Science which I co-authored with Maarten Derksen (Field & Derksen, 2021). The following text fragments are taken from pages 5 and 6 in the original article.

> Research into human beings – their behavior, emotions and cognition – brings with it the issue of psychological reality and the observer being somewhat enmeshed. Though an attempt at divorcing the two will be unsuccessful, awareness of the condition through reflexivity is possible. Reflexivity is the process by which the researcher continually and explicitly engages in self-awareness and analysis of personal influences on the research process (Finlay, 2002). Reflexivity on the part of the researcher allows them to question and adapt their interpretations, based on issues that arise during the study. It lends credibility and realism to the conclusions reported (Clancy, 2013): through reflexivity, a researcher is able to produce a faithful account of the research (Hertz, 1997). Pillow (2003) describes reflexivity as a way to "legitimize, validate, and question research practices and representations", and to call into question our data and methods. It assists us in understanding the social world, and also provides insight into how that knowledge is constructed.
>
> Although the unintentional incorporation of subjectivity into the research process can be seen as a barrier to good scientific practice, reflexivity has the potential to facilitate and even enhance researcher objectivity. Finlay (2002) presents it as an 'opportunity' rather than a 'problem'; that is, subjectivity can be used productively. It can function as a mode through which the researcher can become more aware of their influence on the research. It can provide a means to treat the influence accordingly, whether that means to isolate it, if the quality of the research is being threatened, or to work it into the research:

> reflexive practice can impart a richness to the data and their interpretation which is valuable in furnishing our understanding of human thought and behavior processes: the goal of psychological inquiry (Wacquant & Bourdieu, 1992). For those researchers who consider their subjectivity a threat to the quality of their research, using qualitative methods involving reflexivity may be worthy of consideration. Almost as a remedy to the potential problem subjectivity can cause, Crang and Cook (2007) recommend for researchers to recognize their 'partial and situated' subjectivity. They recommend to 'tap into' it as a resource; to use it for achieving a deeper understanding of the phenomena under study.
>
> Those who engage in qualitative research typically consider the researcher as the linchpin of the entire research process, and create and maintain awareness of their presence in the research process. Reflexivity gives this presence a practical application, and it is woven into the fabric of the study and can be included in the written delivery of the conclusions to the scientific community. To practice reflexivity, the researcher attempts to keep one eye trained on himself, so to speak: he remains aware of his own feelings, thoughts and expectations as he engages in the process of research. In the eyes of some, when employed in this way, reflexive practice can ultimately serve to impart objectivity to the research. In practice, this requires researchers to critically assess their role as the measurement instrument (Lincoln & Denzin, 2003), as well as the person conducting the study, and attempt to understand how their dual-function can impact their conclusions.

In the early stages of my time as a PhD researcher, I had sketched out a plan for my dissertation. 'Reflexivity' had been jotted down as a discussion point in the Methodology chapter of the thesis plan. This is not odd – reflexivity is certainly discussed as if it is a research method. As I have progressed through my study trajectory, however, I have come to realise that reflexivity is more than a method; indeed, it goes beyond practice. I have come to see it as more than a modus operandi, it is, for me, a way of being. So to discuss it in this section without emphasizing its importance in my work would be to sell it short. As I have conducted my research, reflexivity has gained such momentum in my practice that everything I do is imbued with it.

Myself as Researcher and Participant

As I discussed briefly earlier, metascientific research is, by definition, highly reflexive. In my case, I am a reform activist and metascience researcher who has studied the reform community. This interesting position has required deep reflex-

ivity. A concrete part of that reflexive practice has been coming to terms with my stance, and figuring out what the identity of 'participant observer' means for me, and how I navigate it during my ethnography.[7] I explore this now.

My stance

In an article on participant observation, Kawulich (2005) describes the different stances the ethnographer may take when conducting observation. She states that the type and amount of data collected by the observer is partly a function of the degree to which the observer participates in the community under investigation. She summarizes Gold's (1958) different participation roles: the complete participant (who is a member of the community under study, whose researcher status is completely concealed from the community), the participant observer (who is a member of the community under study, whose researcher status is known to the community), the observer as participant (referring to when the researcher is a participant for purposes of data collection, without them being a true member of the community) and the complete observer (who is completely hidden from the community during observation – think wildlife biologist who hides in a realistic man-made hideout in the shape of a hippo to observe animals on the Serengeti).

Above, I mentioned that my own stance falls into the 'participant as observer' category, as I am a member of the science reform/open science/metascience community. People know who I am, through the different roles I play in the community. I am a researcher in the field of metascience, and have been conducting research in that capacity since 2014 when I began conducting my first published study. People in my field cite my work, and engage with it informally. I am invited to speak at conferences and contribute to workshops, and have presented my work in other capacities too. I am a reviewer for, at this point, six journals which handle research that is adjacent and complimentary to my own. I am an editor for two different journals, and I handle articles relating to metascience, open science and reform topics. I have friends and colleagues in metascience, open science and reform groups.

At the same time, I am studying these groups of people. I discuss questions central to my project with them, and observe them as they engage with one another in person and online. A number of them are aware of what I am doing, and I am open to discussing my ethnography with anyone who has asked me. My Twitter biography, during the time I was conducting this research, clearly described my project and disclosed that if people were discussing open science or reform topics, I was observing their exchanges. My bio read: "*PhD researcher*

[7] An additional layer of complexity is introduced by the community itself. They are reflexive in their own, somewhat objectifying way, in that they make a study of the study of their own research process (and those of their scientific peers). They practice, in my view, a kind of collective reflexivity in pursuing solutions to problems within science.

conducting ethnographic research on the reform and open science community. Open science, always. Tweets = Data. OSF: osf.io/fv5nj".

In her review of the literature on stances, Kawulich (2005) repeats an amusing statement from Merriam (1988), who calls the participant observer stance as being schizophrenic, because of the complex nature of balancing participation and objectivity. In my experience, the participant observer role requires a lot of reflexivity, to maintain this balance. Merriam also points out that the stance is difficult because one must participate, then write notes after the activity.

No amount of reflexivity can 'fix' this aspect of practicing the stance – in my experience so far, it is a matter of getting used to multitasking (which leads to developing proficiency with observing and participating at the same time, which is only really possible when you participate passively), training yourself to remember more detail than you otherwise would, making time to write detailed notes as soon as possible, and taking photos where appropriate. This laundry list of points implies that I think myself a professional at maintaining the stance and collecting good data at the same time. This is not the case at all, and I am constantly smacking myself in the forehead trying to recall 'aha' moments and interesting interactions later on during note-taking, or debriefing with my supervisor.

Myself as the data collection tool

Practically every textbook on qualitative research methods emphasises how crucial it is for the ethnographer to be aware of how their personal attributes position them in relation to their subjects. I have worked hard to become and remain acutely aware of how my gender (and sexuality), race and class, among other more subtle attributes, affect the way I observe and interpret the resulting data. As Musante-DeWalt and DeWalt (2010) highlight, for instance, men and women have access to different information during fieldwork, simply by virtue of their gender. The same, naturally, goes for other individual attributes. I believe myself to be of a similar societal class to many of the actors in the community I study (working in academia as a grad student, and attending a symposium such as this is not highly exclusive, though it is certainly a privilege that many do not enjoy), though not all. I am white, and come from a privileged nation (I am an Australian national, but I live and work in the Netherlands).

I maintain awareness of these facts about myself, because of the risk of symbolic violence that may exist between myself and actors I might be engaging with in the field. Although practicing self-awareness and reflexivity does not remove the threat, it should attenuate it, and hopefully will increase the negative effects bias can have (when one is not aware of bias, for instance). Additionally, I am aware that many of the people attending these conferences are selected because of access to funding. European actors participating in the symposium (myself included) are in privileged positions because they can fund a visit to the United

States. Others from Asia and Africa, Canada and Americans who live in the most distant states also can attend due to access to funding for such trips.

5.5 Ethics

Just as my research process required reflexivity and an appreciation of the role of subjectivity, so does it require keeping a good handle on ethics, especially as it relates to the internet and Twitter data. This is not necessarily as straightforward as one might expect, because of the nature of this research (i.e., research that is conducted on content posted online on a publicly available social media website). Social media, say Tinati and colleagues (2014), provides the opportunity for exploring "new social relations that are oriented around digital subjects and objects." (in p. 1151 of Williams, Burnap & Sloan; 2017) While this is a good thing for future research possibilities, it implies that an entire new avenue opens up for complicated issues for research ethics, resulting in headaches for researchers studying social media content and communities.

For one thing, there are no agreed upon norms for this kind of research (Fiesler & Proferes, 2018). Situational ethics (i.e., ethics which are flexible depending on the research context) are usually adopted (M. L. Williams et al., 2017). Previously established guidelines apply to in-person data, or data that are acquired via the 'usual' in-lab methods, and a quick look through the guidelines in place in different countries of the world will tell you that they are outdated to adequately steer ethical research through the quagmire of handling internet-harvested data. As Williams and colleagues say it, the "digital revolution has outpaced parallel developments" (p. 1150) in good ethical practice. Other problems with settling on good ethical guidelines involve philosophical questions, like 'what researcher behavior is most ethical?', and how does one determine what is public and what is private when it comes to text that people publicly post on Twitter?

The virtual researcher has to consider how to balance the hard legal conditions Twitter users implicitly adhere to once they've agreed to the site's Terms of Service (ToS; Fiesler & Proferes, 2018) and data protection laws, with how to conduct their research with respect to participants' expectations and privacy. They have to conduct risk/benefit analyses at different points of the research process, determining what will constitute harmful data use and what is useful for their academic exploration. They must navigate suggestions on whether or not to directly quote Tweets, and whether or not to include identifying data with these quotes. Do they disguise participants' identities as Bruckman (2005) suggests, or should they be as transparent as possible, following Twitter's Broadcast Guidelines which state explicitly that researchers should not obscure any identifying user information, and to include user names and handles alongside Tweet content? If they are prepared to seek consent from users before handling their data

for research purposes, they then have to think about what that 'consent' would look like, given that asking permission and getting informed consent are not the same thing (i.e., people can only give informed consent if they have actually been informed). They should probably take into account that people will be sensitive about how their data are interpreted in the context of academic research, and that some participants may expect compensation for the use of their data.

These are problems that the researcher grapples with, but the fact remains that people who post content online using public social media platforms are probably more vulnerable than they think. The average user does not know how Twitter handles their data. They rarely read ToS documentation carefully, if at all, and they do not realise that their consent is not required by researchers before their content is used in scholarly research. The average user may not realise that their actual audience differs from their imagined audience, and that their Tweets are open for anyone to see and can be easily found via search engines like Google. Users are rarely involved in the development of ethical guidelines that directly affect the use of their data, and their consent is rarely sought by researchers, which makes them vulnerable as research subjects (Proferes, 2017). I avoid using the term 'participants' here, because it is my view that if consent to use their data is not sought by researchers, they are not willing participants. Indeed, according to Fiesler and Proferes (2018), users believe in privacy achieved by obscurity. The problem for users is that research has the potential to disrupt this, and they might not even be aware of that.

My Ethical Guidelines

The reader has probably, by now, started to get a sense of my stance with relation to my use of Twitter data in my research. My sentiments and practice are partly guided by suggestions in the literature on this matter, partly by the expectations and thoughts of the community I study, and, unsurprisingly, partly by my own moral compass. My perusal of the literature has led me to understand what the lay-user thinks about consent and the use of their data, what is 'usually done', and what might be reasonable and defensible practice (the text in the previous few paragraphs sums up most of this). My probing of members of the community under my study via the Twitter poll I discussed earlier in this chapter (April 16, 2021, which yielded 733 responses) and subsequent thread, has revealed many important considerations. Firstly, it has revealed that although most people (around 60%) don't expect to be asked for consent before I quote their Tweets in my research, nearly 28% would like to be asked, and about 8% of them don't think I should quote them at all.

From my community, I have learned that people like to be involved in the discussion and are eager to help guide me toward good ethical practice (e.g., by providing resources for me to use). They think that, although it might not be explicitly required, it is 'polite', 'kind' and 'decent' to ask permission before quoting

Tweets. They posed important questions for me, prompting me to consider how to handle deleted Tweets and content from verified user accounts. They brought up issues of how to handle sensitive data, whether or not Tweet content should be associated with usernames/handles, and questioned whether Tweets are someone's intellectual property or not. Someone suggested I should cite Tweets like I would a scholarly article, and another pointed out that issues only really arise when interpretation of Tweets is involved (users may not agree with the position that is projected onto them). Indeed, concerns about misinterpretation of Tweets were foremost for many people who explicitly commented on the thread.

The poll and thread I opened on Twitter, aimed explicitly at open science and metascience practitioners, gave me pause. These people, largely researchers, or involved in science policy and communication, should not be the average user. Instead, I would expect them to be more informed on privacy laws surrounding personal data, and be somewhat aware of common data handling practices. I expected them to be well-informed, but I think I underestimated the extent to which they are still vulnerable, simply by virtue of being online content-creators. The internet, and, by extension, social media sites, blur the lines between the professional and the private (though this issue is hardly new, or only relevant in this context: as Hammersley and Atkinson put it, "what is public and what is private is rarely clear-cut", p. 267; 1995), and academics using online media are both professional and private persons. They have home lives, pets, houseplants and children. They bake bread, brew beer and cook for their families and friends. They variously possess vulnerabilities relating to demographics and personal identities (which may affect how they expect their data to be used in research, according to the findings of Williams, Burnap and Sloan, 2017).

Indeed, the context of the pandemic has put the issue of private versus public into even sharper relief; the lines of home and work blurred even more than they were before. Online meetings and conferences open up conduits between the different spheres in which researchers exist, and people may find it difficult to maintain previously-held boundaries. For instance, full professors are seen in Zoom calls in hoodies and sweatshirts, and with toddlers or cats on their laps. The low-level mundane chatter that makes up much of Twitter content in non-pandemic times continues as well, adding to the perspectives on real life that the pandemic has prompted. It is especially important during difficult periods as it grounds people and reminds them that life continues. Prominent members of the community share humorous and mundane tweets as much as lesser-known members and grad students.

My poll and accompanying thread has underscored the humanity of the group I have studied. In the thread, members of my community posted, asking me explicitly and politely to respect their privacy, and asserting their right to faithful interpretation. They used emojis and 'please' and 'thank you' to show their vulnerability as people (in contrast with their status as career scientists and professionals), and in doing so gently and unwittingly remind me of the enormous

responsibility I have as an ethnographer and custodian of their Twitter data. My moral compass was already guiding me toward a protocol which leaned heavily toward respecting people's privacy and right to be forgotten, but interaction with my community cemented how I want to approach the issue of ethics in my dissertation and any publications involving Twitter data.

I will follow my own ethical guidelines, which integrate expectations and consensus from my community with legal requirements. These guidelines follow the general tenor of the Belmont Report on ethical guidelines for the protection of human research participants (1979), whose basic ethical principals concern *respect for persons*, *beneficence* and *justice*. I will act with respect for my participants, with a view to protecting their well being, and commit to regularly making risk/benefit analyses to minimize any potential for harm for the community and its individual members. I outline my ethical guidelines in the form of a sort of pledge to my participants, as follows:

- As often as possible, I will directly quote Tweets verbatim and in their entirety (in keeping with Twitter's Broadcast Guidelines: `https://about.twitter.com/company/broadcast`, but counter to the suggestions of e.g., Fiesler and Proferes, 2017).
- Whenever possible, I will include details identifying the owner of the Tweet alongside the content of the Tweet (in keeping with Twitter's Broadcast Guidelines, counter to the suggestions of e.g., Fiesler and Proferes, 2017).
- I will make contact with every Twitter user from whom I would like to quote, show them the context in which the Tweet will be quoted, and seek consent from them to use their intellectual property in such a way. If the user does not grant me permission to quote them, I will not quote them directly or in a way that reveals their identity.
- I will not quote Tweets unless the benefit my research gains from their use clearly outweighs the potential for harm their use may bring upon the owner of the Tweet.
- I will not quote Tweets from private/locked Twitter accounts (in locking their account, these users have explicitly made their content private).
- I will not quote Tweets that have been deleted by the user.

I hope that my participants – my community, my colleagues and my friends – feel that I have treated their intellectual and emotional materials appropriately, and with respect and kindness in this dissertation.

6

Communities of practice – In Theory

> *The remedy here is to recognize that there is no such thing as ethnography that is not guided by theory (albeit vague and lay) and to draw the implications, that is, to work self-consciously to integrate them actively* at every step in the construction of the object *rather than to pretend to discover theory "grounded" in the field, import it wholesale postbellum, or to borrow it ready-made in the form of clichés from policy debates.*
>
> Wacquant, 2002, p. 1523, emphasis in original.

As Geertz (1973) emphasizes, cultures are never entirely mapped out. Anthropologists can, at best, provide partial and incomplete explanations of the people they study. Through the analysis and interpretations I present in **Chapter 7**, I attempt to reach a kind of understanding of the culture of the reform community. I do this through interpreting the words and behavior of the actors and the identities they develop, the relations between the actors, the artifacts they co-produce, and the context in which these elements are embedded. My interpretation is legitimized by the understanding I have built over years of study, but is only one perspective on a multi-faceted and complex group of individuals, who are sustaining a powerful social movement.

My analysis is also grounded in theory, as Wacquant (2002) insists it must be. I draw almost entirely from one theoretical framework – the Community of Practice (henceforth CoP; Lave & Wenger, 1991; E. Wenger, 1998), although one aspect of the analysis is helped along by Bourdieu's social capital theory (1986).

In this chapter, I describe Wenger's CoP framework in some detail, as a grasp of it is necessary for the reader's understanding of the analyses I present in the next chapter, **Chapter 7**, as I use the framework heavily in my interpretations of the material. The research material and theory might not have been woven together "at every step in the construction of the object" as Wacquant suggests, but I have tried to sufficiently link the analyses to Wenger's theory to validate my observations, and help draw out the interpretation of meaning in the findings.

6.1 A Community of Practice

Reading the CoP framework literature is complicated, as the framework has undergone three key developments since it was first introduced by Lave and Wenger in 1991. That can slightly alter one's interpretation of the framework, depending on the source. Lave and Wenger's seminal work, a book entitled *Situated learning: Legitimate peripheral participation*, focuses on interactions between novices and experts, and the way in which newcomers develop their identity in the context of the community. They discuss that learning isn't only a simple matter of receiving information, but is about increasing participation in a community of practice.

Wenger's 1998 book, *Communities of Practice: Learning, Meaning, and Identity*, expands and develops the original framework. It emphasizes identity, and the trajectory of individuals' participation within the community (Wenger compares 'peripheral' with 'core' participation, for instance). In the book *Cultivating Communities of Practice: A Guide to Managing Knowledge* (2002), the CoP framework evolved again into a more practical form, as a tool to improve the competitiveness of an organization.

Li and colleagues (2009) describe how different interpretations of the theory make it difficult to apply it as a framework. For the purposes of using the framework to help analyse my findings, I am primarily focusing on Wenger's 1998 presentation of the framework, though Lave and Wenger's 1991 work will crop up at times. I will briefly introduce the idea of the CoP, and follow with the definitions of the concepts most central to the framework such that the reader is familiar with them when they surface in the analysis of the coming chapter. Although Lave and Wenger's community of practice identified three major components of the framework (community, practice and domain), I will work through the expanded list of concepts as presented in Wenger's book. Where useful, I will elaborate on some definitions, making initial, general, concrete links to the reform and open science community. I keep this section brief, as I prefer to bring these concepts to life through my discussion of the findings rather than abstractly listing them here in isolation of the research material, and to leave more detailed descriptions to Wenger.

Key Concepts

For Wenger, CoPs are everywhere: an integral part of daily life, "so informal and so pervasive that they rarely come into explicit focus…" (1998; p. 6). According to Eckert (2006), the CoP is a "collection of people who engage on an ongoing basis in some common endeavor." They "emerge in response to common interest or position, and play an important role in forming their members' participation in, and orientation to, the world around them." A CoP is a "learning partnership among people who find it useful to learn from and with one another about a

particular domain. They use each other's experience of practice as a learning resource." (p. 9; E. Wenger, Trayner, & De Laat, 2011)

CoPs within academia, say Brownlie, Hewer and Tadajewski (2009), can be compared to the learning communities found within classrooms and are "informally bound together by shared expertise and passion for a joint enterprise" (p. 139; E. C. Wenger & Snyder, 2000). Research outputs like articles represent only a fraction of their collective practices, which are facilitated through conferences, workshops and seminars, networking online, and collaborating, say Brownlie and colleagues (2009).

Domain

Community members have a shared domain of interest and commitment. This common ground motivates the members to participate in the group, guides their development and underpins their actions with meaning and purpose. The domain is what gives the group its identity, setting it apart from other communities.

People in the reform community are united by a clear goal which drives the movement: scientific reform. They achieve this based largely on practices, which themselves are used to facilitate large-scale adoption of open, reproducible and responsible science practices in academia and even beyond. Many are deeply invested in this common goal, and it is what drives reform actors to improve their own research practice. It underpins their actions and plans to innovate academia to accommodate scientific transparency and integrity. Their common goal also drives outreach efforts aimed at broadening and diversifying the reform community, and opening the realm of science to the public.

Learning

A form of social participation. For individuals, it is about engaging in and contributing to the practices of the communities of which they are a part, and for groups, learning is about refining practice and ensuring the continuance of the community, passing on communal knowledge to new generations of members.

Described in the following points, *community, practice, identity* and *meaning* are the four elements comprising Wenger's concept of learning.

Community Members pursue their shared goal through participating in activities, debate, problem-solving, information exchange and networking. Fostering a strong sense of community and belonging encourages continued investment in the group.

Many members of the reform community are actively involved in group activities, discussion, information-exchange, education and networking, formally and informally. On Twitter, for example, the account @OpenResearchCal is a resource for the community which provides a list of open science related events,

and facilitates collaboration, education, networking and support for the reform community.

An example of formal resource development and education for open science is FOSTER Plus (an acronym which stands for Fostering the practical implementation of Open Science in Horizon 2020 and beyond), a Europe-funded project across six countries which aims to "contribute to a real and lasting shift in the behaviour of European researchers to ensure that open science becomes the norm", in providing training and resources to academics.

Practice First and foremost, contends Wenger, practice is a "process by which we can experience the world and our engagement with it as meaningful" (p. 51; 1998). Concretely speaking, members develop, build upon and share the community's repertoire of knowledge, tools, methods, documents, websites and so on, which they in turn use in their own practice. These communal resources help the community address issues they face in their specific context.

Practice, in its most basic definition (i.e., the carrying out, or exercise of something), is at the heart of the reform movement, because research practices are a large part of what the movement seeks to reform. This has set open science-focused conferences and symposia (such as AIMOS, Metascience and SIPS) apart from ordinary academic conferences, in that they focus on actively reforming science – philosophy and practice. The first Metascience Symposium (held in 2019) was focused on discussing reform, and AIMOS, SIPS and the later 2021 Metascience Symposium have emphasized action. Broadly, the reform community facilitates open science practice by operating workshops, giving seminars, and building repositories of materials and resources for others in the community. They actively engage in connecting people, and uniting current reform and open science initiatives.

Identity A way to reference how learning changes us, moulding who we are, what we do and why we do it. Learning helps create "personal histories of becoming in the context of our communities." (p. 4; E. Wenger, 1998) As people engage, they gain knowledge and their identities develop as a result.

The engagement of community members in discourse (on Twitter, for instance), helps develop their identities as members of the community, as well as the identity of the community as a whole. The issue of identity and how it is developed, both for individual actors, and for the community, is important in my interpretation of my findings in the next chapter.

Wenger described three modes of identification, through which community participants express belonging:

- Engagement involves developing relationships, engaging mutually in shared activities, and managing the group's boundaries.

- Imagination involves constructing an image of the community's practice and participants, and seeing the self as part of that group.
- Alignment involves following directions, aligning the self with the community's expectations, and orienting one's energy towards the community's goals and to fit in with the structure of the community.

The creation of learning is achieved through a dynamic combination of these modes, and the community members' identities as participants in that group are affected by the picture they build of that position.

Meaning Meaning is not used by Wenger in the traditional dictionary sense. Rather, he explains, meaning is situated in a process he calls the 'negotiation of meaning'. This process involves the interaction of two other processes: participation and reification (concepts which I will discuss shortly) which form a duality fundamental to our experience of meaning, and, by extension, to the nature of practice. Participation and reification complement and intertwine with each other, and the negotiation of meaning can only work when they are both involved. Note that the negotiation of meaning is not necessarily a goal-oriented process (although it can be). In a way, it can be seen as 'just what happens' when members of a CoP engage with one another and the community's enterprise.

Participation

Wenger uses the term participation to describe the "social experience of living in the world" (*ibid*; p. 55), in terms of being part of social communities and enterprises. It involves acting and engaging in that context, and is a complex process encompassing our 'whole person', and all that entails (talking, thinking, feeling etc.).

Reification

Unlike most terms wrapped up in Wenger's CoP framework, the term reification can largely be taken at its face value, referring to when something abstract is made into something concrete (even if that thing is not truly a material, or concrete object).[1] Reification in this context is the process by which the knowledge accumulated by the community is projected into concrete artifacts, or tools, symbols, concepts, or stories.

Examples of reification in the reform/open space would be FORRT's open science glossary of terms, SIPS's code of conduct (which is a reification of SIPS's identity), or any one of the reform initiatives produced by the community. These

[1]Though it is worth noting that Wenger uses it to refer to both a process and its product, somewhat complicating his use of the term.

are not material things, but they are the formalization of ideals and expectations of the community; values made concrete by documentation.

Knowledge

CoP members acquire knowledge through participation, which ultimately leads to a communal sense of how to grapple with issues that face the community. This knowledge is dynamic, social and individual, explicit and tacit.

Associating Practice and Community

Associating practice and community, says Wenger, gives tractability to the otherwise potentially fuzzy term of 'practice', and helps define the CoP as a special kind of community. Wenger describes three different dimensions of the association between practice and community – expectations by which CoP members recognize participation, and establish guidelines as to who is an in-group member, an out-group member, or something in between. These three criteria are: *mutual engagement, a joint enterprise,* and *a shared repertoire*.

Mutual engagement Practice exists because actors engage in actions, negotiating meaning with each other. That is what defines the group, or community. Mutual engagement consists of different elements of engagement, from simple interactions among members, to the establishment of relationships and group norms. The inclusion of a member in what matters in a community is important because it is a requirement for being engaged in the practice of a community.

Wenger (1998) emphasizes that one should not assume that mutual engagement entails peaceful coexistence, or allegiance between members in the context of a CoP. "Peace, happiness, and harmony are therefore not necessary properties of a community of practice." (p. 76) He furthers this idea, pointing out that interpersonal challenges, competition and disagreements are all forms of participation (being part of the social enterprise).

Joint Enterprise The concept of a joint enterprise is relatively simple, and can be defined as the community's collective understanding of the purpose of the CoP. Each individual's understanding of the community's enterprise and its meaning for them does not need to be uniform for it to be a "collective product", says Wenger (p. 79). This point will be discussed in some detail in the following chapter's analysis, as it is relevant for many of my observations.

Shared Repertoire During the course of time, as a community engages in the pursuit of a common goal, it generates *stuff*: trappings that link the community to its joint enterprise. The members contribute to these communal resources, which help the community negotiate meaning, in what Wenger terms shared repertoire.

The elements of this repertoire can be random, gaining coherence because they are part of the community as it pursues its enterprise.

In my capacity as an observer of the reform and open science spaces, I have come to see that blog posts about questionable research practices and templates for preregistration documents are as much a part of this community's practice as photos of sourdough posted on Twitter, or SIPS stickers on laptops. According to Wenger's framework, a CoP's repertoire involves stories, gestures, tools, protocols and concepts the community has developed or adopted over time, which have become part of its practice. Linking back to the negotiation of meaning, the community's shared repertoire combines participation and reification, meaning that it includes the "discourse by which members create meaningful statements about the world, as well as the styles by which they express their forms of membership and their identities as members." (p. 83; 1998) The importance of this aspect of the framework to understanding and explaining some of my observations will become apparent to the reader later.

6.2 A Constellation of Practices

Some collectives, by their structure and nature, lend themselves to being described as communities of practice, but other configurations are too diverse, diffuse and broad in their scope for that label to be useful. Such configurations, in Wenger's framework are referred to as constellations of practices. That is, groupings of CoPs which have two identities. They can be identified as single, unique CoPs, but they also can be identified as a part within the broader group. The CoPs within a constellation may share historical roots and have related enterprises, serve a single cause, have members in common, have overlapping discourses and compete for the same resources (p. 127; E. Wenger, 1998).

Gherardi and Nicolini (2002) tell us that a constellation is an emerging discursive community which makes room for multiple competing discourses. In this conceptualisation, the constellation opens up to a kind of learning that is negotiated by the comparison of different perspectives.

Boundaries, Multimembership and Brokering

People tend to hold 'multimembership', that is, to belong to more than one CoP. This means that they are, during their day-to-day activities, crossing boundaries between those communities. As we negotiate these boundaries, juggling our identities in these different contexts, we highlight the differences between communities, emphasizing that they are unique: Histories, repertoires, kinds of communication, for instance.

Most academics occupy many different fields or spaces as part of their jobs. For instance, most senior academics are peer-reviewers, editors and committee

members, as well as contributing to teaching and supervising students, on top of their research work.

Crossing boundaries between different CoPs presents opportunities for what Wenger terms 'brokering', which is the process of transferring elements across practices. Connections are developed between CoPs by people who can port elements of one CoP into another. When brokering is done well, learning is caused as knowledge travels both ways across community boundaries; that is, it is both imported and exported.

Brokering can be uncomfortable work, and this is particularly true in the reform space, because often people are embedded in CoPs that are at odds with the goals of the reform movement. For instance, Stavroula Kousta is the Editor in Chief at *Nature Human Behavior*. She is committed to driving reform in science, and eager to learn how to be more effective in that role (something that has been clear to me in the times I have interacted with her during my study). At the same time, she is the editor-in-chief of one of the most prestigious behavioral and social sciences journals, which represents the traditional academic publishing system (which some in the movement have labeled toxic and predatory), and that has been at the heart of much recent controversy in the reform movement for charging high article processing fees for open access. I will go into more detail on this issue in the next chapter, but it is not difficult to understand with this example as an illustration, how tensions for brokers in the context of the current community under study can be especially high.

7

Qualitative Material Analysis

For years, I watched this community of people. During long night hours nursing first my infant daughter, and then my infant son, I would scroll through Twitter, taking note of exchanges between users and reading through members' blog posts. Tired, but not so much that I could not immerse myself in the world of the reform movement, and learn from what I saw. I made sure I attended as many conferences and events as I could, negotiating childcare with my husband and flight times with the university's travel agent, even as I negotiated my multiple identities: an ethnographer, a member of the reform community. . . a participant-observer.

At conferences and workshops, I drank the same conference-venue filter-coffee as many others did, ate the same "continental breakfast", felt the same jet-lag symptoms. I networked like everyone else, that is, with a warm half-full bottle of beer in my hand in poster session after poster session. During the later period, in the coronavirus pandemic, I took notes while attending online conferences and workshops, one eye trained on the chat; the other on the faces of the attendees who had their cameras on.

Twitter was always open on a tab in my laptop's browser. The app always open on my phone. The first thing I would check upon waking, and the last thing I would see before sleeping at night. For years. This is how I embedded myself in their – our – community.

And yet, I felt I was not one of them. Not really, because all the while I watched, and listened, absorbing each scrap of interesting or relevant information, teasing important details from the otherwise phatic chit-chat. I would slip questions about community and boundaries in to otherwise mundane discussions with members, remembering their responses and scribbling them down as soon as I got back to my hotel room. [This excerpt is a reflection taken from my field notes.]

I have been told that ethnographers do not really speak of data in the way that quantitative researchers might. In the way I tend to. Now I sit back, considering the fruit of my ethnographic work, I can see why. It is not so much data as it is lived experience. The produce of my work (i.e., my notes, my photos and observations) is

infused with my experience. Like postcards sent to myself from a trip abroad – the words with which I record my travels are imbued with my experience of those far-off places. At the same time, they function quasi-legally as proof of my presence. A way of saying "I was there". In this chapter, I share my collection of postcards with you, and point out the ones most salient to me and my experience with this community.

In what follows, I describe the results of the analysis I conducted on my research materials. To give the reader a kind of 'roadmap' for how I have structured this chapter: I have grouped major themes together when they relate to one another (the names of which correspond to central concepts within Wenger's 1998 framework). There are five groupings, denoted by sections **7.1–7.5**. Within each of these sections, dominant themes are denoted by subsection headings, and subthemes are nested within those. These themes and subthemes bear titles relating to the study-specific content they contain. In each theme and subtheme, I explore part of the material, giving excerpts to illustrate them where appropriate, then interpret the meaning of the observations I make with reference to Wenger's framework.

7.1 Negotiating Enterprise and Identity

Describing the Community

As I demonstrated in **Chapter 4**, a wealth of different definitions exist for the concept of community. They are related, and overlap, yet they each capture a different aspect of the community concept. They are also all descriptions of community from the perspective of the *outsider*. But what of the people within the group themselves? How do they describe their own community and its aims and ideals, and how do they want to be described by others? This section deals with this issue, which ultimately helps us understand how the reform and open science community negotiates their joint enterprise, their group's identity, from which they derive individual identity, to some degree.

Participant Descriptions of Community: Formal and Informal

The individual actors within the reform/open science community as well as the group itself have been described using a range of adjectives, by in-group members, out-group members and distant observers, via tweets, during interviews and in documentation. Individuals discussing open science and reform groups on Twitter have described the group as warm, collaborative, welcoming, approachable, supportive, principled, reflective, self-reflective and inclusive. One person shared that, in their experience, the vast majority of the group was welcoming to questions, open to criticism, and supportive of those who challenge ideas.

People who think of the group as a community consider that one of the group's functions is to provide moral and emotional support, especially when it comes to conducting science and struggling with problems associated with the process of research, and concerns about traditional academia. Chris Hartgerink, whom I interviewed in 2019 described one possible role of the community from his perspective: "It's good to know other people have the same struggles, so you don't feel alone." In a 2021 interview, Nosek described the community as being "a feeling of fellowship with others, as a result of sharing common attitudes, interests, and goals".

These experiences and perspectives are very much in line with how the majority of the group would like to be seen. This 'ideal' characterization is reified by the INOSC (International Network of Open Science and Scholarship Communities) Starter Kit documentation (which provides interested individuals with information and ideas for how to begin your own OSC sub-community). On p. 13, it lists the following positive characteristics of people who have started OSC subcommunities in their localities, which projects a sense of the 'ideal' average OSC member. These members are enthusiastic, outgoing, optimistic, inspire others, inclusive, innovative, team players and well-connected within their institution (or able to establish new connections).

The documentation goes further, to detail how condescension and a judgmental tone will work counter to the goals of building open science communities: "It is crucial not to be condescending about this. People who are particularly passionate about OS might consider their workflows superior to those of others and express such feelings. Being exposed to such perspectives can alienate newcomers and lead to unnecessary resistance by creating a counterproductive "us versus them" distinction. Newcomers are more likely to join a community where they can explore and formulate their opinions and doubts, as opposed to an 'expert club' that is judgemental on the workflows of others. It is therefore critical to strike the right tone. Make sure that in all your communication you are inclusive, humble, and respectful. Stress that the OSC is a learning community and that communication is bidirectional: newcomers and experienced peers learn from each other and identify bottlenecks and opportunities to make the transition to OS more fun and fluent, together." (p. 18)

The SIPS mission statement sets out core values that it says are central to the society, which give one a sense of what the reform community prioritizes, or thinks it does: Self-improvement (framed as the aim to improve the rigor of psychological methods and accuracy of empirical claims), transparency and openness, critical evaluation (SIPS purports to foster a space where skepticism and criticism are central to scientific discourse), civil dialogue (the mission statement notes that incivility interferes with healthy scientific critique, and mentions the role of power differentials in the community) and inclusivity are all listed in the document (p. 2).

In contrast with these self-descriptions published by formal organizational ele-

ments of the community, some actors embedded within the community have said some of their fellows are aggressive and 'gate-keepery'. Other, more peripheral actors have compared them to evangelists, a cult, and vegans. One reformer described the group in social psychology terms, speaking of small group dynamics, and referring to in-group and out-group members.

All of the negative adjectives above are used on social media in tweets. As a form of communication, tweets are often criticized for being too short to allow much depth of expression or nuance. As such, actors' use of Twitter can belie their careful and deep analyses of these problems and their possible causes. More concretely, some who are participating in the discourse, for instance, have criticized the group for their lack of intellectual humility, and have shared the perspective that hubris has replaced scientific curiosity. In an email interview, one source [Source 1, henceforth *S1*] wrote: "I think open science activism at times loses sight of the actual science and forgets that we are in the business of asking questions." They elaborate, describing their view that some of the same problematic patterns of thought and behavior that led to the current problems in science are perpetuated in the reform movement. They speak of an "inherited epistemological stance" – that of traditional academia – which impacts on the content of reform activity, and the ways in which proposals are made. It is possible that this observation reflects the difficulty the reform community faces in uprooting itself from the traditional academic world. I will return to this point later.

S1 gives clear examples of the kinds of problematic behavior that are being perpetuated in the reform space: "Not carefully vetting research questions, not clearly formulating assumptions about or models of phenomena, designing research based on expected results and conducting empirical studies to get those results, thinking of complex phenomena in terms of simple linear models with forced categories, assuming that a naive empiricist approach is the only way of tackling scientific questions, looking for clear cut, black-and-white answers to contrived questions and a tendency for dichotomous thinking in general (good/bad, right/wrong, true/false), a desire for making science a monolith, and mistaking substantively weak but emotionally compelling verbal arguments for scientific evidence in support of a claim. . . " These issues go beyond intellectual laziness, says the source, in no uncertain terms: "These are the norms of the field. The publication incentives and citation norms perpetuate it."

Description: A negotiation of identity

For many people who support and engage with reform and practice open science, methodology and initiatives are the primary focus because improving scientific methods is central to the group's enterprise. Others in the group, however, choose to treat open science and reform as a social identity in that they identify personally with some of the values of the collective. The idea of a community certainly

implies a shared identity, but some actors discuss the idea directly and ruminate over what that might mean for them, and the group as a whole.

Nosek weighs up the benefits and risks of treating reform and open science as a social identity in a 2021 Twitter thread. On one hand, he writes, treating open science as a social identity benefits actors in that they can rally around a common cause, find others with similar interests, can provide meaning and can help with articulating what it means to adopt the practices for some who are interested but not yet practicing. On the other hand, says Nosek, potential risks are that stereotyping of "open science" and "non-open science" people can occur, creating an us-versus-them mentality. Other risks are raising barriers to inclusion, polarization, and shifting the focus from principles and behaviors to being about belonging. Nosek warns that these risks are deepened by the community's use of group-based language (he gives the examples of "open scientists, SIPS, social psychologists") when individuals criticize behavior. He writes that "Inevitably, the groups are more heterogeneous than the basis of the criticism, widening the conflict." (Nosek, B. [@BrianNosek], 2021a) In another, later tweet, he extends the discussion about the problems with treating open science as a social identity: "So, if we derive some identity from open science as a movement, where does it get attached? It can be orgs and others when available, but in the murky realm, when bad actions occur in the name of open science, then we are forced to ask ourselves "Am I part of that?!"" (Nosek, B. [@BrianNosek], 2021b)

The use of the "Open Science" label and its salience as an identity signaler is another point of debate in the group. One actor shares the perspective that the utility of the Open Science label is 'outward facing', and "agnostic to the community's internal heterogeneity." Charlie Ebersole, in a discussion with me at MS2019 said that he thinks the label is useful because it speaks to the formation of institution, which can drive the community to do useful things. He mentioned that the establishment of the institution would facilitate training for students, which is useful because skills relating to open and meta-science are highly applicable even if the future of fields like metascience are not certain.

As I discussed above, *S1* observed what can be interpreted as the friction that negotiating a new, 'reform' identity is generating in the greater academic sphere in which the reform movement exists. They recognize that actors are attempting to transform an academic system in which most of them are still enmeshed. The emergent reform community is attempting to uproot itself; estranging itself from a traditional academic parentage, and redefining itself as the generator of a new scientific way of thinking, and way of doing. Wenger states that CoPs are not self-contained entities; rather, that they develop within larger contexts. Along with that come shifting, dynamic boundaries and tensions between the larger context and the CoP as it establishes itself.

Part of this process of pulling away entails the establishment of a new identity, and what we see in the material I have analyzed is evidence of a negotiation of that *emerging* identity. All the examples I gave earlier – of what the community

is like, and how they behave, or lists of member characteristics, codes of conduct and mission statements – are attempts by members and groups of members within the CoP to reify their experience of the collective identity of the CoP, to define their joint enterprise; that of reforming traditional academia. They are part of the CoP's negotiation of identity. Of this, Wenger writes that members' "... identities form in this kind of tension between our investment in various forms of belonging and our ability to negotiate the meanings that matter in those contexts." (1998; p. 188) What matters to this community? What are its priorities? Indeed, that this seeming tension between ideals and reality is evident reflects the fact that the community's members are invested in shaping their identity, and are committed to their enterprise. Wenger calls this reification of "aspects of accountability", and it is all part of the practice.

7.2 Participation and Reification

Community Priorities

One point of discussion in the group which has become more salient to me over the course of my ethnographic study concerns what actors prioritize in the reform community, and what they think the group as a whole should prioritize. Though a simplification of a complex issue, I visualize the breadth of possibilities as sitting along a kind of spectrum, or continuum. One actor referred to a continuum with values on one end, and implementation on the other. On one end of this line are those people who are driven primarily by "reform ideology" (as one actor called it), and who prioritize values and value diversity in the group. On the other end of the spectrum are those who are driven primarily by reform methods, and see open science and reform activity as being largely about refining scientific practice.[1] Looking at the whole spectrum of priorities, or drives, we can see scientific reform as a social movement, represented in all its facets and dimensions; the people of the reform space represented as a heterogeneous group, contributing diverse and complimentary perspectives and skills to the movement.

I engaged people on the issue in interviews, explicitly asking them what they thought of this idea. Nosek, when describing priorities for the community (which included accessibility, inclusivity, openness, transparency, rigorousness, and reproduciblity), reinforced that people greatly vary in what each of these things means to them and how they prioritize them in the movement, but pointed out that there is still plenty of overlap in the group: "... if we were to conduct a cluster analysis of a rich measure of open science attitudes, interests, and goals, we would find collections of individuals with overlapping views." *S2* spoke of

[1] I recognize that speaking of such a spectrum gives an overly simple impression of a very complex group of people, who are involved in a intricate and dynamic social movement. The majority of the people I observed do not conform truly to either extreme of this spectrum. Rather, as with the left-right political spectrum, people *lean* one way or another.

how the idea and the use of open science tools are correlated, but how they are completely independent.

Most members are somewhere in the middle of this imagined spectrum, and the vast majority of actors in the group acknowledge the importance to balancing both of its extremes when it comes to reforming science. Paola Masuzzo (@pcmasuzzo) tweeted: "Open science is not just about improving the way we share data and methods; it is also about improving the way we think, work and interact with each other. It's about technology enabling social infrastructure that can promote inclusivity to create kinder science." (2021) When confronting another actor about his comments on a Twitter thread, Nosek directly counterbalanced values and practice, stating that inclusivity is just as much a part of 'open' as transparency. I will draw this out further, to explain that inclusivity is can be seen primarily as a *value* of open science, while transparency is *enacted* through people's engagement with open science and reform practices. One actor expounded on Nosek's tweet, stating in a 2019 post, that while the main goal of reform is to be anti-exclusionary, you need to be able to identify and solve problems with practices (she gave the example of open access) to achieve that goal. Here, she highlights the interplay between values and methods, and the idea that values-driven aims of the community can be achieved through working on practice-related barriers.

Another actor pointed out that how one sees open science influences how we think about different approaches to open science and reform within the group. If we see OS as a behavior (that is, a practice more so than a value), she says, we should accept that it will not always be (and does not need to be) ideal, or perfect, and that 'open' will be manifested differently depending on when it is being practiced, by whom and in what contexts. *S1* shared a similar reflection: "So to me, open is an adjective that qualifies science. Science is the common thread that brings many people together. As researchers, many of us think about, engage with, consume, and actively contribute to science to answer some questions and to gain knowledge. "Open" is a qualifier that may refer to different things in different contexts, and for different people."

Those who are more focused on the proximal goals of adoption and implementation of practices tend toward having a single-minded drive to improve methods and statistics. Some view the methods of the more extreme actors for provoking change as too inflammatory, but others believe that the work some of them do for the community is worth too much for them to be excluded from it. Two of my interview sources (*S2* and Source 3; *S3*) made clear statements to this effect when discussing people in the group who were known for being highly critical of others' work. *S2* said that, all-in-all, they have brought huge benefit to the field. *S3* said that the helpful contributions dominate the problematic ones. Other actors have spoken about a 'greater good' on Twitter, in the context of explaining and justifying the actions and words of those who expect others to conform rigorously and wholly to reform practices.

The opposing view, that scientific contributions as an end do not justify the means (i.e., the 'good' done by bullies in the community does not justify the destructive way in which the good is achieved), is also represented by actors in the group. People engaging on Twitter over this issue with this view argue that no one is above the 'law' no matter whether or not you're a good scientist, and that we should have faith in the collective genius of scientists to think that we can still conduct good research without "elevating assholes". The overarching goal of science reform will prevail, even if bullies who do good work for the cause are removed from positions in which they can cause harm. One interview source explained her view that the reform movement's moral context creates an environment where bullying can be tolerated, conveniently overlooked, or even facilitated when it comes to promoting 'the cause'.

Belonging to a Community Versus Using their Methods

By now, it is clear to see that the question of whether or not we can speak of a 'community' (described as a tight-knit group which provides members with a sense of belonging and support, rather than a diffuse group with nothing connecting them but a particular interest in common), is as central to debates within the group (and between them and outsiders) as it is difficult to answer. On one hand, many actors will strenuously argue that there is not a community, but that the enterprise of open science and reform is about practices and changing long-standing problematic incentive and reward structures in science. In a 2017 post, one vocal figure in the group takes a strong stance against the idea of a community, writing that open science isn't a group of people separated into member and non-member categories. Improving science (the enterprise of the group), they go on, is about research and not belonging. In 2022, I checked back with this source, asking whether they still held such a view, and they helped give the original post some nuance. In their view, while there is a kind of community it is very diffuse in nature, not like a sports team where there are clear members and non-members. The source went on to say that the community is not synonymous with the original enterprise around which it formed. People can be committed to certain open and reform practices but be disconnected from others who engage with those same practices. The opposite can be true too, according to the source: people might be fully embedded in the community and its culture, while having little motivation to actually practise 'open science'. In the context of the reform community, says the source, any community constructed around reform and open science goals and principles are superstructures.

This perspective is shared by other participants. For instance, a 2021 interviewee [Source 2; *S2*] spoke with me about how the idea and the use of reform and open science tools are correlated, yet completely independent. They mentioned a recent event on Twitter (involving several people in the group aggressively criticizing a scientific article): "I think in the last few days on Twitter, we saw there are

people who use all the tools, but they are not in any way invested in the culture, or diversity statement or whatever, they just don't care." In a sense, this source indicates that they believe there is no community per se, rather the enterprise is just about tools and their application.

Others have distanced themselves actively from the community because of their principles. They do not want to be associated with it because of its problems, and the positive values and norms that the community are not enough to secure their alignment with the group. For instance, Alex Danvers (@Alex_Danvers; 2021) tweeted about his decision to distinguish between his own goals and the values of the community: "Today I changed my profile to indicate I'm interested in science reform, not open science. It's been hard for me to come to grips with the idea that the community I originally found under the open science label (in 2012) is different from the way the label is being used now." Katie Corker (@katiecorker)shared the perspective of some others who might not want to be associated with the open science community, while at the same time happily and openly engaging in its practices: "'Alternative explanation: not just self serving bias but identity politics. People don't want to be seen as a "replicators." I see the same pattern with other practices. A: are you an open science person? B: no way! A: do you share your data, etc? B: of course!'" (2019). Another actor shared their concerns with someone else using the open science label to describe them. Although they recognized that being described as an "open science person" wasn't factually incorrect, they was concerned that the label was too simplistic, and failed to capture their complex views and approach to using open science tools to conduct rigorous science.

This problem – being labeled as part of a community when for some reason you do not wish to have such an association – has been explored by Gee (2005), who advances the idea of *affinity spaces*. These are locations, physical or virtual, where groups of individuals are drawn together to engage in a shared interest. Gee's idea of affinity spaces, which focuses on the *space*, rather than membership in a community as a central concept underpinning shared learning, is an alternative to Wenger's CoP framework, he says. Rightly, Gee notes that the use of any kinds of labels, particularly when it comes to community labels (presumably because of the identity you share with a community when you are a member) is challenging. He writes:

> We face vexatious issues over which people are in and which are out of the group, how far they are in or out and when they are in or out. The answers to these questions vary – even their very answerability varies – greatly across different social groupings. If we start with the notion of a "community" we cannot go any further until we have defined who is in and who is not, since otherwise we cannot identify the community. Yet it is often issues of participation, membership and boundaries that are problematic in the first place. (p. 215; Gee, 2005).

Gee, in this quote, perfectly captures one of the key 'problems' I have pursued in this research project. More importantly for the current context of interpreting the findings, he clearly sets out the intractability of *community*, which may explain why some of the people I have observed and spoken with are uneasy with the idea.

Regardless of what you think of the community, most people seem to agree there is a community of sorts, though most of those will concede that such a community will have very loose and undefined boundaries. Nosek, in my interview with him, suggested that because the community is diffuse, membership is "vague and self-defined". Others call the community informal, and have commented that this status means shifting boundaries for norms and behavior. Two interview sources stated that they thought of the group as a community in its earlier days, when it was smaller and more cohesive. In their perspective, it has grown out of the 'community' descriptor as it has grown in size, and that its bounds have become fuzzy. There seem to be costs involved with a growing community with soft bounds. One of my sources mentioned that you need to watch what you say much more in a larger group because with so many personalities and views in play, you are more at risk of saying the wrong thing. Another pointed out that as groups become larger and gain more power, they are likely to see more discord and disagreement among actors. This is a natural and expected progression, and not a surprising turn of events for the open science community, the source concluded.

Membership Styles: All-In and Buffet

The values-implementation continuum is somewhat abstract, so I will describe a recurring theme in the discourse as a means of illustrating discussions of the continuum. A subset of the people who are more extreme about their prioritization of the implementation of practices seem to be of the opinion that people practicing open science need to go all-in, implementing every initiative that the community has come up with. This all-in approach is difficult for many researchers, despite how deeply they support the open science and reform philosophies, because it takes time and effort to learn and integrate new practices into ones research process. Moreover, for some researchers the practices are not applicable to the research conducted in their discipline.

Those who prioritize values strenuously push against the all-in approach, because it disproportionately disadvantages some people in the group, and raises barriers for many, where open science and reform should be lowering as many barriers as possible (a point *S1* emphasized in our interview). For instance, consider a case described by another interviewee, Source 4 (*S4*). They explained that an ECR friend wanted to pivot to transparent reproducible quantitative science, but that her supervisor wouldn't allow it. Not everyone is in a work or lab environment where they can easily just adopt every practice. *S4* pointed out that some people are not in positions where they can spend a lot of time adopting

practices, because they have caregiver roles or lack other resources to make it feasible.

This all-or-nothing attitude affects different fields too, as some open science initiatives and practices (e.g., data sharing) largely target quantitative research pipelines, leaving qualitative researchers questioning whether they are welcome, or whether the practices are meant for them. The all-or-nothing stance, with all its dogmatism, has a strong value component. Its focus is on practice (methods, statistics, etc.), but in some cases it becomes too extreme to be pragmatic.

A strong voice of opposition to this extreme approach has been gaining momentum, with people rejecting it as 'gate-keepery', and discussing the counter-idea of an 'open science buffet'. I explored this idea in detail in **Chapter 3**, where I discuss open science practices, but I will briefly reiterate the idea here too. The idea of viewing practices and implementation as a buffet was first introduced by reformer Christina Bergmann in a talk she delivered in 2018. Just like there are many dishes at a restaurant buffet, the actions of the reform movement and open science communities have provided researchers with many practices to implement in their research. She emphasized this in a tweet: "Seems a good moment to remind people of the Buffet approach to open science. Not everything works for everyone, and it is ok to pick practices." (@chbergma; Bergmann, C. [@chbergma], 2021)

Others in the movement have taken the idea and run with it. One points out that the all-or-nothing perspective is the result of poor open science PR rather than objective truths, and recommends that people take what helps them given their unique research needs, and ignore the rest. A tweet by Andrea Howard (@DrAndreaHoward; 2021) makes a similar point, reinforcing the feeling of exclusivity that some feel: "In a setting where to many, open science already feels closed off to only a select group of Super Science insiders, the last thing we need is another outlet for self-congratulatory posturing. Just . . . do the science in as open a way as you can, when you can." Another poster tweets that open science should acknowledge the many types of scholarly contributorship that people in the group can offer.

The all-or-nothing versus buffet stances are concrete examples of how people in the group perceive the relative importance of practices and values; how they prioritize different things in their approach to the group enterprise. Those who focus heavily on methods and practices at times seem to argue that open science and reform is a zero-sum game; that giving more time to inclusivity and diversity efforts will mean less resources go to improving methods. Those who are focused on values, on the other hand, seem to focus more on the long-game of the reform movement – distal goals for the future of science. They argue that making science more kind, diverse, accessible and inclusive can be balanced with implementing practices for better science. They demonstrate that these values can be developed and upheld alongside methodological reforms activism.

Linking Reform Methods and Participation in the Community

The clear lack of agreement over whether or not there is *an* open science community reflects the fact that community is, in many ways, subjective. How we view the meaning of community, and how we perceive our own roles and belonging determine how we define community, and influence whether we agree to being classified as being a community member. This subjective view of the concept of the community gets more complex for some people in the group, who want to do practices but actively exclude themselves from a community proper, not because they focus on practices and tools, but because they do not wish to identify with the community's values (and/or because they reject the idea of a strictly defined community). It is also intersubjective – the community emerges from interactions including its interactions with other communities (what Wenger calls 'boundary work'), the inclusion and exclusion of certain people, defining and making concrete its goals and values by reification, setting norms and distinguishing itself from other communities.

The observation of *S2* and others – that there is a possible subgroup of people who are not interested in the community, along with the cultural and social aspects of engagement with it, while still using the tools – deserves some attention in the context of the CoP framework. In short, Wenger's framework does not allow for such a separation between a community and its enterprise (at least, not the kind of separation that the word 'superstructure' implies). While such a subgroup might not adhere to what Wenger calls reified "aspects of accountability" (p. 81) such as community rules, standards and policies, that they engage in the community's practices is enough for them to be part of the community. This is because when individuals engage with the tools of a community (whether in their construction or use), a transaction opens between them and the community. Their identity is shaped by their engagement with the community's tools, and the community and tools are likewise shaped by their engagement. This is because, to Wenger, participation is itself a *source* of identity: "The transformative potential goes both ways", he declares (p. 56).

To put this more concretely, this subgroup is furthering the goals of the reform community by practicing transparent science, and rejecting traditional scientific 'ways of doing' along with the community even if they do not engage with others in the community. Now, it can be said that these people may be peripheral in terms of their embeddedness in the community, and their participation may be less intense than that of those who are deeply enmeshed with others in the movement. To push this point even further, in the CoP framework, participation is a complex process combining functions like thinking, talking and doing. Even if some people's engagement with the community is limited to the use of its tools, they are still participating in the sense that they are thinking about applying the tools to their own academic practice, doing it, and likely are, at times, having to discuss this thinking and doing with others in their surrounds. At some point,

these people have considered the benefits of using a novel set of tools in their research workflow which go sometimes counter to tradition and which are sometimes not easy to apply or straightforward to learn. They have chosen to apply the tools to their academic practice, despite these boundaries. People in their research group, for example, might ask why they preregister their projects, or upload their study data or software code on an online repository.

If you will allow me to exercise this idea just a bit further, we might even go so far as to consider this subgroup themselves a community. In their choice to use the tools of the community, while eschewing engagement with the group themselves, they are a community adjacent to the reform community. The products of enterprise reification such as tools can cross boundaries between them, but there is too much discontinuity between the communities for them to be considered as one. I will return to this idea at a later point.

The Duality of Participation and Reification

The tension between the tools and their meaning in the overarching enterprise of reform can be analyzed in the light of Wenger's discussion of the tension and duality of participation and reification. The practices and tools of the reform and open science community are a vital part of the community reaching its goals. At the same time, participation of members of the group is required to give meaning to the tools, and keep their use check (avoiding the issue of the use of the tools becoming the goals in and of themselves).

Take the use of open science badges, for instance. For some years, journals have awarded badges to articles which reach certain standards of openness. An article can get a badge when its author has made the article's data available on a public repository, and it can get one when the paper's methods are likewise shared. Some journals also award them when an article has been preregistered. Journals do not always check that the authors fulfil the conditions of the badges (i.e., that what they declare is open has actually been made so), but tend to rely on a trust system, believing authors who say they have made their data and materials public, or that they have preregistered the study.

These badges have become a point of much debate in the community. To many, they are a step forward for improving science, and help increase transparency in research practices, and trust in scientific literature and researchers themselves (see Kidwell and colleagues and Schneider, Rosman, Kelava & Merk, respectively 2016; 2021), both of which are concerns for the community after the crisis of confidence. As each new journal adopts badges, people post excitedly about it on Twitter, and applaud the journal's progression in the new, open science order. Some have been critical of the potential for abuse of the badge system, however, raising the concern that people might try to get the badges without doing the necessary work for them in reality. Other points raised suggest that human error and a lack of standardization across fields might mean that people

do not always meet the conditions of the badges. Some see the badges as an example of gamification of open science, finding it condescending and infantilizing. We need infrastructure to support open science, not brownie points and walks of shame, tweeted one reformer in 2021. Some dispute whether they work at all as a means to incentivize transparency (see Bastian, 2017).

The badge system reifies the goals of the open science enterprise, projecting them into scientific practice, and creating points of focus around which actors negotiate what matters to the enterprise. Reification can be too succinct, portable, persistent, and focused, emphasizes Wenger. When the focus is too much on the badges themselves, we lose sight of *participation*, the overarching point of them in the first place, which is to increase transparency, and, in turn, increase the quality of research output. In other words, when a measure becomes a target in and of itself, it risks losing its value as a measure.[2] Participation is needed to avoid such potential 'misalignments of reification' (p. 63; E. Wenger, 1998). When one combines the use of reification such as the badge system with participation in the community's enterprise, meaning is negotiated.

To animate this idea of meaning negotiation, consider the example of replication. Replication as a topic has been explored a great deal by reformers, and continues to be a salient point of debate. Reformers discuss what replication is and is not. They haggle over how to define different kinds, purposes and motivations for replication. They debate whether it is really useful or not as a way of establishing scientific truth. Continuous interactions of this nature work the group toward a particular kind of meaning; the group jointly reaches the meaning of reform.

Every time meaning is negotiated in this way, where reification of science reform ideals or concepts meets participation in interpreting their meaning to the enterprise, what science reform *is* in practice is established anew. What practicing reform *means* to the members of the group is re-established. Wenger's concept of meaning negotiation is a process of transactions; of feedback between actors and the group, and between sub-groups of actors. By practicing science reform, reformers do not just impose their ideas of how that process should be performed on the rest of the movement in isolation. Nor are practices of reform simply imposed upon them. There is a continual building up of what reform *means*, as things like openness badges and replication are debated and practiced within the community.

The Danger of Reification

Wenger warns of the 'double edge' of reification. Reification can be a powerful process by which attention can be focused and support galvanized in a movement. At the same time, it carries the risk of becoming a substitute for commitment

[2] Some readers might recognize this as a description of Goodhart's Law.

to and understanding of the goals of the movement. The enterprise of the reform movement is reified by tools such as preregistration, unification of national agencies across Europe to ensure wide-scale open access publishing (Plan S), and labels such as those discussed above which indicate group identity, to briefly return to that theme. However, people can adopt the open science label, practice the initiatives and plaster their laptops with open science stickers, without truly engaging meaningfully with the community and its enterprise, as I hinted before.

To further drive my argument, I will flip Ebersole's earlier point (about the open science label being agnostic to complexities in the group) on its head. On one hand, 'open science' reflects an overarching and unifying goal, as Ebersole implies. On the other, it obscures the important differences between people in that group from view. The label fails to describe the complexities and diversity of approaches to the group's enterprise that are represented in this community. These differences are especially masked from outsiders, who contain potential adopters; possibly the most important people the group would like to attract.

Priorities: Negotiating the Joint Enterprise

The tension between different groups within the community pulling for different priorities says something about the process involved in defining a joint enterprise. It is a process, writes Wenger, rather than a static agreement, which "produces relations of accountability" that cannot just be boiled down to norms. Part of defining a joint enterprise is when the enterprise is reified. Regardless of whether you advocate an all-in approach or prefer the buffet, the debate takes place about reified elements of the enterprise – tools, norms and so-on. But the practice does not only include these elements of the shared repertoire. Crucially, for Wenger's framework, the practice also includes ways in which actors interpret reified aspects of the enterprise, and integrate them into their participation in the community. So, practice involves reification as well as the negotiated interpretation of those reified elements. The reform community's joint enterprise is partly defined by the duality of reification and participation, as I discussed at the end of the previous section, but also the negotiation of the interpretation of reification and participation. Reconciling conflict and friction over interpretations is what the enterprise is all about. How should the community prioritize certain aspects of open science and reform, and how should they ideally be integrated into participation?

The enterprise of reform is constantly evolving in response to exogenous as well as endogenous change, which is an important aspect of the process of negotiating a community's shared enterprise. In this case, consider changes from the wider scientific community and institutions (for instance, consider the uptake of data sharing, preregistration and registered reports), as well as from within the reform community itself (in response to activism, feedback and critiques from reform actors). This process for this community in particular is highly reflexive

and complex because what they are setting out to achieve *is* change itself: change within the reform community, as well as change within the academic community within which it is situated. This means that, in practice, the joint enterprise is perhaps more malleable and dynamic than the average CoP, and more volatile or responsive to disruptions.

7.3 Conflict, Power and Disrupted Participation

Critique and Criticism

This section describes behavior by reform actors which has been judged as negative by other reform actors. It refers to actions and words targeting other reform actors, as well as targeting people outside of the group (whether they are outgroup reformers or people who have conducted research that is the focus of criticism). It can be called "bad behavior" or "bullying", and people in the movement tend to describe it in terms of problematic tone. I am referring to behavior that to others may look too mean or aggressive. It may involve outright ad hominem attacks, or may manifest in criticism of a research paper that seems unkind or unfair in some fashion. Actions and words that are intended to exclude other people are clear examples of problematic behavior. Bad behavior in the group can be defined more formally too, by SIPS and the formal OSCs, for instance. They both list unacceptable behaviors such as "discriminating, racist, intimidating, harassing, lewd, demeaning, bullying, stalking, or threatening speech or actions" and "sustained disruption of speakers or events (verbally or physically)". They also note that advocating or encouraging these behaviors in others, is in itself unacceptable behavior.

Note that I have categorized it based largely on how it is or can be *perceived* by reform actors, rather than on whether or not it is objectively classifiable as bad behavior, to the extent that is possible. Unsurprisingly, perspectives on what behavior is bad and appropriate vary greatly, depending on to whom you speak.

The effects of negative behavior from individual actors and small groups are made worse by the Twitter platform. As one interviewee lamented, "... Twitter just somehow brings out the worst in people". It's designed that way, to polarize and inflame people, opined another. An interviewee I spoke with in 2019 put it plainly: "Being disliked is likely a function of Twitter as a platform as its algorithm favors and amplifies negative content." Alex Holcombe (@ceptional) explained it in the context of the effects of bullying specifically: "Just that occasional twitter users get the wrong impression of bullying being more representative, due to the algorithms." (2018)

Indeed, Twitter's algorithms are intentionally written to amplify partisan and incendiary rhetoric (Stark, Stegmann, Magin, & Jürgens, 2020). Radical and extreme Tweets are prioritized to appear on many people's feeds, at times even inserted into the feeds of users who do not follow the accounts responsible for the

tweets. Some of the effects of Twitter's algorithms seem to vary depending on the content, too. A large-scale study conducted in 2020 involving millions of Twitter users revealed that the "... political right enjoys higher amplification compared to the political left" (p. 1; 2022) Twitter's meddling, driven by sophisticated machine-learning systems, can affect hundreds of millions of tweets every day, according to a Twitter Company blog post (J. Williams & Chowdhury, 2021). This is a well-documented phenomenon, which gained great exposure during the 2016 US presidential election (see Darcy's CNN Business article for a brief discussion of the role of Twitter's algorithms in driving political rhetoric; 2019), and it affects all corners of the Twittersphere, including the group of people at the heart of this ethnography.

Bad Apples or Systemic Rot?

Whether the perpetrators of problematic behavior are 'bad apples' (i.e., such people are few, and likely randomly distributed in the reform population) or whether the behavior of some people is a reflection of symptoms of a pervasive, or systemic problem (i.e., they are manifestations of a sexist, racist academic culture), is very much a topic of debate in the reform space. The majority of actors that I have personally engaged with are of the opinion that the problem is a matter of a few rotten actors spoiling an otherwise healthy field. "There are two and a half people who misbehave on Twitter", declared one actor in a session at the 2019 SIPS conference, held in Rotterdam (SIPS2019). "It's fairly few people who tend to cross the line by most folks' estimation" summarized the session leader in the notes written up afterwards. I chatted with Charlie Ebersole, a prominent reform actor at MS2019, who seemed to confirm this, referring "a few crazy outspoken actors" as being the root of much of the discord.

The bad apples perspective carries the risk of obscuring bigger concerns with culture in the community. For one thing, many bad apple actors have large platforms on Twitter, or are otherwise well-known or senior in the community. In response to one of the many negative interactions between one such actor and others in the community (2019), one reformer tweeted that the incident was the latest case of senior people with large followings not understanding (or caring) that their grandstanding against dissenters is destructive to the community. One actor's 2018 post adds complexity to this observation. They suggested that how representative bullies are of a community is not the point. Rather, they argue, it is a matter of those bullies having disproportionate power in a community, especially over those with relatively less power. These tweets highlight that even though the genuine 'bad apples' may be few, the impact of their behavior on others in the community should not be dismissed. In a 2021 tweet, another prominent actor wrote that while there are relatively few bullies, there are many people who enable them, either actively or passively. This comment highlights another problem: That it is not just the behavior of the bullies, but the (lack of) response of

the greater group that can inflict damage.

One vocal senior reform actor explicitly criticizes the 'bad apples' perspective of behavior in the community. They post that to boil the problem down to a few bad apples is to perpetuate the bigger problem of toxic group culture. They follow up by saying that the perspective of the problem being a matter of a few bad apples represents a greater culture problem which leadership fosters. It is a way of dodging the responsibility of actively fostering an inclusive culture, they argue in their posts.

A vocal minority have brought a counter perspective to the discourse, ascribing behavior issues in the community to academic culture. This perspective is reflected in the previous tweet. A prominent voice in the online reform discussion clarifies this stance in a 2018 post. They don't believe that the problem is down to a few problematic "bros", they said. Rather, the issue is that toxic behaviors like bullying stem from systemic social problems in science. They emphasize that these issues are present regardless of whether it is the 'open science' community, or just the regular science community. *S1* explained the perspective in more detail: "The whole affair is driven by assumed epistemic privilege, moral grandstanding, policing, and gate keeping who is allowed to do science and how. That's a patriarchal, authoritarian mindset in essence so there is nothing surprising about the prevalence of bros and bro culture in the movement."

Lisa Feldman-Barrett at MS2019 shared concerns about speaking out against bullies, regardless of which perspective you share. She pointed out that if people imply, even unwittingly, the field is a cesspool of bad people, no one will listen to you. Tweets I have collected give weight to this. In a session at SIPS2019, another reform actor acknowledged these risks. Speaking truth to power, she said, does sometimes alienate people. She concluded with the declaration that such outcomes are inevitable.

Critique: It's Just Business

Another angle on the systemic explanation for unwelcome behavior was mentioned by a few different sources in the course of interviews: a few people put down behavior issues to the 'business model' of the reform community, and the fact that the nature of the movement attracts a certain kind of person. Reforming science can be difficult and costly, pointed out *S1*, because it involves taking on the role of criticizing the work of others: "it's their business model, it's criticizing", he shrugged. Another actor shared in a tweet that not every reform will be a win-win situation, thanks to powerful, persistent, and consistently enforced norms against reform action. *S2* offered their simple explanation: "I think it's a fairly general mechanism, where if you have a community that kind of develops outside of the mainstream, in part founded on the assumption that we want to criticize people…".

That same source went on to highlight a possible link between the goals of the reform movement and the kinds of people that are attracted to it. "I don't think it's surprising at all that people who are fairly critical and fairly disagreeable show up more in the OSC because obviously they didn't feel very at home in the mainstream community." *S1* called it a 'third variable', speculating that the same variable that attracts people to open science and reform activism is also responsible for certain behavior when it comes to critique and criticism. The possibility of reform 'attracting' certain people has also been brought up on Twitter. In 2021, one actor stated that open science attracts non-conformists who are high in openness but low on agreeableness. He went on to point out a distinction that is often glossed over in Twitter debate: not all disagreeable actors are bullies, particularly when they fight power. Another actor was less charitable in speaking about reform attracting particular characters. A movement focusing on overturning traditional, accepted scientific practice will inevitably attract many people who are inclined to bully or be bigots. Openness, they concluded, means openness to bullying and aggressive behavior.

Other users were less nuanced in their criticism of reformers. For instance, one wrote that people who are hated in their fields responded by starting Twitter accounts and advocating for open science on them. This kind of criticism is not without push-back from others in the group, however. In 2018, one person tweeted that their intent to take open science back from "assholes" on behalf of others in the movement who care about scientific credibility, and know that people must support each others' efforts to do better science if they are in a community focused on such a goal. Laura Scherer (@ldscherer) tweeted a similar statement, in the context of sharing a negative experience with another actor in the group: "Just one final thought: In spite of this experience, open science and science reform are really important to me. And it won't stop being important to me because some people who have an "open science" identity have bad takes, or are even total assholes, on twitter". (2021) Statements like this reflect an acknowledgment on the part of reform actors that the goals and values of the reform movement are important to uphold and persist with, despite the actions of some problematic characters damaging the image of the movement.

Impacts of Bad Behavior

Irrespective of whether bad behavior is perpetrated by a few bad apples, or whether its roots lie in deeper, systemic issues of culture, the majority of the group recognize that it carries serious consequences. These consequences affect the in-group, as well as actors who would potentially join the reform space, or associated practice communities (potential adopters).

Negative behavior affects the in-group. The behavior itself can create discord and feelings of alienation or exclusion, and it can make people afraid to engage with the group. In a 2019 post, one reformer wrote that the negative behavior of

some in the community won't stop people who are committed to open science and have the resources to invest in it, but that it might fracture the community and create negativity. This, declared the poster, is the opposite of what open science means.

Negative behavior can affect individuals indirectly, too, and not in ways that one might expect: For example, Lisa DeBruine (@LisaDeBruine) shared that the gap between her own experience and others labeling negative behavior in the community was taking its toll: "Tonight I'm feeling weirdly excluded from the open science community, not by people in it, but by people outside insisting it's all "bros", which is so far from my experience that I have to assume they must be excluding me (and tons of awesome women and men) from the community." (2018)

Negative behavior can have the effect of disproportionately affecting early-career researchers (ECRs), as they have relatively less career security and experience, and can potentially be more easily intimidated. Importantly for the goals of the movement, it may discourage them from engaging in the discourse. In 2019, one disillusioned person posted that, for them (an ECR observing from the sidelines) it was disappointing to see senior scientists shutting down those whose opinions counter their own. If dissenting leads to being mocked on Twitter, warned the poster, the response by many ECRs might be to just keep quiet next time. It is not just in-group ECRs that might be discouraged from involvement with the group either: bad behavior from within the movement will doubtless drive many potential adopters away, regardless of their academic career status. This is a concern that is regularly expressed by reform actors and those on the outside looking in alike.

Crucially, negative behavior from the in-group can impact the likelihood of others joining the movement, or adopting its practices. As most tweets are publicly available, people on the fringes and outside of the reform group can watch group's Twitter engagements as they happen, or use Twitter's search engine to go through the threads. It appears that many people are not happy with the face of the movement that outsiders are likely to have seen. Tweets like the following give a sense of this concern: "Wow Open Science^{TM} seems like such an accepting and kind group of people #sarcasm"[3] (Blommaert, J. [@drjulie_b], 2017), and "I had to deal w/a lot of trolls today in my feed. But I feel worst for people who replied and then had to put up w/the random racist, sexist trolling that has become standard fare in the "open" science community. Just look at the progression of this shit. Ya'll got a problem." (Rai, T. [@tage_rai], 2020)

[3]Note here the use of the trademark – this is often used by people on Twitter as a means of drawing attention to negative behavior relating to the OSC, as one reform actor suggests in a reply to someone querying its use that it is a sardonic and witty poke at negative behaviors from open science advocates. It also indicates that it is a trademark of the community, that the community is known for this and is part of its identity.

One reformer, who has been involved with reform activity from a point in time early in the 'crisis of confidence era', sets the issue out very simply in a tweet from 2020. They posted that in-fighting in the reform movement has led to the erosion of the communal mission. An interviewee I spoke with in 2021 echoes this concern, framing it in terms of the role the group can have in encouraging people to join the movement with kindness and empathy: "I think there are better ways. . . I think about the way you learn a new thing or be convinced by something, you would want to be convinced or taught in a way that is kind and empathetic. . . it's not the best use of our time to yell at other, especially at early career researchers who might not be in a position to implement these changes." Another Twitter user flips the issue around, explaining in a 2017 post that the warmth and openness of the community's culture plays a crucial role in encouraging promotion of open science and reform, for those who believe open science and reform are important. An eloquent post by William Ngiam (@will_ngiam) echoes their sentiment: "Advocating for transparent and open science involves empowering researchers to engage in reproducible practices and lowering barriers to do so, such as by creating accessible tools, collaboration and guides – not to bully and shame them into it." (2021)

How Should the Community Handle Bad Actors?

As one might expect, how the group reacts to and engages with one another on the issue of bullying and misbehavior is a point of heavy contention. As with any other context, from the schoolyard to the workplace, people can address objectionable behavior in two main ways. People can either ignore the behavior, and act as though the perpetrator is not there, or engage with the behavior directly, confronting the perpetrator, even perhaps to the point of actively excluding them from the group. These two approaches are reflected in my findings.

Brian Nosek (@BrianNosek), the person behind COS and someone people widely regard as a major figure in the movement, opts for the first approach: "The behavior is corrosive and counterproductive. It isn't obvious to me whether there is anything to do but ignore it – give it no extra oxygen." (2019) In an interview with me in 2021, he provided some more reasoning behind this approach: "Personally, I try to live by the values, goals, and style that I believe are important and I aspire for the open science community. Given the structure of twitter, I rarely think that engaging directly with misbehavior leads to productive shaping of positive norms and values for the community. There are many exceptions of course. But mostly, I think positive, productive norms are set in that medium by ignoring people that behave poorly so that their misbehavior is not very visible and the actor's behavior is not reinforced by giving them attention." Despite Nosek's tweets, many in the group are dissatisfied with what they see as passivity.

Their discussions on Twitter demonstrate clearly that they want a more active response to bad behavior. Several members of the group have spoken out

against bullies, discussing 'weeding', or calling them out individually for their behavior. As one Twitter user put it, bullies can't just be allowed to hijack the movement – they should be called out directly. Sophia Crüwell (@cruwelli) has been vocal about her thoughts on handling bullies: "We should call people out, but individually (as has been done before). Blanket statements dismissing the entire movement and its practices are just as unhelpful as the bros."; "But someone needs to call out individual bad actors, otherwise it's just smoke? And I've said this somewhere else but I truly do not think that the community is tolerating bullies - people are called out for any truly bad behaviour, and they do see (some) consequences for this." (2018) Another person framed their approach differently, stating that because bullies thrive when they are ignored, their behavior needs to be explicitly and directly addressed. It is more important to condemn the behavior, says the source, than to worry about whether or not bullies represent the community more generally. This perspective provides a counterpoint to Nosek's reasoning for ignoring bad actors.

Problematic Actors- Limitations of the CoP Framework

At this juncture, it is perhaps useful to consider some limitations to Wenger's CoP framework as it is applied to the reform community. Firstly, as Hong and O (2009) point out, Wenger's CoP framework does not thoroughly address conflict and disagreement in communities. To some extent, they write, this is because previous studies have tended to focus on single occupational groups which share idiosyncratic cultures. This is a limitation of the framework, because many communities contain individuals with very different worldviews and identities, some of which conflict. Additionally, power is unevenly distributed across the community.

Hong and O highlight the role that power differentials can play in conflict in communities of practice. In their study, "intercommunity relations were structured and sustained to the advantage of the more powerful group" (2009; p. 322). They argue that dominant groups in the community pursue their own brand of the shared enterprise, potentially at the expense of other, less powerful groups through controlling rhetoric and resources. This has the potential to strongly influence the participation of certain members and sub-groups in the community in the shared enterprise. The reform movement is just like any other community, in this sense: there are more powerful groups and less powerful groups that exist.

This raises another shortcoming of Wenger's CoP for its use in this analysis. Fox (2000) correctly states that Wenger's 1998 framework does not really tackle issues of unequal power distribution (power is addressed in brief footnotes, rather than head on in the main text, writes Fox). The framework handles the issue of power as a part of identity formation, though this is not surprising given the focus on identity in the book. I will briefly discuss power in the reform community through the lens of Bourdieu's social capital theory (Bourdieu, 1986).

Power can mean different things in the reform community, but, as with other communities, it tends to be in the hands of people with the loudest voices, and with the most privilege. Some power can be transferred from the wider academic community to the reform community. For instance, senior white men and women from wealthy backgrounds in academia may be able to step into the reform community and carry some of that capital with them. They might have a 'feel for the game' already (in Bourdieu's social capital theory, this is called cultural capital; Bourdieu, 1990) and may benefit from that knowledge and intuition in the reform community. Power might also be conferred to people who have had a notable role in founding and contributing to the reform movement. For instance, Brian Nosek founded the Center for Open Science (COS) and led the now-iconic large-scale open science collaboration paper of 2015 which reported that more than a third of social and cognitive psychology papers did not replicate. Arguably, he is an actor in the community who wields a great deal of power. In this case, economic capital figures in, as Nosek and his team at COS have been the beneficiary of over 20 million USD since its foundation in 2013. This allows him and others associated with COS to gain influence in the community as they feature on high-profile projects and publications, gain cites from these and recognition from peers. They gain legitimacy in (as well as outside of) the community in this manner.

Interestingly, Nosek is a good example of transferring power from the wider community to the reform community: He already had gained significant social capital not only as a white man, but also from his work developing the now-influential Implicit Association Test. This work led to Nosek and two others (Greenwald and Banaji) founding Project Implicit, which is a large non-profit organization with an international research team which collaborate to meet the organization's aims to educate the public about implicit human biases, and collect data on the topic (see `https://www.projectimplicit.net`).

Yet others are powerful in the community because of the nature of their own social standing, or popularity and their deep embeddedness in the community. Bourdieu (p. 248; 1986) considered someone with a high degree of social capital to have access to "a durable network of more or less institutionalized relationships of mutual acquaintance and recognition". Some in the reform and open science space are well-known figures because of a high degree of participation in the community. They contribute to several of the enterprise's artifacts – tools or practices, articles, workshops as well as the public discourse we see on (for instance) Twitter. That is, they are visible, and become trusted, knowledgeable figures in the community, and gain status and legitimacy in that sense.

Part of the community's activity aims to redistribute power. The bropenscience hashtag has been pegged as a way of drawing attention to, and helping dismantle structural power imbalances in the scientific community. A session at SIPS2021 focused on this issue: "an overview of how science... upholds colonialism and perpetuates harmful patterns of extraction and power imbalance. We will... discuss anti-racist, open science, and community-based methodologies that

can change the way we do science and mitigate these harms." Despite attempts like these, plenty of skeptics exist who argue that open science and the scientific reform movement replicates the power imbalances and inequalities that exist in the traditional research system.

Bad behavior as a Disruptor

Wenger's 1998 CoP framework does not address problematic or disruptive actors, but I can put an analysis of their potential impact into terms of the framework myself. First, gate-keeping and bullying by some actors is likely to affect others' perceptions of belonging by undermining their perceptions of their own legitimacy and centrality in the community, and by making them question whether or not their individual identity is aligned with the group identity and greater enterprise. It could discourage greater participation in the community which in turn could inhibit the community's practice and stymie its progress, if we think about it in the extreme. Ultimately, bullying within the community can push people to the edges of the community, where it does not take much for one to feel like a complete outsider. Wenger discusses peripherality in a CoP, calling it an ambiguous position, and writes that peripherality is a position where actors are kept from moving further inward, toward participation: "Practice can be guarded or available, membership can be daunting or a welcoming invitation, a CoP can be a fortress just as it can be an open door." (p. 120)

The literature on social group dynamics highlights the significance of problematic group members on the characteristics and function of the group. Barrick, Stewart, Neubert, and Mount (1998), for instance, demonstrated that "worst" member's personality attributes were much stronger predictors of group cohesion (and performance) than the "best" member's personality attributes or the group's mean personality scores. In other words, the group's 'personality' is operationalized by the personality of the worst group member. As Felps, Mitchell and Byington (2006) put it, "a dysfunctional member's behavior inhibits essential group functions, processes, and goals." (p. 180) Of course, not all group members feel as though they are part of a team, meaning that the group's cohesion and performance in terms of a movement's progress are unimportant (or, at least, of a lower priority than the individual's goals). Many do, however, and the concern that their community's image is tarnished by the actions and words of some disagreeable members can cause much anguish. One actor posted that they were troubled about the idea of people associating them with a community that includes and tolerates bullies.

7.4 Reform and Open Science: A Constellation of Practices

As I have already discussed, Wenger's CoP framework describes two kinds of co-operative enterprises: communities of practice and constellations of practice. The analysis of this chapter was carried out under the implicit assumption that the reform and open community was a single CoP, and my interpretations reflect that assumption. However, it is my opinion that my analysis of the community can go further – that more detail about the group will surface – if it is treated as a constellation of practices. For one thing, treating the community as a constellation will explicitly address the discontinuities that I have come to believe are an integral part of their structure. To acknowledge that the community is too diffuse and divided by nature may better explain some of the findings this ethnography has yielded, and shed light on interesting and fundamental features of the community that treating it as a single, unified CoP might gloss over.

Open Sciences, not Open Science

My observation that the community under my study is a constellation of practices is reflected in the opinions of some in the community, and many feel that multiple communities exist under the umbrella of 'open' and the reform movement. In a thread about prioritizing values in the community, Gjalt-Jorn Peters made a link between prioritization of values and departing from the idea of a single, unified community: "And I'm increasingly feeling that if a community is to some degree about values/ideals, there are inevitably people who disagree. We don't need one community with everybody; we can have lots of communities, too; and community- less/-fluid people, too." (2020b) He continued: "It leaves more room for diversification, live and let live, etc, where being in any single community doesn't matter so much any more. People can just be as 'communal' as they feel comfortable with, and each community can include/exclude as they deem consistent with their values." (2020a) *S4* described the open science community as not being a single community, but instead being "little pockets of people" with different practical interests. *S3* shared similar thoughts, that the community is diffuse with shifting boundaries. They also suggest that there are sub-communities, and that these might be defined by the platform they use to communicate. Interestingly, they also integrate the element of time, and how the group has moved to various different platforms as it has grown and matured: "I think in the early days there was a lot of stuff on Facebook, and the 'Psych Methods' discussion group and then on Twitter, and so in a sort of public sphere. . . and now I think some of the more interesting discourse is moving to Slacks and direct messages, partly because people are more cautious about what they say publicly.[4]

[4]Slack is a platform that is gaining popularity among academic groups. It is used to coordinate events and collaborations, and acts a little like instant-messaging among the group. Slack channels

So I would say these are sub-communities. So we have the people on my blog, and we talk privately, and what we say is less guarded."

Another person I spoke with informally (Source 5; *S5*) shared the view that there is not just one brand of reform. To this source, people participate in the community in different ways depending on things like their scientific discipline and geographical location, which affect their nature and degree of participation with the community, and how they interpret the collective enterprise. In the mind of another informant, *S2*, as the group has grown and developed since its initial formation, it has become less and less like a single, cohesive community: "It's very loosely defined... I think in the beginning I think it was a very small group of people, then it expanded, and I think the SIPS conference in the Netherlands, was like 800 participants or so, that was the moment where you could really speak of a community. Because everyone was doing it in some way or other. And by doing it I mean, using open science principles." I think it is likely that, as it grows and develops, this group is becoming increasingly more like a constellation. Perhaps, as the group has grown, it has attracted different kinds of researchers, and a plurality of its goals and dimensions has emerged, creating a constellation-like structure with distinct and somewhat independent groups operating in parallel towards the communal goal.

Heterogeneity in the Community

When we consider describing the reform and open science community as a constellation of practices, we open up the possibility of highlighting differences and divisions among the group. Analyzing these differences and what they can mean for the group enterprise becomes fruitful when the lens of the constellation is applied. In the context of talking about problems within the community, Hartgerink shared an analysis of differences between in-group members and what it may mean for the movement's progress: "These differences emerged... they come to the forefront, because as a community starts making progress, there's more internal discussion possible because you have less of an urgency... the low hanging fruit gets taken away..." In Hartgerink's view, we might see divisions and factions in the community as signs of progress. To break from the idea of a homogeneous, unified community to give people space to diverge from a single set of priorities is likely to also give the movement some air to allow it to grow and progress, and perhaps mature.

The topic of the group's heterogeneity has been discussed at length among the group, in the context of its identity. Many have stated that the group is more diverse and heterogeneous than others say it is, and than it appears to be on Twitter (although, one actor did share his concern that the group is at risk of falling into

are accessible to people who have an invitation to join the channel. Some groups make the link publicly available, meaning that the only barrier to participation is having a Slack account and joining with the link, but others limit who can join thereby keeping the group's chat private.

the trap of saying they're diverse without working to ensure the group are diverse in reality). The INOSC Starter Kit documentation explicitly instructs potential OSC-forming groups to begin with a heterogeneous group of initial members: *"In the initial stages of an OSC, attracting members is most effective by word of mouth... Note that it matters a lot who your initial members are. If all your initial members are from one faculty, it will be more difficult to attract members from other faculties at a later stage, because they will think that this community is not for them. The same goes for OS expertise. If you only have OS experts as members, newcomers will be more reluctant to join. Thus, make sure that your initial members are a heterogeneous group."*

Heterogeneity and Productivity

In Wenger's framework, diversity is just as important as homogeneity in terms of how engagement in practice is made productive, and even possible in the first place. Certainly, the reform/open community is an ill-defined group in many senses. They represent most scientific disciplines, are from all stages in the academic career trajectory, are of different genders and racial backgrounds, and hold a multitude of different views on the scientific enterprise and how and why it should be reformed. But because they are all responsive to problems in the traditional scientific enterprise, they are connected by mutual engagement. Let me emphasize here, that what seems to connect all of these CoPs in the constellation (i.e., what makes them a *constellation* – a cluster of visible entities which form some kind of outline or pattern) is their engagement with problems. The energy of each of the CoPs in the constellation is used to solve serious problems in science and academia; that is their *affinity space* (Gee, 2005).

Wenger also highlights that working together in their capacity as reformers can create differences in and of itself. They specialize and focus on different reform practices and approaches to the problems, gain reputations, and distinguish themselves in the community. Each person makes their own niche in the group, which becomes more integrated as well as defined during their engagement. This is why, argues Wenger, "homogeneity is neither a requirement for, nor the result of, the development of a community of practice." (p. 75) A key point Wenger makes in considering heterogeneity, is that peace and harmony are not requirements for, nor the results of a CoP. Observations that *S2* and *S3* made, namely that most cases in which sustained interpersonal interactions take place generate conflict and friction, push this idea further. To them, conflict is nearly inevitable; almost an expectation or a concession to be made.

7.5 Boundaries and Discontinuity in a Constellation

A constellation of practices includes multiple overlapping interdependent communities, which implies multiple boundaries, or sociocultural differences (barriers

and interfaces, even) between practices which lead to discontinuities in actions or interactions (Kislov, 2014). These boundaries, argues Kislov, are unavoidable because they are underpinned by differences between communities: diverging competences, histories, styles, and discourses and, perhaps most importantly, identities. Crucially, we must recognise that such boundaries are the result of *boundary work*, meaning that they are actively constructed. Evidence of boundary discontinuity can be found in abundance in the material my ethnographic work has generated, and can be clearly seen in my analyses in this chapter.

Consider, for instance, differences in how the group interprets the collective enterprise, particularly in its approach to reforming science. For some communities in the constellation (including the all-in actors), the enterprise is all about the aims and objectives of the movement in terms of coming up with concrete practices and tools, and using them (as many as possible) to achieve better, more transparent science. For others, such as the 'buffet' advocates, the idea centers more around bettering science holistically, and the enterprise should target all facets of that task, from diversity and inclusion to transparency and quality. These differences reflect the coexistence of separate 'reform identities', rather than one single collective identity.

The analyses of this chapter reveal evidence of preexisting boundaries within the constellation that are structural in that they line up with different scientific disciplines or traditions. One can also see, however, that boundaries have become erected between groups in a somewhat organic fashion. These boundaries (and the discontinuities, or ruptures in practice that they can cause) have materialized largely in response to disagreements over how bullies should be handled, and what the nature of scientific critique should be. Kislov argues that the structure of a constellation can promote the reproduction of existing boundaries, and I find that a likely possibility in this case because preexisting disciplinary/traditional differences are carried over into the group's negotiation of meaning.

Consider, for instance, the discussion about replication. Replication (and indeed, the lack of replication success), as I mentioned in a previous chapter, is a key element of the open science enterprise. This is problematic for some reform actors, however, because replication is not a concept which is easily 'translated' into qualitative research because of the lack of reliance on quantitative data (and nor should it be, argue some). And yet, many of the tools developed by the reform and open community center on increasing replication success rates, and much of the literature discusses that high levels of reproducibility are desirable, and a mark of a healthy scientific field. Additionally, much of the Twitter discourse I have observed informally advances this narrative. Evidently the divisions between quantitative and qualitative disciplines in academia are reflected, perpetuated or even exaggerated in the reform constellation. Consider too, theory-focused reformers, who tend to argue that things like replication and preregistration should not have the central place in reform that it is usually given.

Forces to Help or Hinder Science Reform

Kislov (2014) explores the duality and ambiguity that arises when we consider the nature of boundaries in the context of a constellation. Kislov argues that boundaries have an inherent duality about how they affect practice across constellations. On the one hand, he says, boundaries can be responsible for "innovation, learning and cross-fertilization between practices" (p. 3). In the way that interdisciplinary research can yield rich and valuable scientific insight because of the combination of methods and analysis approaches from different fields, so too can different CoPs in a constellation learn from and enrich one another. The 'metascience'-focused CoPs, for instance, would benefit from insights from science and technology studies (STS). Their often highly applied approach, with a focus on data, might gain legitimacy from deeper engagement with existing theory.

On the other hand, says Kislov, boundaries can be the cause of "separation, fragmentation, and disconnection" (p. 3). As became apparent to me (and several others) at MS2019, a fracture has opened up between metascience and STS. When the organizers of MS2019 opened the conference with claims of establishing a new field. The conference website stated its objectives as follows:

> During this decade, we have witnessed the emergence of a new discipline called metascience, metaresearch, or the science of science. Most exciting was the fact that this is emerging as a truly interdisciplinary enterprise with contributors from every domain of research. This symposium served as a formative meeting for metascience as a discipline.

The conference website further shared the aim of bringing together "leading scholars" who are investigating issues such as how scientists generate ideas, how they interpret and treat evidence, and what the cultures and norms of science are. These claims needled many STS scholars, understandably; STS has been grappling with questions like this for decades, and it would seem, based on the conference website and program, as though most metascience researchers have not made a meaningful attempt to engage with their work on the topic. The Twitter discussion was fraught with frustration on the part of some vocal STS'ers using Twitter. The responses ranged from musings that metascience was reinventing the wheel or rebooting the Vienna Circle, to confusion over why metascience might set itself apart from STS, to outrage at the idea that metascience was re-branding STS as if STS didn't already exist.

From the perspective of the reform-minded STS-grounded CoP, the metascience CoP opened a chasm between them by not engaging with STS scholarship prior to claiming their new 'territory'. This disconnection between these CoPs within the constellation shows that boundaries can be tricky to navigate and that misunderstandings and slights can easily lead to fragmentation.

This implies that depending on what forces are in play, differences between groups can be productive (e.g., by cross-fertilization, in Kislov's words) or a hindrance (e.g., by fragmentation) to the shared enterprise. Several forces vie for control in this constellation. Some forces are neither 'good' (such as ones which are constructive, focused on building community inclusivity, for instance) nor 'bad' (e.g., destructive ones, such as bullying). Most of them are complex in practice and can lead to either discontinuities or connections between communities in a constellation depending on a number of factors.

One dominant force to consider is the Twitter platform itself. Although Twitter provides a platform for much informal exchange between co-participants (which should help establish continuity within the constellation, implies Kislov; 2014), the character limit of tweets does not allow for nuanced or lengthy explanations of perspectives, which can fuel tensions, and contribute to the widening of ruptures in the fabric of the constellation between communities.

We also must consider boundary objects and interactions (Star & Griesemer, 1989). Boundary objects are artifacts, processes and concepts that bridge boundaries, and can be used by communities in collaboration with one another. Star and Griesemer discuss these in the context of the various organizations and bodies present in the scientific enterprise. They explain that scientists have to grapple with keeping important information and its meaning intact in the face of the great diversity represented in science (p. 388, *ibid*).

Wenger noted that boundary objects are malleable and vague enough to function in different contexts and cultures, yet robust enough to allow communication *across* boundaries. Examples of reform constellation boundary objects are concepts like validity and scientific integrity. These concepts can mean different things depending on their application. For instance, validity is a highly malleable concept, varying in its meaning depending on the discipline applying it. Validity is assessed and achieved differently for qualitative versus quantitative studies, however it is easily recognizable as a concept across all of science. Boundary interactions can take the form of discussions, meetings and cross-boundary projects. Conferences such as AIMOS and Metascience are examples of reform constellation boundary interactions.

The reform constellation is rich with boundary objects, and I posit that this is because each of the communities in the constellation are deeply rooted within the same concerns: of the poor state of the scientific enterprise. I argue that the crisis of confidence has been a common starting point for many if not most subcommunities under my study, and that even though the communities disagree about approaches to reform and rules of engagement, they agree that *something* must be done to improve science, going forward. Whether it is to focus on improving how we develop theory and link it to empirical study, or to make all research material open, or to fight questionable research practices through the use of checks, systems and software. This is a good example of a boundary object. Everyone agrees that *something* must be done (which is the core of the boundary

object) but different groups can and do interpret this *something* differently, and have different competencies to apply when trying to resolve it. Boundary objects serve to coordinate the perspectives of the communities in the constellation, and reification of the joint enterprise is at the heart of this.

Boundary objects can act as barriers to progress, though, Kislov cautions, if they are used to reinforce established power structures, or solidify actor status. It is possible that this occurs to some degree in the reform constellation. Coming back to an earlier point about the emphasis on replication, for instance, it is likely that some boundary objects are used to perpetuate power differentials among CoPs. Historically, replication has been prioritized in the movement. Replication, in the minds of many prominent actors, is an important tool for diagnosing flimsy effects in the literature, and as a test for new ideas before they are released into published record. But, it is not always easy or possible to apply to all disciplines. It is malleable to a degree – LeCompte and Goetz, among others, talk about replicating qualitative research (LeCompte & Goetz, 1982) – but it is much more easily achieved in quantitative research studies. In prioritizing replication (through funding and rhetoric, for instance), positivist fields are prioritized over those focused on interpretation, and can maintain their place at the top of the hierarchy in science.[5] Those who happen to study and perform replications can gain capital over those who are in, say, ethnographic or phenomenological traditions.

Boundary interactions can generate and reinforce connections between people from different CoPs, and they are plentiful in the reform constellation as is the case for boundary objects. This is perhaps in part because the reform constellation is situated within the broader context of academia, and boundary interactions like conferences, meetings, visits and projects occur between different academic CoPs all the time. Conferences are a good example of this, at which interdisciplinary and inter-institutional meetings between scientists occur. The emergence of new boundary practices, during such interactions, is accompanied by the transformation of identity and meaning, and requires continuous collaborative work across boundaries, says Kislov. It is unclear to me whether different CoPs within the reform constellation are willing to exert the kind of effort that the negotiation of joint identity and meaning between CoPs would demand.

Clear barriers to the fruitfulness of the joint enterprise exist: take trust and respect as examples. I have observed mistrust and lack of respect between some groups in the constellation, in the way that people refer to others, and in the way that people wield power. As a concrete example, some actors choose not to cite others' work even though it is clear that they are aware of it, and that it should be cited. *S1* spoke to me about this, finding that the lack of willingness by some

[5]This is in comparison with qualitative research which has long been relegated to the 'lower rung' of the quality ladder (Sandelowski, 2000). An attitude due, at least in part, to perceived problems with reliability, validity and generalizability (LeCompte & Goetz, 1982), and negative attitudes towards subjectivity (Maykut & Morehouse, 2002; Nau, 1995).

in the reform space to appreciate the scientific labor of others, and to gauge its scientific impact properly, is a pervasive problem in the reform literature.

Another example is incentives, or a lack thereof. From what I can tell, no obvious incentives exist to establish and reinforce cross-boundary co-operation or knowledge-sharing. In organizational contexts (which are usually studied in the CoP literature), systems to facilitate sharing can be set up and resources funnelled into such efforts. Because the reform constellation spans the globe and is not embedded in a single institution, such a thing is difficult. Sub-communities are not incentivized to bridge the ruptures in the constellation, and so it is more straightforward to see differences than to work to realize similarities.

I have so far discussed two kinds of 'boundary bridges': boundary objects and boundary interactions, but as I outlined in the previous chapter, Wenger's framework refers to a third kind: brokers. Brokers are actors who hold membership in multiple CoPs in a constellation, and act to facilitate interactions and coordinate practice between those CoPs. In a sense, brokers themselves are boundary objects, because they are flexible in their identity as they must renegotiate it depending on which community they are participating at a given time. That flexibility is manifested in brokers' needs to have an understanding of each CoP in which they are members, as well as legitimacy as negotiators and interest in the role. Informal brokering goes on within the reform constellation, as people straddling different CoPs attempt to navigate their different identities, and broker connections between them.

I have observed what can be interpreted as brokering in discussions on Twitter. Brokers often act like mediators in 'arguments', deescalating debates, and smoothing over friction and tension between actors. I have been able to 'tell' who some of the reform constellation brokers are by the fact that their perspectives are entertained, or 'listened to' by the different arguing parties. Arguments on Twitter often end up in people disengaging from the debate or blocking one another, but brokers are sometimes able to maintain the discussion and bring about some kind of truce, at least in some cases. This is because they are taken seriously by both sides, because they have legitimacy and authority in both groups.

In the previous chapter, I discussed Stavroula Kousta as a concrete example of what I argue is a broker between two clearly distinguishable CoPs. As the editor-in-chief of *Nature: Human Behavior* (*NHB*), she has a defined leadership role in a CoP which represents the traditional science enterprise the reform movement is trying to change. Naturally, this may categorize Kousta as the enemy, or at least, not 'one of us'. Despite her role in *NHB*, she is interested in engaging with open science practices and takes time to attend science reform conferences and workshops (like, for instance, MS2021 and the Lorentz Center Perspectives on Error Workshop of 2021, at which she was a speaker on publishing reform).

Her interest in the role of brokering is, to me, obvious. She penned the following in a 2022 *NHB* feature article on highlights of the first five years of the journal's existence: *"When I took on the role of launching Nature Human Behaviour,*

I wanted to create a multidisciplinary journal that stood for rigorous science that makes a difference in the real world. Over the years, we've published several important research papers that embody this vision. [...] Two Perspectives helped to set the tone for the journal from our very first issue. "A manifesto for reproducible science", the product of a collaboration among ten metascientists, has become a textbook reference for the open science community since its publication. The manifesto went beyond enumerating the ways in which science has been failing to distil steps and initiatives required to support credible science. These views and ideas are at the heart of the journal's identity and have become key in the transformation of a reproducibility crisis to a credibility revolution in science." (p. 12; Antusch et al., 2022)

That Kousta chose to highlight what is now a somewhat iconic paper for the reform and open science literature (already cited over 1900 times at the time of this writing, in April 2022, and authored by a range of high-profile reform actors including Brian Nosek, Dorothy Bishop, Chris Chambers and Eric-Jan Wagenmakers) shows her interest in the topic, and that she wishes to act as a conduit between the reform community and her *Nature* outlet.

8

Network Analysis

8.1 General Introduction and Challenges

What is the open science and reform community? How can we delineate its bounds; that is, who are members of the community, and who are non-members? What are their characteristics? How are the connections between the group structured? Are there clear, cohesive subgroups, or is the group generally cohesive?

These questions are central to the work this dissertation presents, but finding answers has been a challenge. The root of the difficulty lies in the fact that no single *formal* community exists which encompasses everyone who has an interest in or practices open science, or who is part of the scientific reform movement. They are not like a football team or the staff of a supermarket, who wear physical evidence of their affiliation to help people identify them. They are not registered on an online roll, or in a database on a server somewhere, and they do not meet regularly in the same place. They are spread across the world, occupying every continent, have representatives in most universites and research institutes. Many people choose to openly identify with the group, and do so in various ways, while others prefer not to identify with the group at all and reject others' attempts to associate them. In other words, delineating the community, or determining what the 'sample' is, is difficult.

I attempted to work around this issue of delineation through in-depth study of the possible open science and reform group, approaching the study from qualitative and quantitative angles. In this chapter, I explore a few possible facets of the online community, using a network analysis approach to describe patterns of connections among some actors within the open science and reform space which are captured by Twitter. I consider the findings in the light of the conclusions of the previous chapter: The complexities and structure of the open and reform community can be most easily revealed, and their activity analyzed when they are considered a *constellation* of communities of practice, rather than a single community of practice.

It is also important to mention that the research I describe in this chapter is highly exploratory. I do not attempt to test any hypotheses or investigate explicit research questions using the data I describe here. As a researcher who has been thinking about this community and its possible structural configuration for some time, I naturally have my own ideas and expectations, although I refrain from formally explicating them. I opt to interpret patterns in the data instead, and link them to other observations and findings in this dissertation where appropriate. Second, the population parameters relating to the group of people under study are unknown. As I discussed previously, communities – particularly those online – can be notoriously difficult to delineate. This is the case for this community (or constellation thereof), and we carry the limitations this implies with us throughout the following descriptions and analyses. I attempt to grapple with the identification of a *sample* of the open science, metascience and reform presence online, with the full knowledge that the validity of this attempt hinges on the methods I use to collect my data. I explore this dissertation's limitations in detail in the final section, where I underline them in relation to my findings.

8.2 Introduction to Social Network Analysis

Data are everywhere. We generate data constantly, even when going through the most mundane motions of everyday life: when we note down a callback number left on a voicemail, scribble a unintelligible shopping list on the back of an old cashier's receipt, leave a post-it note to a spouse reminding them to get things for dinner. As we shop online, playing the role of the consumer, we generate chains of different data. Our purchases will also train Google's targeted advertising to market similar products and services to us. With the internet, these data can be generated and transmitted in less time as it takes to take a deep breath and regret that last impulse purchase.

It is not only in the role of consumer that we contribute to big data. In an age of user-generated online content and online social media, the lives of most people are heavily influenced by online activity. This includes social interactions, which involves the use of online social media platforms. One such platform is Twitter, one of the most visited websites since 2013. Twitter is a tool which allows for the generation of great amounts of data. As I mentioned in **Chapter 3**, approximately 500 million tweets are posted by the platform's users per day. In addition to being a repository for a huge archive of tweets, Twitter also retains demographic and geographical data for each of its registered users.

A third kind of data handled by Twitter is in the form of social ties. As I will discuss later in this chapter, a Twitter user can 'follow' and be followed by other users. When you follow someone on Twitter, their posts and replies are put into your timeline feed (which displays a personalized stream of content based on the posts of people you follow and suggestions made by Twitter's algorithms), and

when they follow you, the content you produce appears in their timeline. People on Twitter follow each other for a variety of different reasons, including: to share information and support, to facilitate networking and scientific collaboration, for educational reasons and to develop social capital. By accessing Twitter's enormous database via its application programming interface (API), anyone is able to 'mine' information about these ties, and use that information to describe links between Twitter users in the form of a social network. While the previous chapter focused on defining and describing the open science and reform community, this chapter describes a more quantitative exploration of the structure of part of this community on Twitter's platform, using social network analysis (SNA) to explore the data I scraped from Twitter.

Background

Where there are humans, there are connections. Familial ties, friendship ties, cultural and racial ties. Each person is embedded in a set of networks, which change and evolve throughout the lifespan. Some of these networks are largely set, such as families into which we are born. Others are constructed by our participation in certain activities such as sports and social media sites like Twitter. All of these networks, no matter the nature of the ties, can be measured, analyzed and visualized with network analysis tools.

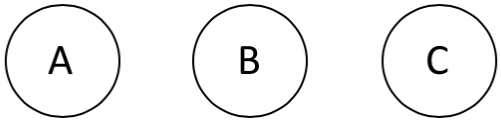
Freeman's 2004 text locates the roots of social network analysis (SNA) in the late 19th century work of Simmel and Durkheim, who highlighted the value of studying patterns of relationships between people. It wasn't until the 1930's, however, that researchers began to develop analytical methods to study network data. Since then, the popularity of network analysis has greatly increased. SNA is a valuable approach to quantifying the dimension of social capital (Rehm & Notten, 2016) – describing and analysing the nature of the ties between people. Freeman notes that social network analysis is driven by the "notion that the patterning of social ties in which actors are embedded has important consequences for those actors." (p.2; 2004) Network analysis methods are used to reveal the patterns underlying these social ties and attempt to define when and why the patterns arise, and in some cases assess what the consequences of them are on the community (and, of course, the individuals it is comprised of) under study.

With the guidance of Wasserman and Faust's canonical book (Wasserman & Faust, 1994), I will now introduce important concepts in SNA theory and practise which will come up throughout my exploration of the structure of the open and reform community. Key terms are indicated by the use of boldface type. I will not provide a complete catalogue of SNA concepts, rather I will provide what information is relevant for the community under study. Please note that the following explanation of SNA's background is simplified and conceptual. SNA is associated with a large and well-established field of study; I only skim its very surface in this chapter.

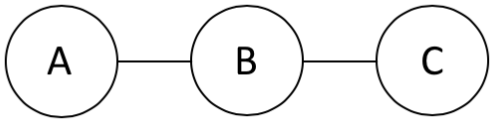
8.3 Terminology and Concepts

Network Elements

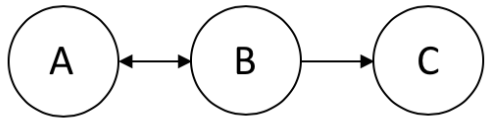
Beginning at the level of the individual – the most basic level of analysis – are **actors**, also known as **nodes**. Although in this context the word 'actor' refers to a person (within the community), in some applications and for some levels of analysis, the word can have different meanings. When we speak of relationships or shared preferences or experiences between two or more actors, we speak of **ties**, also known as **edges**. In the figure below, A, B and C are actors.



A tie between actors (two, which forms a unit known as a dyad or pair, three, which forms a triad, or more) might indicate that they know one another. Ties can be undirected or directed. An undirected tie indicates that the relationship means the same thing to both actors, for example, that they coauthored a paper. The figure below depicts an undirected triad.

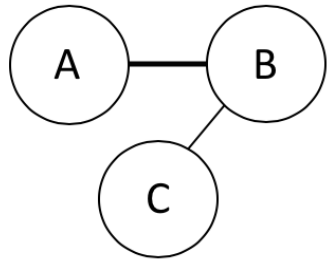


A directed tie indicates that there is the possibility for some imbalance in the relationship. For instance, on Twitter, ties do not have to be reciprocal (unlike on Facebook where two actors must agree to be friends before a tie can be formed), meaning that a user can follow another user without being followed back by that user. Directed ties can be uni- or bi-directional. Actor A and B follow one another on Twitter, so the directed tie between them is bi-directional indicating that it is equal. In comparison, B follows C on Twitter, however C does not follow B so the tie between them is unidirectional. Usually, directed ties are represented in network visualisations by lines tipped with arrows; undirected ties without, as the undirected graph *above* depicts.



Zooming out one more time gives us a **network**, or **graph**, which refers to a group of actors and the ties between them. Networks are usually either directed or undirected, depending on the kinds of ties between the actors within them, however other more complex networks exist.

Some networks contain information about tie strengths, or **weights** too. For instance, consider an academic department at a university. Researcher A and B in the figure collaborate frequently together (indicated by the thickness of the edge joining them), however Researcher C has only collaborated once with B and never with A. Note that the triad depicted in this example is undirected. This is because collaboration is always reciprocal by nature. In this example, it is not possible to collaborate with another researcher on a paper without them also collaborating with you. Weights can indicate different things, depending on the network in question. For instance, it can provide information on the extent of the strength of a relationship in terms of intensity ('best' friends, versus acquaintances), or the frequency of contact between people, or, in the case of the example above, the frequency of scholarly collaboration.



Network and Node Properties

The analysis of network data involves the calculation of three categories of statistics: those which focus on the whole network, those which focus on the individual nodes within that network, and those which focus on dyads, triads and subgroups. I describe these in turn.

Network Properties

Networks tend to be explored first in terms of their overall characteristics, and there are many metrics which can be calculated on whole networks. Analysts consider, for example, network size and density. A network's size is quantified by how many nodes are present. Density is defined as "the extent to which links which could possibly exist among persons do in fact exist" (p. 2; Marsden, 1993). The density coefficient, which quantifies how many of the possible ties within a network exist has a possible range of 0 to 1, where 0 would describe a 'network' of isolates, and 1 would indicate that every user in the network was following and followed by every other user.

Analysts also calculate network diameter and average path length, which are used to measure the network. These two metrics use path length, which is the number of edges that the shortest path between two given nodes contains. Diameter is calculated as the "longest distance between any two nodes in the network" (p. 1; Takes & Kosters, 2011). This concept describes the 'length' of a network, based on the distance between the two most far-apart nodes. To apply this idea, a low diameter in a network of co-authors might indicate a research group (Gaskó, Bota, Suciu, & Lung, 2020).

A closely related concept (it is closely related in that path lengths are used in the calculation of the diameter metric) is average path length, calculated as the sum of the shortest path between all pairs of nodes, divided by the total number of pairs (Macià & Garcia, 2017). Conceptually, this tells us, on average, the number of steps it takes to get from one node in a network to another. An average path length of 3 in the current context indicates that if one user wanted a piece of interesting information to get to another user, it would on average have to go through two other users before reaching its destination.

Lastly, average degree is a commonly calculated network metric. It refers to the average number of edges per node in a network, and tells us how many social relations people have on average with others in a given network (Wasserman & Faust, 1994).

Node Properties

When network analysts describe and analyze networks, they use a set of metrics to numerically quantify the position and influence of nodes within the network, and the different aspects of engagements of a group's actors with one another. These metrics give insight into patterns which indicate structure, such as hierarchy, and can indicate the presence of cliques (defined as a network subset in which the actors are more closely tied to each another than to other actors in the network) within the greater community.

A basic element to describing a social group concerns how embedded or immersed actors are within their network. This is also an important concept that

network analysis grapples with, and it is referred to as **centrality**. A central actor, or node, is defined by Wasserman and Faust as one which is "involved in many ties" (p. 173; 1994). Centrality refers to the different measures of the importance of an individual actor within a network. Borgatti, Everett and Johnson (2018) distinguish two key measures of centrality. **Degree centrality** which is the number of connections of one node, and **betweenness centrality** which is the number of times that any given node falls along the shortest path between any two other actors.

In a directed network, degree centrality is further broken down into in- and out-degree. **In-degree** refers to the count of *ingoing* ties to a node, while **out-degree** refers to the count of *outgoing* ties from a node. In the context of the directed Twitter follow network, a user's in-degree counts the amount of followers that node has *in that network*, and out-degree the amount of people they follow *in that network*. Note this is a different figure to the followers/friends, which refers to how many *total* followers and friends someone has. Take the example of the Twitter account of the Open Science Framework (@OSFramework). In the follow network I will soon explore, this account has an in-degree of 965 and an out-degree of 169 (it is followed by 965 other users in this network, and follows 169 others). It has a degree of 1134, which equals the total of all in- and outgoing ties connecting this node within the network (i.e., degree = in-degree + out-degree). At the time these network data were collected, @OSFramework had 33,510 *total* followers on Twitter and was following 1,259.

Determining which nodes are most central in a network helps us identify and analyze well-connected actors and describe aspects of being well-connected in a network. Central nodes are in a strategically advantageous position for a number of reasons. They dominate links between others, may experience a high level of exposure (e.g., to information shared in the community) in the network and/or may have relatively high degree of influence over other actors. They are also more likely to be influenced by others (have high out-degree scores), and/or to be more popular (have high in-degree scores). This is because they are more likely to be exposed to the thoughts and information others share than isolated nodes.[1]

On the opposite end of the spectrum to central actors are those who are peripheral. These are actors with few connections to the rest of the network, and rely on contact with central actors for access to others. In extreme cases, we talk of isolates, actors with no connections to the rest of the network (more formally, an isolate is identified if the in- and out-degree indices of a node are zero).

Continuing to explore centrality measures, we come across two key concepts which relate to how individual nodes are placed structurally within the network. **Betweenness centrality**, which I defined earlier, is a network metric which, in the

[1]Note that all of this depends on whether a node has high centrality due to in-going edges or out-going edges. That is, if an account has a high betweenness centrality score because it is has many followers, it will not have much exposure to information within the network. In this way, whether or not a network is directed or not matters in the interpretation of centrality.

context of social network data, gives a sense of how strong actors' interpersonal ties are within the network. Actors with high betweenness centrality (provided that they have high in-degrees, high out-degrees and are very active) are thought to have influence over the flow of information within the network, and their removal tends to be highly disruptive to the network's structure because removal would break a great many ties and leave structural holes in the network. In practice, this can disrupt information flow across the network (Yap et al., 2011). For social network data, the flow of information is a very important concept. For the current topic, this is especially salient, because the flow of information affects how quickly rumours can spread on Twitter (about cases of fraud, or new reform initiatives, for example), how effectively new practices are adopted, or how well groups of scientists can work together with one another (Takes & Kosters, 2011).

Concretely, as I mentioned before, betweenness centrality is calculated as the number of times a node is present on the shortest path between two others, and highlights users which are in a bridge or broker position between subcommunities in the graph (assuming that high betweenness centrality is associated with a high in-degree, a high out-degree and high twitter activity). So, in the context of the Twitter follow graph I describe in this chapter, I use the example of @OSFramework. This account's betweenness centrality is 153,683, meaning that the @OSFramework node intercepted 153,683 paths between other dyads (i.e., pairs of nodes) in the network. It has both a high in-degree as well as a high out-degree, indicating that the interpretation of high betweenness centrality indicating brokerage is likely applicable.

A related network statistic to highlight in the context of the network under study is that of **eigenvector centrality**. This statistic, like betweenness centrality, denotes influential nodes, though the calculation produces coefficients in a more sophisticated manner. Golbeck explains that eigenvector centrality measures the importance of a node, at the same time giving consideration to the importance of neighbouring nodes (2013). That is, a high eigenvector centrality score is associated with nodes that are connected to others who *themselves* have high scores. Ties with high-scoring nodes will be weighted more in the calculation of another node's score than will ties with low-scoring nodes.

Dyad-, Triad- and Subgroup-based Network Properties

Other parameters for a network concerns relations between groups of actors (from the dyad level of analysis upwards). For instance, the extent to which people in a network reciprocate ties – **reciprocity** – is of great interest to network analysts. The reciprocity coefficient is a ratio of the number of edges in a graph which point in both directions, to the total number of edges in the graph. It has a possible range between 0-1. For Twitter following activity, calculating such a statistic will give us information on which proportion of the follows are mutual follows in a given graph. In a friendly, close network, we would expect

that, by far, most follows are mutual follows. Surprisingly, the same holds for a network which has warring factions when those factions have close friendships *within* them and no relations *between* them. An unbalanced, hierarchical network where a lot of users do not follow their fellow network members, on the other hand, will have low reciprocity. On a global level, reciprocity can be seen as a measure of social solidarity in a community or in its sub-communities. Yoon and Park (2014) write that reciprocity can be evidence of a social ritual of following behavior.

Transitivity is another important statistic to calculate, especially for social networks which tend to have high values of transitivity. Transitivity is a fraction of all possible triads that are actual *closed* triads, or triangles. To give this explanation some concreteness, consider, as in the example above, scholarly collaboration. If a network consisting of researchers in a particular field or discipline were to produce a transitivity coefficient of 0.20, it would mean that there is a one in five chance that two academics would have a collaborator in common. This metric is another way of getting a sense of how tightly clustered a community is internally.[2]

Homophily in a network is defined as the tendency to form ties with similar others, or proximity on 'individual level' variables (attributes, preferences or personal qualities, for instance). Homophily induces reciprocity, and in some cases can be induced by the structure of the network. Different levels of depth of homophily exist in SNA, described as **value homophily** (where individuals are linked by meaningful similarities like belief systems, for example), and **status homophily** (where sameness exists on a superficial level only). A related but distinct concept is **propinquity** which refers to geographical or physical proximity. Propinquity often facilitates interactions (and eventually relationship, or tie formation) between actors.

SNA in the Context of Twitter Data

The introduction of modern computers (with their incredible computing power) and the internet has made mining for and analyzing the data of large and varying kinds of networks possible and relatively easy. Until the last few decades, people have tended to apply SNA tools to the study of in-person communities, though with the introduction of the social web has come a new avenue for study in the form of social networking platforms such as Twitter.

SNA methods originally developed for the analysis of in-person networks are largely appropriate for application to social media network data, though there are nevertheless some translation issues (especially as SNA applies to Twitter data) to be noted. On a basic level, SNA tends to describe the qualitative aspects of networks. In many contexts, ties indicate friendships with others. On Twitter,

[2]Note that the meaning of transitivity becomes harder to grasp in case of directed graphs, like Twitter follow graphs, however.

as mutual following is not required (unlike other sites such as Facebook, as I mentioned earlier), and ties are often forged to facilitate information exchange (not just between friends or as social support), it is difficult to know how to categorize ties between people beyond saying that they're mutual followers on Twitter. This leads into a related issue involving multiplexity (which refers to the different kinds of connections that might exist at one time between two people; Bliemel, McCarthy, & Maine, 2014). When people's relational spheres overlap (when people are friends but also share a tie of a different kind, such as working as colleagues in the same university department), the different types of ties between them can be hard to disentangle, especially concerning Twitter network data which contains a degree of multiplexity, though little information about different ties can usually be found. For instance, although many academics indicate which university employs them in their Twitter bio, it is difficult to detect whether some of these people are also friends. Moreover, while some people do provide information about affiliations and geography in their bio, it is not standard or required, nor is it verified or always accurate, and therefore does not provide reliable information on which ties can be categorized. I discuss the concept of the Twitter biography in detail in **Section 7.5**.

Another point to consider is the ending of ties. In person, people's friendships with others change as a function of time, often to the point where they discontinue. If a person reports their ties with others, they are unlikely to report broken ones unless specifically asked about their history with that person (but they will not usually report a broken or non-existent tie). In contrast, social media network data often does not faithfully reflect broken ties. Although unfollowing on Twitter is used (liberally by some users) as a way to indicate the wish to cut a tie with another person, it does not always happen. You do not get any kind of notification when someone unfollows you either, so even if you wished to reciprocate the unfollow, effectively completely (rather than asymmetrically) severing the tie, you would need to know that that person had unfollowed you through other means (such as them telling you, for instance). People who are not very active on Twitter might 'slip under the radar', so while their tie with someone on Twitter no longer exists in actuality, they might forget to unfollow them (or the other person might forget to unfollow). Blocking is another way to cut a tie with someone, however, information about who has blocked whom is private and so can't be mined for our purposes.

General Methodological Challenges

Some methodological issues that apply generally to SNA also apply to Twitter network analysis. One such example is that of boundary separation: the question of "Who belongs in the network and who does not?" is difficult to answer. Indeed, it is difficult to answer with respect to many in-person networks, as the ethnographic findings I presented in the previous chapter highlighted. Twitter data are

no different. The problem of defining the boundaries of the community, which I have already considered earlier in this dissertation, exists also in this context, as there is no single index or indicator of a person's membership or affiliation with the open and reform constellation of communities. I have made choices about whom to study (I will describe and justify my choices later on when I describe my methodology), but it is with knowledge of the boundary separation issue that I apply caution to my exploration of the network data presented in this chapter.

Other methodological challenges which can affect in-person SNA are less relevant for Twitter network data. For instance, consider problems relating to informant accuracy and reporting biases. As in-person network data is often produced when an actor is asked to report their ties with others, the integrity of the data is contingent on the memory and trustworthiness of the informant. People tend to forget about lost relations, and to bias towards expansiveness and making themselves seem more socially attractive (Feld & Carter, 2002). These problems are largely circumvented in Twitter data because ties are not reported, rather they just are (i.e., people follow others, which establishes a link that normally persists without the person having to remember that they have made it). For the same reasons, people cannot really misreport their ties with others; we either follow people or we don't; we are followed or we are not.

8.4 Data Sources and Approach to Data Collection

As I mentioned before, the work in this chapter is based heavily on the open science/science reform follow graph, which I obtained by modeling the following-relationships which exist between Twitter users who have chosen to include certain keywords in their Twitter profiles (see **Section 7.5**). Additionally, I describe a network of likely open science community members, gathering network data in real time from Twitter using the search query #openscience for a period of two weeks over December 2020/January 2021. Finally, I describe a network based on the activity of Twitter users who 'attended' and discussed on Twitter the 2020 iteration of the Interdisciplinary Meta-research and Open Science conference (AIMOS2020), during the period in which it was held (in the first week of December, 2020). For the exploration of these networks and a discussion of my motivation for choosing them, see **Section 7.6**.

As this chapter focuses on Twitter user networks, the data collected centers around Twitter users which are potential open science and reform community members, and their tie-formation patterns (i.e., their followers and friends).[3] Ideally, only information on 'real person' or personal Twitter accounts would be col-

[3]The word friend in the context of Twitter follow graphs is different to the word friend in typical usage. A Twitter 'friend' is someone who you follow, while a 'follower' is someone who follows you. At the risk of confusing people, I will use the word friend in this context to refer to accounts a user follows from here on in this chapter.

lected, but there is no easy way to filter out bots and organizational accounts. Bot accounts are automated accounts which, in this context at least, are programmed primarily to function as information amplifiers. They tend to follow a large, specific corpus themselves and, depending on their programming, search for posts containing news items or publications and preprints and retweet them. As bots are easy to program, even for beginners, they are ubiquitous on Twitter and are sometimes very difficult to detect for humans. To my knowledge, there is no way to reliably filter them out using automation either – any such process would be manual.

Organizational accounts represent corporations, institutions, agencies, news outlets and 'common interest groups' (Oentaryo, Low, & Lim, 2015). These too, are ubiquitous on Twitter. They account for about ten percent of Twitter users, and are usually run by employees or members of the organization in question. In this corpus, such accounts are usually tied to research or academic institutions or common interest groups. For instance, many regions and countries around the world host informal, or 'grass-roots' open science communities (OSCs), many of which run Twitter accounts as I mentioned before. As with bot accounts, organizational accounts cannot be automatically filtered out, though they are easy to identify manually, as they usually identify clearly as organizational accounts in their screen names, profile bios and often include an organizational logo in their profile picture.

Although including bot and organizational accounts in the networks I explore is not necessarily a threat to the validity of my methodology, their inclusion may introduce noise into the data and my descriptions, as the behavior and thought driving their tie formation behavior (as well as that of their followers) is difficult to understand and define, and likely the result of different social mechanisms than personal user accounts on Twitter. In the results section, I will present data 'cleaned' of such noisy accounts, as well as the full data set.

There are multiple ways to collect data from Twitter. One can access Twitter through using its API directly, which allows one to interact with Twitter's backend, and request data from its servers. In this chapter, I do this via Python to obtain the main follow graph of the open/reform community. I go into detail about this approach in **Section 7.5**. Other approaches involve the use of applications and software which handle making calls to the Twitter via the API (and handling the retrieved data) for you. I use one such program in this chapter (Gephi, which I will also discuss in **Section 7.5**) in order to obtain data for a very specific purpose.

8.5 Open, Metascience and Science Reform Follow Network

In this section I analyze the first set of network data I retrieved from Twitter. The goal of the analysis I present is to explore the network of following-relationships

between Twitter users who are interested in the topics relating to open science, metascience and science reform, as indicated by the keywords in their biographies.

Biographies and Keyword Searching

Biographies (bios, for short) are user-defined text blocks, located on a user's profile (or 'home') page. They are limited to 160 characters, and function as short 'about me' or personal blurb. Academic Twitter users typically use their bio to describe their academic activity and position at their institute, philosophies and affiliations, and sometimes nationalities as well as other personal characteristics or identifiers, and emoji.

Bios are an important part of fashioning one's online persona, and what words people choose can potentially reflect interesting things about them and how they connect with others online. I make such an assumption, to some degree, by using bio-related keywords to collect network data and to capture characteristics of sub-communities of users qualitatively. Despite the importance of how people write their bios, and the fact that Twitter has existed for some time, relatively little research is available on the topic. When others such as Grandjean (2016) and Grant, Moon and Busby Grant (2010) have used bio keywords to explore groups of Twitter users they have glossed over the issue of why people might write what they do, and how the information given in bios might function to categorize them into social groups. Kohana, Okamoto and Kaneko (2013) also discuss categorizing people into groups based on their biographies, stating that Twitter uses bio keywords to suggest new people for other users to follow. They take a technical approach involving clustering to categorize users, however, and also do not address the issue of what motivates users to use the words they do, and how carefully people craft this part of their online identity. Another study focused on analysing users' bios to characterize self-description on Twitter, without going in to the word choice process.

Early in my fieldwork, I noticed that many open and reform community members use the bio to self-describe as open science practitioners or advocates or 'science reformers'; some use specific keywords relating to open science, reproducibility, reform, and metascience to do so. This is not surprising, given that for many reformers, keywords like these describe topics of interest or of expertise. Some people seem to have taken care to write their bios, thinking hard about how perceptions of them are influenced by this short text while others write very basic factual bios, including just their affiliation and a few hobbies. On March 27, 2022, I asked people how much time they spent on crafting their Twitter biographies using a Twitter poll. The poll, answered by 156 users, indicated that around 14% of people said their text was "carefully thought out", while approximately 43% answered that they typed "just the basics". A third of responders reported to have given their biographies some thought. Two of the replies to the poll clarified

that they change the text from time to time to reflect what they are doing and how the text represents that, and that they change the text to reflect their mood.

Clearly, users who include hashtags in their bios are doing so purposely, to explicitly associate themselves with open science, reform and related practices. At the same time, I speculate that others who do not include these hashtags have a wide variety of reasons behind their choice (they might do it to *dis*associate themselves from the open science, they might not think it is important, they might not have thought about it, etc.). I will discuss the issue of keyword use, community proxies and group membership in more depth in the discussion section of this chapter.

Although there is precedent for defining the OSC through bio keywords as I mentioned above, it is only a proxy for open science and reform community membership and has limited utility. This is because the method relies on community members including the keywords in their bios, and will necessarily miss those who do not. For instance, Simine Vazire – arguably a very influential science reformer, and a very active Twitter user – does not include any of the keywords I used in my actor-selection process (I list these in the paragraph below) in her bio and so will not appear in the list of people originally pulled from the API.

I have identified a set of keywords based on hashtags and commonly used terminology in the in the group, which I used as queries in calls to the Twitter API, and collect user data.[4] These are: *openscience*, *open science*, *open research*, *reproducibility*, *reproducible science*, *reproducible research*, *metascience*, *meta science*, *metaresearch*, *meta research*, *Center for Open Science* and *science reform*.[5] The last two keywords were added at a later date because it became clear who explicitly self-identify as 'reformers', and people who are associated with the Center for Open Science (COS) were not included in the first run.[6]

The Twitter user information of each person whose bio contained any one of these keywords on the dates of the searches was added to a list of over 2900 users.[7]

[4]The Twitter API allows a user to request data about other users and their activity based on parameters and queries set by the user from Twitter servers. For the purposes of the description and analysis I present here, I used the Twitter API's `search_user` method. API calls using this method permit an authorized Twitter user to retrieve information about other Twitter users, based on the text of their personal user account biographies.

[5]To the uninitiated this keyword looks funny, as 'open science' in both spoken and written form should contain a space, however hashtags cannot contain spaces as a rule, and so people often use the hashtag version of the keyword in their bio. For making API calls of this nature using a programming language, queries require the passing of exact strings as parameters, and so the code I used to collect my data will show these as 12 keywords, even though there are only really ten.

[6]The COS is the first and biggest organization of the open science community, the work of whose founders and employees is central to initiatives in the open science community (like preregistration, preprints in psychology, and data and materials sharing)

[7]The code and initial data-set for the first ten keywords listed date from early December, 2020. However I re-ran the code and collected the data again between 17 and 20 February, 2021 because the data file it produced was corrupted, and I could not verify my findings against it. I included

The code then established whether or not ties existed between users on the list, and pairs were created from this list (between each keyword-identified user and each of their friends). An edge list was then generated from the pairs, where each row in the edge list represents a tie between two users. While most users in the original list followed or were followed by at least one other person in the list, a large subset of the sample had no ties to others at all.

Note that the graph produced by this list of nodes and the edges between them contained a list of people and the connections between them within the list. This corpus does not contain the original list of users plus *all* the people they follow (such a method would produce an enormous graph, and would contain a great many nodes with no relation to open science or reform), it only includes the original list and the connections that exist between *them*.

Additionally, attributes for each node were mined from Twitter, including their friends and follower numbers, and their Twitter handles.[8] The code and data generated from it are available at this project's OSF page at `https://osf.io/6yr72/`. The code for this project was written in Python, with much help from Vincent Barbay.

Once I obtained an edge list and node attribute files, I imported the data to the network analysis software Gephi for analysis and visualization. Gephi is open-source software for network analysis and data visualization, first released in 2008 (Jacomy, Venturini, Heymann, & Bastian, 2014). It features a graphical user interface with many options for handling the data itself, network analysis statistics, and the visualization of network graphs. For the research in this chapter, I used the 2017 updated version 9 (0.9.2).

Data Analysis

The full data set contained 2958 nodes. Although most nodes were tied to at least one other node, a large number of isolates – 470 nodes, to be precise – exist in this network (if the reader recalls, isolates are identified by not having any in-going or out-going ties, i.e., a degree of 0). Although some of these isolates are genuine (i.e., they're part of the community based on their bio keywords, they're not connected to anyone else in the user list), most are not (by this I mean that they should not have been part of the list obtained from Twitter, because they are not actually interested in open science reform).

I have randomly checked 20 percent of the 470 with a degree of 0 (i.e., about 96 users), and found that most of them (84 users) should not be included in the data set. I elaborate on the possible reasons for this in the limitations section in the discussion. Based on this finding, I chose to exclude all of them. I ex-

two more keywords in this re-run which increased the sample by approximately 90.

[8] All information for a given user is returned by the Twitter API for methods such as `search_user`, and is saved as a .JSON file. The code for these analyses pulls out from these files information I thought might be relevant, though I do not use all information.

cluded a further 242 nodes, as they had a degree (total followers and friends) of four or less, for the same reason.[9] I checked half of these accounts, and as with the isolates, there was no good reason for them to be considered a part of the open/reform corpus. While this excludes potential community members, it will not exclude many or central ones. The data set with these 712 nodes excluded contained 2246 nodes and 100128 edges.

Using Gephi, I calculated a range of metrics on both the node level and on the network level: Average degree, eigenvector centrality, network diameter and density, and average path length.

Gephi, though it has its strengths, does not produce all network statistics a user might wish to calculate (such as reciprocity and transitivity). I wanted to calculate reciprocity and transitivity for each of the graphs, so I used the Python library NetworkX to calculate these. I also used libraries Seaborn, Numpy and Pandas to produce the plots accompanying each network I describe. All data and accompanying code can be found on this project's OSF page.

SNA Results

Open Science and Reform Community Follow Graph

As I mentioned, the follow graph (which captures keyword-selected accounts and the connections between all of them) generated by the data I imported to Gephi produced a network containing 2958 nodes. I excluded a subset of the nodes which I discussed in **Section 7.5**, leaving a filtered sample of 2246 nodes and 100128 edges. I then used Gephi and NetworkX to produce relevant node- and network-specific metrics which can help give a sense of the network's characteristics, and plots to visualize the data. Table 8.1 lists the calculated network-level metrics, which I discuss in some detail shortly.

Using Table 8.1, one can see that on average, a given user in this corpus has around 45 followers and friends. Once again, this figure refers to how many followers and friends an actor has *in this corpus*, not overall (the average user on Twitter has approximately 453 followers; McCarthy, 2016).[10] The median user follows 709 other users on Twitter, and is followed by 885.

Like most Twitter follow graphs, the in- and out-degree distributions for the network are heavily skewed. Specifically, only 74 accounts, or three percent of the total, *are followed by* more than 10% (i.e., 224 users) of the corpus. Even more extremely, only 33 accounts in the network *follow* more than 10% of the corpus.

[9] Although some might argue that leaving them in there might not have affected my later analyses, I considered there to be enough of them that removing the noise they create might be beneficial, and possibly lead to the calculation of less noisy metrics.

[10] This is not the true average, it is the trimmed average which removes extreme outliers (leaving 99.04% of the sample) like former US president Barak Obama and celebrity Katy Perry who each have over 100 million followers.

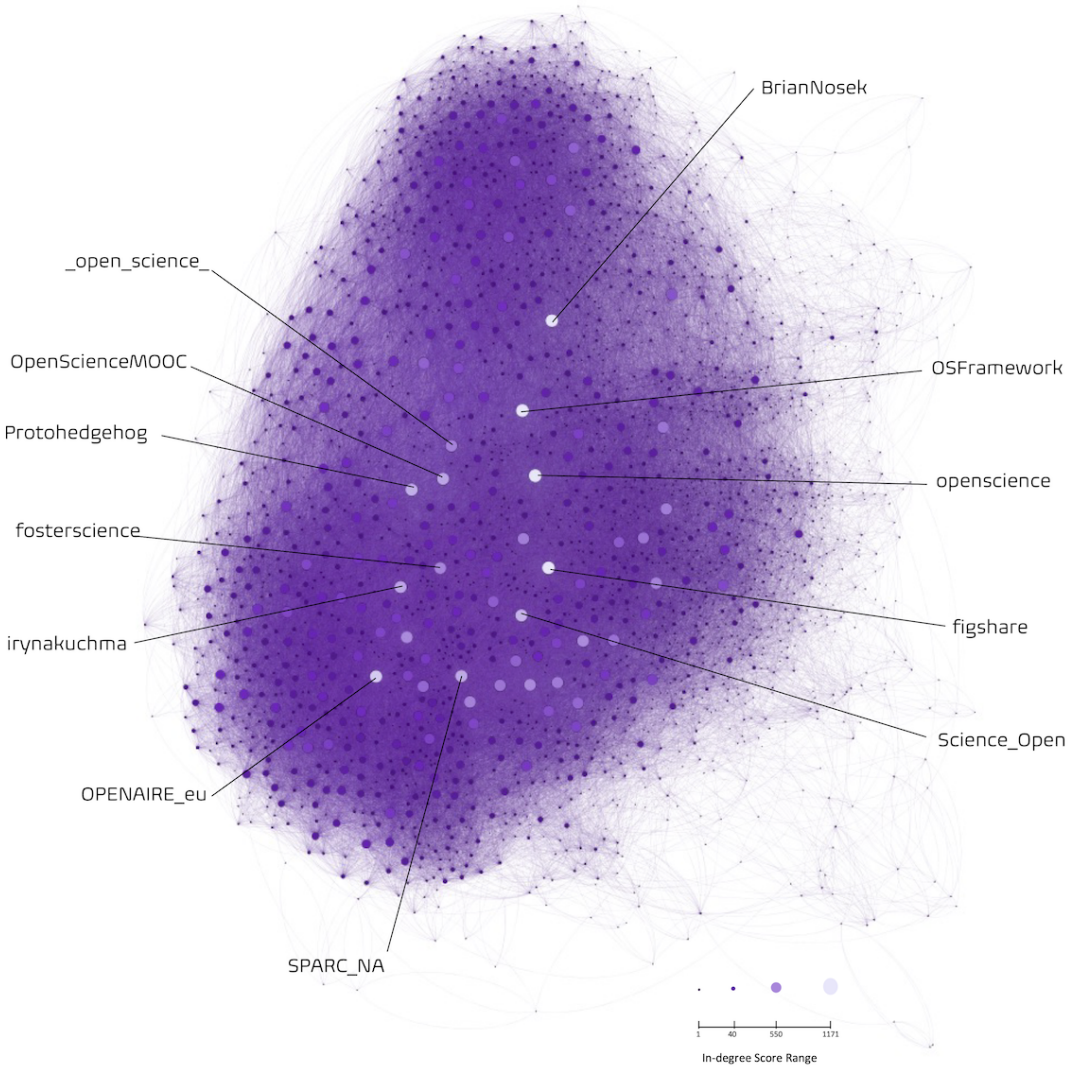


Figure 8.1: Visualization of the complete follow graph ($N = 2246$). Node size and shade denotes in-degree: larger, paler nodes indicate that the node has a higher in-degree score, that is, that they have a relatively high number of in-going connections (they are followed by many in the corpus).

Metric	Value
Nodes	2246
Edges	100128
Ave. Degree	44.581
Diameter	8
Ave. Path	2.661
Density	0.02
Transitivity	0.22
Reciprocity	0.49

Table 8.1: Network-level statistics for the open and reform follow graph.

This shows evidence of a phenomenon typically observed in social and information networks like Twitter: A small number of users, or nodes in the graph, take up most of the 'resources'. That is, most accounts in the network have few followers, and follow few people, relatively speaking, compared with a few accounts which follow many and are themselves followed by many. It is well known that, typically, nodes initiate ties with more nodes than the other way around (i.e., the average Twitter user follows more people than they are followed; see, for instance, Myers, Sharma, Gupta & Lin's exploration of the structure of Twitter follow graphs: 2014). This phenomenon is well-known and heavily discussed in the SNA literature as the power-law. Mislove and colleagues (2007) discuss how in- and out-degree distributions of networks such as Twitter and YouTube are usually heavily positively skewed. The same goes for other metrics which quantify forms of social capital (like popularity and influence, for example).

Naturally, some of the accounts in this corpus exaggerate this effect, as they are the kinds of users who occupy some kind of authority position, or privilege in the community. These kinds of accounts belong to journals and institutions, societies, platforms and well-known actors. For instance, @NatureComms, the account belonging to the journal *Nature Communications*, OSC founder @BrianNosek and platform @OSFramework are two such accounts. A high degree score (in-degree, out-degree or total degree) does not necessarily mean that the account is especially influential in the network; some such accounts are automated by bots which follow a great number of other accounts. Take the example of the account @openscience, which retweets any content containing #openscience – it has 71,500 followers and follows 16,000 other accounts.

To further study the properties of the network, I inspected the network's diameter, which gives us an idea of the shortest distance between the two most distant nodes (which is 8). The average path length (2.661 for this network) describes the average shortest path between two users. In the analysis of in-person social networks, this metric can give us a sense of how efficiently information can be

passed between nodes in the network.

The transitivity index for this network is 0.22, meaning that the proportion of pairs of user accounts that share a mutual friend or follower (in directed networks, the direction of the tie is ignored) is around 1 in 5. Despite this high count, the network is sparse, with a density coefficient of 0.02 (meaning that around 2% of all possible connections between the nodes in this network have been made). A density of 0.17 is a common density threshold among communities of friends and family in physical space (Wellman, 1979). Additionally, this is a relatively large network, and density has been shown to dramatically drop as a function of network size (see, for instance, Figure 7 in Stephens & Poorthuis, which uses a number of data sets to demonstrate that density is almost always below 0.05 once a network contains more than 1500 nodes Stephens & Poorthuis, 2015).

The reciprocity index (which I defined in **Section 7.3**) for this network is 0.49 for this network. It quantifies the tendency of users in a network to reciprocate followers. In the case of this network, around half the ties are reciprocated or 'mutuals'. This index is not high or low in and of itself, necessarily, and it complies with both hierarchical elements and a certain degree of 'horizontality' in the pattern of connections in this network.

The finding of a sparse network, although not surprising given its size (and the fact that it is a directed network),[11] highlights how network visualization can be misleading. As Grandjean (2015) reminds us, networks are (sometimes strongly suggestive) visual representations of data, whose complexity is not limited to a simple graphical rendering.[12] The graph in Figure 8.1 *looks* denser than it is, given the number of *actual* ties compared to the number of *possible* ties. That said, I did expect that a community would be denser than this follow network seems to be. Is there more to the structure of this community of people than this preliminary exploration suggests?

Interim Conclusion

Ultimately, this initial exploration of the data does not tell us very much about the open and reform network that it – to some degree, at least – represents. The 'hairball'[13] visualization and modularity value (which I explore in detail in the next subsection) indicates that there is complexity and richness in these data that

[11]The density of an undirected graph with a given number of nodes will be two times that of the same graph but with directed edges. This is because two edges can exist between every pair of nodes in a directed network – one incoming and one outgoing – compared with an undirected network which only has one kind of edge between nodes (p. 127; Abraham & Hassanien, 2012).

[12]I will discuss the limitations and possibilities of graph visualization in the conclusion.

[13]The hairball effect is a well-known issue in network visualization. I will let Schulz and Herter explain: "... every visualization researcher and practitioner knows the painful experience of a beautifully designed network layout breaking down once the input graph scales up to realistic node and edge counts. The resulting "hairball" suffers from cluttering and overplotting to an extreme that renders it unusable for any practical purposes." (p. 1; 2013)

this bird's-eye view cannot reveal. The retrieved statistics tell us some things: the graph is sparse. With a diameter of 8, and a long average path length, this is not a small-world network, 'where everyone knows everyone'. Usually small-world networks have shorter diameters (around 6, see Kleinberg, 2000) and relatively short average path lengths. The distributions of in versus out degrees is skewed, and on the whole, one in four dyads will share a mutual follower or friend, and about half the network tend to reciprocate a follow.

I now turn to an exploration of the community structure in this network. I hope, in breaking down the network down into smaller components, to see if some more structure and meaning can be extracted from the data.

Modularity Within the Open/Science Reform Follow Graph

In keeping with the previous chapter's data analysis, I considered the network from the perspective of Wenger, treating the greater network as a constellation of many smaller ones. I remind the reader of a passage found on page 126 of Wenger's 1998 book: "Some configurations are too far removed from the scope of engagement of participants, too broad, too diverse, or too diffuse to be usefully treated as single communities of practice." Throughout the following exploration of possible sub-communities, I hope to provide a new angle on my argument in the previous chapter that it is most fruitful to think of this group as a constellation or grouping of sub-communities, rather than a single homogeneous community of practice.

Community Detection

A property common to many networks is community structure, the "division of network nodes into groups within which the network connections are dense, but between which they are sparser" (p. 1 Newman & Girvan, 2004). Although community structure in networks has been studied for some time (since the late 70's, according to Newman and Girvan), methods for doing so up until the early 2000s have yielded mixed results. This is, say Newman and Girvan, largely due to the fact that earlier methods relied on hierarchical clustering techniques. These techniques, which are "aimed at discovering natural divisions of (social) networks into groups, based on various metrics of similarity or strength of connection between vertices" (p. 1, *ibid*), have limitations which result in unreliable network partitioning. Newman and Girvan describe two main shortcomings. First, they regularly fail to find a valid community structure in networks where the community structure is known a priori (and so it is unclear whether they partition a network correctly in other cases where the structure is not known in advance). Second, they tend to focus on the cores of communities, neglecting to assess peripheral nodes (which end up getting placed in the wrong communities, even

in simple networks where their sub-community membership is clear, even under simple visual inspection).

Newman and Girvan's (2004) article presents a method for community detection[14], which forms the basis for 'modern' community 'detection' algorithms. As Newman and Girvan (2004) demonstrate, their method retrieves known community structures with a high rate of reliability. However, it is crucial to have a way to determine, for any method, whether the communities detected are valid or meaningful. Certainly, as Newman and Girvan emphasize, community 'detection' algorithms will always produce partitions, even in randomly generated networks with no meaningful sub-structures. This is where **modularity** comes in. Modularity, also known as the quality function (Traag, Waltman, & Van Eck, 2019), quantifies the *quality* of the community structure which has been detected by an algorithm. More precisely, it "measures the fraction of the edges in the network that connect vertices of the same type (i.e., within-community edges) minus the expected value of the same quantity in a network with the same community divisions but random connections between the vertices." (p. 7; Newman & Girvan, 2004)

Newman and Girvan (2004) state that in cases where the number of intra-community edges is no better than what would be expected in a randomly generated network, modularity (commonly denoted as Q), is equal to 0. Q approaches 1 the stronger the community structure is in a network (i.e., in a network that is comprised of very tightly connected sub-communities with few ties connecting them to one another). According to Newman and Girvan, values of Q tend to fall between 0.3 and 0.7 (higher values than this are rare).

Community Structure in the Open Science/Reform Follow Network The default network structure-finding method in Gephi is the so-called Louvain algorithm (created by Blondel, Guillaume, Lambiotte, & Lefebvre, 2008). The Louvain algorithm's approach to detecting communities, based on modularity (discussed above), involves two key steps which are repeated: the 'detection' of small communities within the larger network by optimizing modularity on all nodes, then each of the small communities is grouped into one node. The idea is to connect communities which, when combined, produce the largest increase in Q.

According to Traag, Waltman and van Eck (2019), the Louvain algorithm is one of the most commonly used algorithms to optimize modularity, as it is, relatively speaking, one of the fastest and best-performing algorithms tested. However, Traag and colleagues have released an improvement on Louvain (which they named Leiden). In their article, they demonstrate two improvements on Louvain: 1) speed and 2) the quality of the connections of communities detected. These

[14]The word detection here implies more precision than may be warranted given the potential shortfalls of these kinds of methods. I use the word as that is the word commonly used in the literature, however it should not be taken to imply that the method necessarily yields very precise partitions.

improvements point to limitations of Louvain, and differences between the performance on both algorithms in terms of quality become apparent in some cases, but remain small (less than .02), and differences in speed mainly become an issue for very large networks (such as networks with 1,000,000 or more nodes; see Traag and colleagues' Figures 5 and 6 for the results; 2019).

Despite this, it was prudent to check the partition results between Louvain and Leiden for the data I describe in this section just in case meaningful differences in the partitioning would indicate problems with the validity of my analyses. I updated Gephi (as a new version, 0.9.3, had just become available early in April, 2022) and installed a plugin created by Traag, Waltman and van Eck (2019) to run the Leiden algorithm[15] and check the results against those produced by Louvain. I re-ran both algorithms (using the modularity quality function) multiple times, in an alternating fashion. The results of these runs demonstrate that Leiden produces markedly similar partitions to Louvain (this project's OSF page (DOI: `10.17605/OSF.IO/SGQRC`) contains the results of this check). Values of Q are roughly the same (they are all between 0.322 and 0.326), virtually the same numbers of communities are detected (usually 6), and the communities detected are of comparable sizes (4 or 5 large ones and one small one). Importantly, mostly the same nodes with highest betweenness are assigned to the same communities. These findings serve to validate the outcomes of the Louvain algorithm with the network I explore in this section, and corroborate the community structure I describe in the following subsections, at least in terms of the most central nodes in each sub-community.

Communities

Description I ran the modularity algorithm for the whole 2246-node network, using the default setting of 1.0 for the γ parameter[16] and not using edge weights (which is checked by default, but irrelevant for these data since they are not weighted). I left the 'randomize' option checked (which refers to how the algorithm chooses which nodes to use in its 'detection' process, which is also checked by default).

The algorithm yielded a solution with five communities within the network,

[15]This plugin had been available on the previous version, 0.9.2, however a bug in Gephi's GUI made it impossible to select 'modularity' as the quality function when choosing options for the algorithm. Only in version 0.9.3 could I actually select all of the desired options, and successfully run Leiden.

[16]The resolution parameter, denoted by γ, adjusts the number of of sub-communities the algorithm produces. Gephi sets this resolution parameter to 1.0 by default. Inputting values higher than this will ask the algorithm to attempt to produce fewer modules (larger ones), and setting the value at less than 1.0 will likely produce more, but smaller modules. Adjusting this parameter can result in badly partitioned networks, with low values of *Q*, if you set the resolution to high, and artificially high levels of *Q* if you set the resolution too low.

with a *Q* value of 0.323, which indicates that the network is somewhat modular.[17] This is not surprising to me, given that my qualitative analysis supports the idea of a constellation of communities, rather than a single, homogeneous one. The sub-communities detected by the algorithm varied in size, but the largest four (390, 605, 578 and 631 nodes) were much larger than the fifth community detected (42 nodes). Table 8.2 provides the network statistics for all the communities derived by the algorithm side by side, which allows us to compare the communities with one another on their different characteristics.

	Community				
Statistic	1	2	3	4	5
Nodes	390	605	578	631	42
Edges	4793	8733	20671	29346	171
Ave. Degree	12.29	14.435	35.763	46.507	4.071
Diameter	9	7	6	5	8
Ave. Path	3.112	2.818	2.337	2.193	2.821
Density	0.032	0.024	0.062	0.074	0.099
Transitivity	0.28	0.23	0.30	0.31	0.46
Reciprocity	0.53	0.42	0.58	0.54	0.77
Q (communities)	0.329 (5)	0.284 (6)	0.21 (5)	0.218 (5)	0.525 (7)

Table 8.2: Table showing network-level statistics for each of the five sub-communities in the follow graph, detected by the algorithm.

Analysis and Comparison Table 8.2 gives a sense of how connected the nodes within each network are. Community 4 is the most densely connected – the average node within it is connected to 46 other nodes, it has the highest density score of all the communities, and the highest transitivity index. It is still relatively sparse, however. The connections within this network show skew – only a few nodes have a large number of connections. Networks of this kind are described as a 'small world' network, by the relevant literature. These networks are "generally sparse: the total number of links... is very small compared to the maximum number of links..." and are characterized by short average path length (p. 1; Takes & Kosters, 2011), which for this community is 2.193. Small world networks also tend to be skewed, with only a "few nodes with a very large amount

[17]It is common to talk about 'detection' of a community structure, but it is important to note that the result of this process simply comes down to algorithm yielding a partitioning of nodes into 5 clusters (likely) yielding (about) the highest modularity of all possible partitionings into five clusters. Of course, it is not possible to assess whether these clusters serve as real sub-communities, clearly distinct from each other. I argue, however, that they may serve as approximations to partly independent sub-communities, and therefore consider them as a description of a structure that possibly quite well resembles an actual constellation of CoPs.

of connections, the so-called hubs, and there are many nodes with relatively few connections." (p. 2; *ibid*)

Community 2, based on the statistics given in Table 8.2, is the least cohesive of the five. It has a markedly low average degree score. Average degree refers to how many edges exist in a network in relation to how many nodes there are. This, in combination with the lowest density and transitivity index, the highest number of sub-communities, and the second-highest node count indicates that Community 2 is even more loosely connected and sparse than Community 4, and is unlikely to be a community where people closely relate to one another. They might be quite a fragmented, and possible diverse group, with small sub-groups within it that do not have much in common with the others. Alternatively, they might have more or less random connections to one another, but relatively few of them.

Comparing the communities on overall eigenvector centrality (i.e., seeing which of the most influential actors within the whole community has been assigned to which sub-community) is interesting, because it can give us a sense of which community might be most influential in the constellation, based on the average actor, or node within each sub-network. Table 8.2 shows five histograms capturing the distributions of the eigenvector centrality variable for each of the five communities.[18] The colours correspond to each of the communities – Community 1 is indicated by blue, Community 2 by green, 3 by yellow, 4 by magenta, and 5 by dark purple. We can see that Communities 3 and 4 contain a comparably large number of highly influential nodes, while the range of the eigenvector centrality scores in Community 1 does not even go above 0.5.

Two questions that naturally arise at this point are: is the partitioning of the greater network reasonable, and, if so, who are the people in these communities? It is difficult to quantify the former, beyond noting that the quality function Q of 0.323 is reasonable (Newman & Girvan, 2004), but exploring each of the five communities might provide a qualitative look at whether there is a natural partition in the network that the modularity algorithm has uncovered.

To that end, I separated the greater network out into each sub-community to generate graphical representations of the node clusters themselves, and to see what words people used in their bios to describe themselves. I copied each of the users' bios for each of the sub-communities separately into an online 'word cloud' generating service (`https://www.wordclouds.com`), to visualize the relative frequencies of words in the bios. In each word cloud, the size of the words is proportional to the frequency of the words in the list. Larger words are relatively more frequent than smaller words. The word clouds I generated are an interesting visual exploration of the words in the users' biographies, which complement

[18] It is crucial to note that this plot shows us the distributions of eigenvector centrality scores calculated for the whole 2246-node community split up by community, and *not* the eigenvector centrality scores recalculated for each sub-community (which is what Table 8.2 reports).

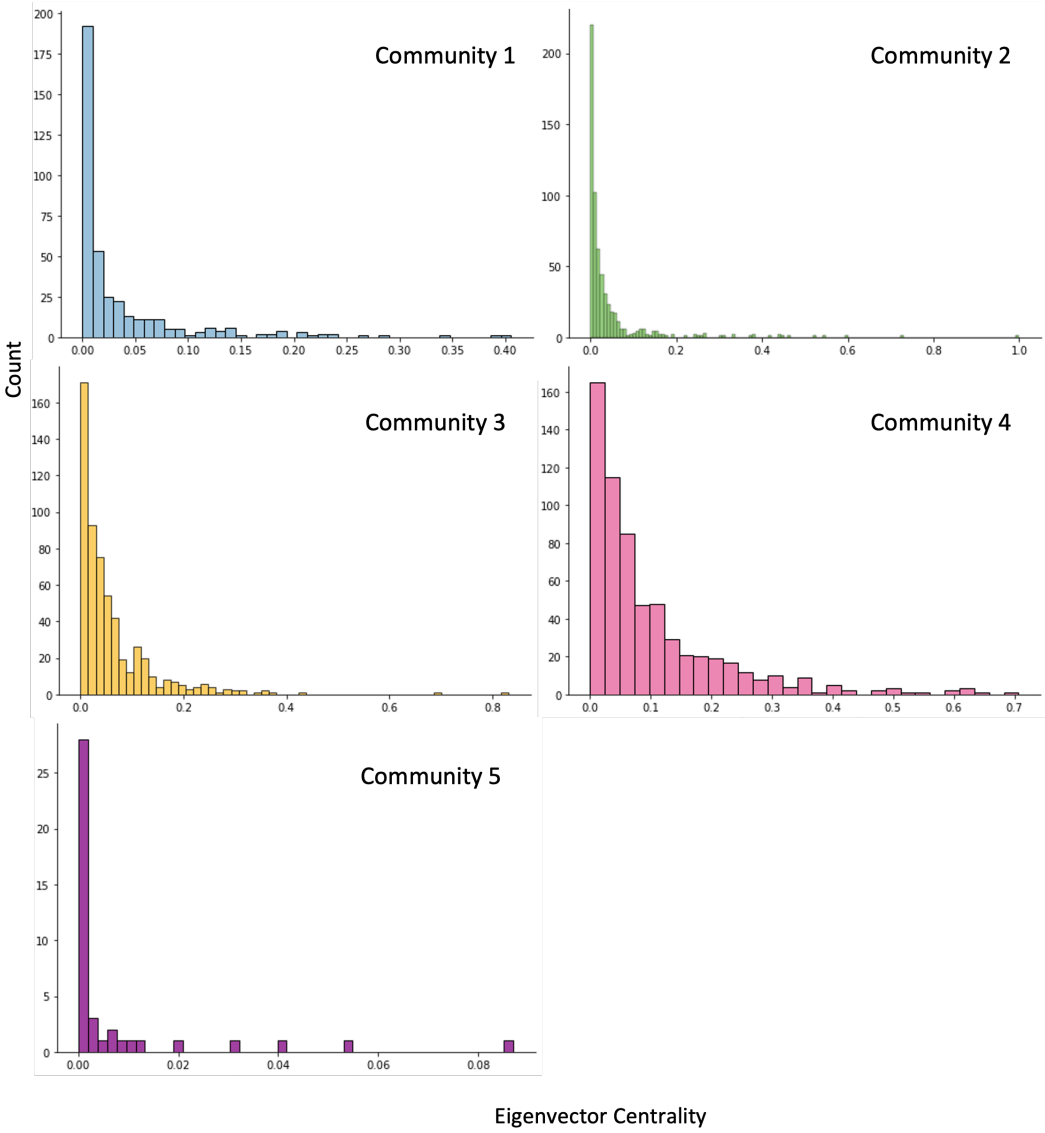


Figure 8.2: Histograms showing the skewed distributions of eigenvector centrality scores, broken down and color-coded for each of the five communities.

the qualitative descriptions I provide for each community.

The word-cloud for each of the communities captures the interests, skills and characteristics of this group as they self-describe in their bios. In exploring the word-clouds, I highlight a few of the most salient (relatively more frequent words) and discuss them to highlight certain points in the exploration that I thought were interesting, and I point to words that are present in some sub-communities word clouds which are absent in others. I also calculated statistics like eigenvector centrality, degree and betweenness centrality for each of the networks separately, to highlight potentially interesting actors within each sub-community. I will explore each one in turn now.

Community 1 The first corpus detected by the algorithm may be the most diverse and heterogeneous of the five, but it is still interesting to see if qualitatively assessing the characteristics of the member nodes validates the grouping of these nodes together by the algorithm. The word-cloud for this community, as you can see, highlights words like *software, open-source, reproducible, rstats, python and computational*, as well as *engineer, phd, professor, postdoc, scientist* and *tech*. The biographies of the most influential figures[19] in this network lend a face to these self-assigned descriptors. For instance, the account with the highest eigenvector centrality score, @rOpenSci has the bio: *"rOpenSci develops #RStats based tools to facilitate open science and access to open data."*. User @juliesquid, ranked 2nd, writes of herself in her bio: *"Championing open data science for kinder, better science for future us.*... Users @cboettig, @choldgraf and @noamross (ranked 3rd, 5th and 6th) have similar bios, and share that they are interested in data science, R, Python and computational research. @carlystrasser, formerly in academia, is an open science program manager at the Chan Zuckerberg Initiative.[20]

Some themes are clear, for this community. Though they seem to be the least cohesive group of the sub-communities, many in the group seem to be interested in the more technical side of the open and reform science enterprise, focusing on practices involving statistics, other software and platforms, as well as general research infrastructures. They represent the academic sciences (indicated by high frequencies of words relating to academia like *professor and PhD*), as well as industry research (*co-founder, manager, tech, engineer*), and appear to share an interest in reproducibility and openness.

[19] Some readers might wonder why eigenvector centrality scores are more interesting to use here than, for instance, in-degree. However, the use of in-degree for a measure of influence in social media networks is confounded by the fact that following behavior is highly subject to being gamed by some accounts (like bots). Eigenvector centrality on the other hand is more robust to issues like this, and is commonly used to determine a node's authority or influence (see e.g., Parand, Rahimi, & Gorzin, 2016) because of the information its calculation takes into account (described earlier in this chapter), which is why I use it here.

[20] The CZ Initiative is a philanthropic organization with a focus on science and education, among other things, co-founded and run by the founder and owner of Facebook, Mark Zuckerberg, and his wife.

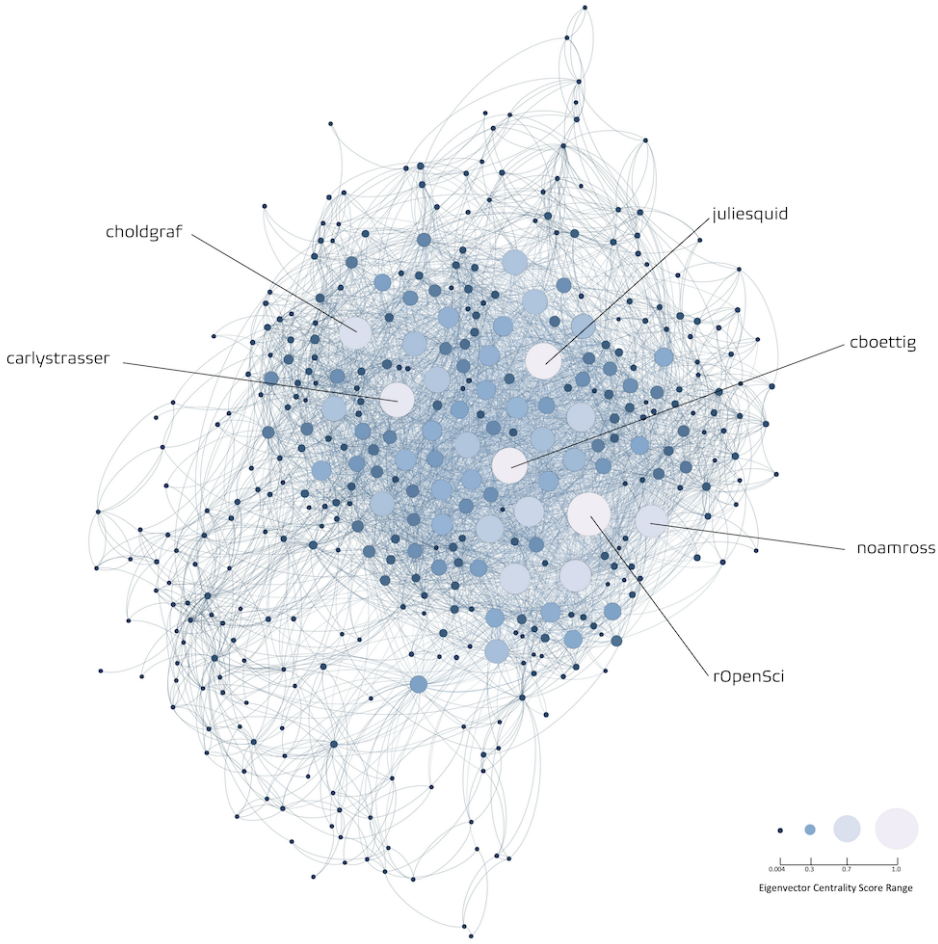


Figure 8.3: Network visualization of Community 1 detected by the algorithm. Node size and colour indicate most influential nodes, where paler, larger nodes are highly influential, and smaller, darker ones are less influential. the most influential nodes are textually labeled.

Figure 8.4: Word cloud generated using adjectives listed in Community 1's members' biographies. Larger words indicate higher frequency in the word list.

Community 2 The second sub-community the algorithm detected is similarly sparse and unconnected, but much larger in terms of node count than the first. This network has the lowest transitivity index of all of them, and is the least dense. The word-cloud generated for the bios of this group show some strong themes, however. This group, it would seem, contains many people interested in the natural and life sciences, given the relative presence of words like *biology, genome, bioinformatics, molecular* and *health* (which are not visible in the other word clouds, meaning that they are either not present or very infrequent). Interestingly, the words *journal, access, publishing* and *openaccess* are also prominent. A look at the most influential nodes' bios aligns well with the relatively high frequencies of such words. The account @figshare is an account representing an online, open access platform, which gives researchers a place to store and share

their research output. The @F1000Research platform provides a similar research infrastructure in the community.

User @eLife is an account for a journal: *"The funder-researcher collaboration and open-access journal for promising research in the life and biomedical sciences. . . "*. The bio of the account for journal PLOSONE is not dissimilar: "*PLOS ONE is an international, peer-reviewed, open-access, online science publication.*" Actors @mbeisen and @phylogenomics are both academics in the life sciences, who study genomics and other microbiology topics. Other notable accounts are for journals like the BMC series: "*A group of open access, peer reviewed journals spanning biological, physical, engineering and medical research. BMC series – Part of @BioMedCentral", Nature Communications, BioMed Central and Open Pharma*.

The thus identified themes for this group may explain why it was indicated as a sub-community. Open-access scientific journals with a focus on life and natural sciences can be expected to naturally be grouped together with research scientists with similar research tendencies. Again, the overarching connection among them is openness, particularly as it relates to publishing platforms – open access, platforms and services to facilitate openness and sharing. Interestingly, reproducibility, which is an important element of the joint enterprise, is a relatively less frequent word within this community compared with Community 1's biographies.

Community 3 The third community boasts an average degree that is quite a bit higher than the previous two communities I explored, and has a small diameter and short average path length, especially given its large size. Pairing these characteristics with a moderate level of transitivity and density suggests that this community is relatively well-connected and cohesive.

The histograms of eigenvector centrality distributions for each of the communities in (Figure 8.2) indicates that some of the highest eigenvector centrality scores are from members of this community. The table shows that users @OSFramework and @BrianNosek are ranked 2nd and 5th in the list for this metric.

The word cloud I generated for the bios of this community shows that highest relative frequencies are for words like *psychology, psychologist, social, cognitive, and development* dominate the visualization alongside *university, PhD, postdoc, professor, academic, fellow* and *reproducibility, meta-research, transparency and integrity*. Clearly, in this group words associated with social sciences and psychology are relatively frequent, which seems to indicate that relatively many of its members work in those areas. Unsurprisingly (as psychology was arguably the first to raise concerns about the crisis of confidence and replication failures back in 2011), this combines with a strong emphasis on reproducibility and metascience. I would consider this group to contain many 'original' open science community members – people who might have attended the first SIPS meetings, and might

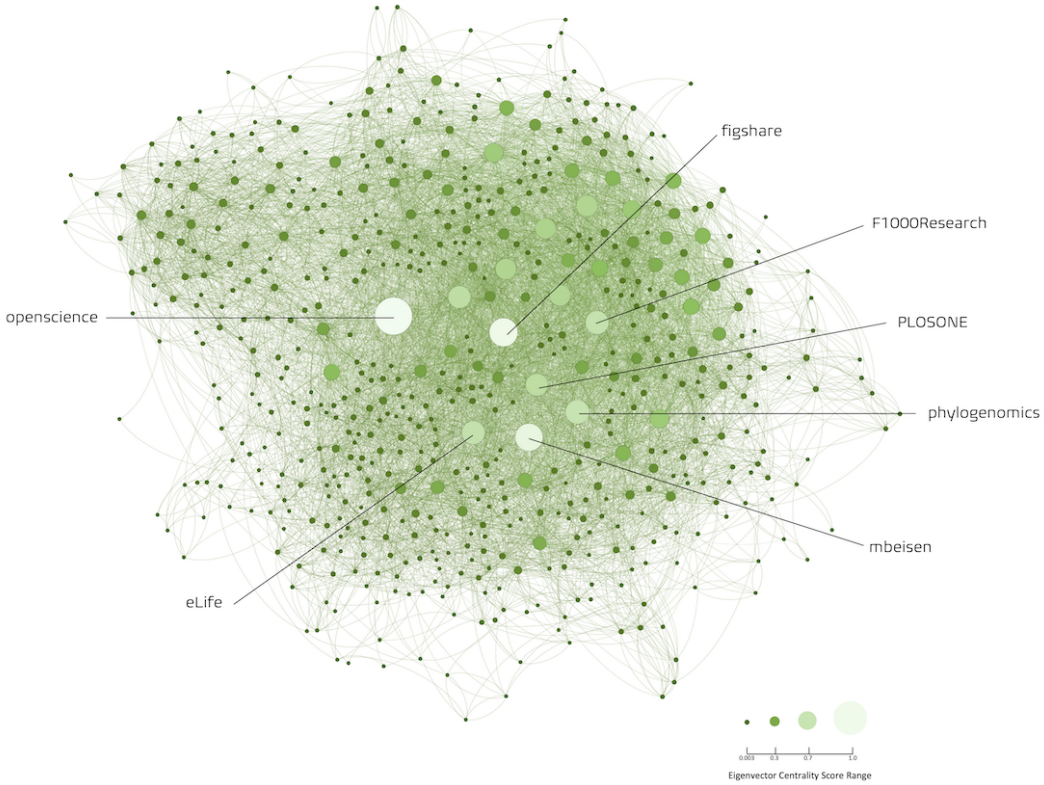


Figure 8.5: Network visualization of Community 2 detected by the algorithm. Node size and colour indicate most influential nodes, where paler, larger nodes are highly influential, and smaller, darker ones are less influential. the most influential nodes are textually labeled.

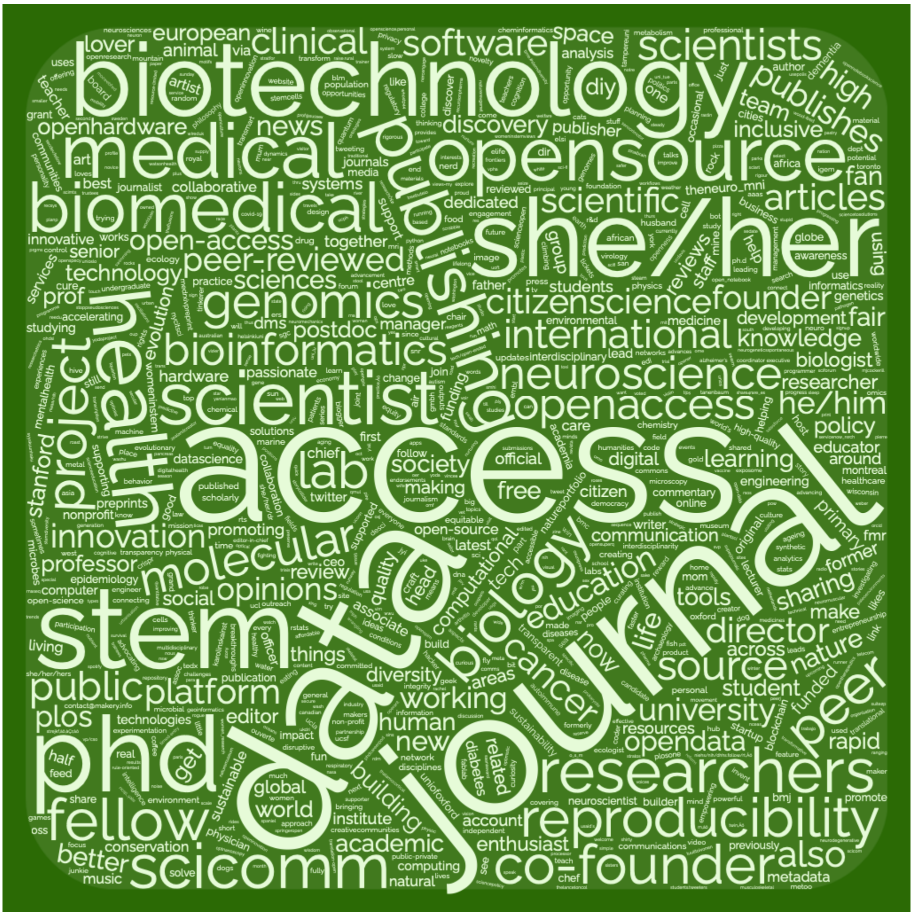


Figure 8.6: Word cloud generated using adjectives listed in Community 2's members' biographies. Larger words indicate higher frequency in the word list.

have been among the first to study metascientific topics alongside their psychology content areas.

User @BrianNosek – founder of the Center for Open Science (COS) and first author of the now-iconic 2015 open science collaboration article which first profiled problems with reproducibility in psychology – is the most influential in this network. Nosek was also one of two guest editors of a highly impactful and controversial special issue: *Replications of Important Results in Social Psychology*, which contained only replication studies. The account @OSFramework, for the Open Science Framework (OSF), is the second most influential node. The OSF, run by the COS, is possibly the most popular platform for sharing study data, materials and preprints (through the PsyArXiv repository, hosted by the COS) in psychology. The Psychological Science Accelerator is a global network of psy-

chological science research groups (represented by the account @PsySciAcc) that focus on large-scale replication studies, and is also an influential node in this network.

Actors @hardsci, @MicheleNuijten, @EikoFried and @fidlerfm (Sanjay Srivastava, Michele Nuijten, Eiko Fried and Fiona Fidler), who are also in the top ten most influential nodes in this network are individual researchers who have each been on the 'open science' scene since very early on, and are well-known actors among the SIPS community. They study topics like psychology and research methods from academic faculty positions. Others like early-career researchers @cruwelli, @DenOlmo, @chartgerink and @peder_isager have held similarly central positions in the open science/SIPS community during the time that I have conducted my ethnography. Europe is strongly represented among this group of actors, with six of the eight actors I mentioned in this paragraph being from either Germany or the Netherlands (Srivastava and Fidler are based in the U.S.A. and Australia, respectively).

Reproducibility, *metascience* and *academic psychology* are among the more popular words within this community. I would argue that this sub-community is at least partly comprised of what S2 and S3 referred to as being the initial or 'early' adopters of open science in psychology.

Community 4 As I discussed earlier in this section, the fourth community identified by the algorithm seems to be the most internally well-connected and clustered. The word-cloud associated with it would suggest that this is a community comprised of researchers and scientists in the digital humanities and science and technology studies, and other people and groups interested in scientific policy, open knowledge and responsible research and innovation. Some salient words listed in these nodes' bios include *openaccess, library, digital, services* and *support*. Words like *european, innovation, rri, eosc, policy, citizenscience, humanities* and *h2020* are also present, while they are absent in the other communities' word clouds.

When looking at the most influential nodes within this community, one could conclude that these mainly represent Europe and are related by their interst in Open access @OpenAIRE_eu, @SPARC_NA, @fosterscience, @resdatall, @OKFN, @OPERASEU and @GOFAIRofficial are all accounts which represent platforms, coalitions and projects that support and foster open knowledge, within academia as well as for the public. Among these groups is an emphasis on innovation, frictionless data and reforming research infrastructure to enable open access for all possible research consumers (not just those in academia with access to well-resourced university libraries).

The most influential actor accounts are @irynakuchma, @jeroenbosman, @RickyPo, @Protohedgehog and @McDawg. According to their bios, these users have the non-faculty research-related roles of: open access program manager,

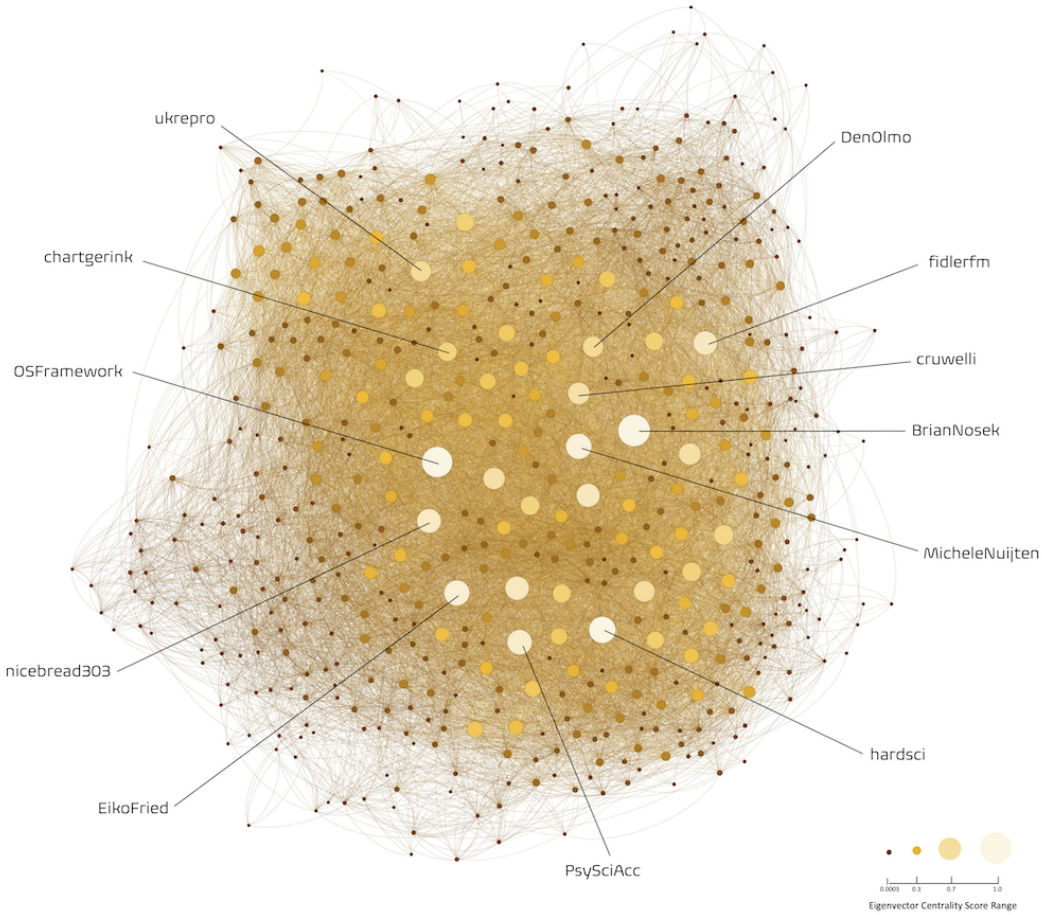


Figure 8.7: Network visualization of Community 3 detected by the algorithm. Node size and colour indicate most influential nodes, where paler, larger nodes are highly influential, and smaller, darker ones are less influential. the most influential nodes are textually labeled.

Figure 8.8: Word cloud generated using adjectives listed in Community 3's members' biographies. Larger words indicate higher frequency in the word list.

university librarian, scientific journalist, independent scientist affiliated with the Institute for Globally Distributed Open Research and Education (IGDORE)[21], and publishing consultant focusing on open access.

An interest in opening up knowledge, for academia and for the public, is what apparently connects most of this corpus. Many of the influential accounts in the group are public-facing in their enterprise, and the top actors are in non-academic/faculty jobs. This is reflected in the relatively lower frequency of some academia-related words (such as *professor, postdoc* and *academic*, for instance), compared to the previous community discussed. Again, as with Community 2, note the relatively low frequency of words like *reproducibility* or *reproducible*.

[21]The correct tense to be used for this individual is past, as he passed away in 2019.

Those words, in fact, are not even visible on this word cloud and demonstrates that their priorities (as expressed in their bios) are focused on openness rather than reproducibility.

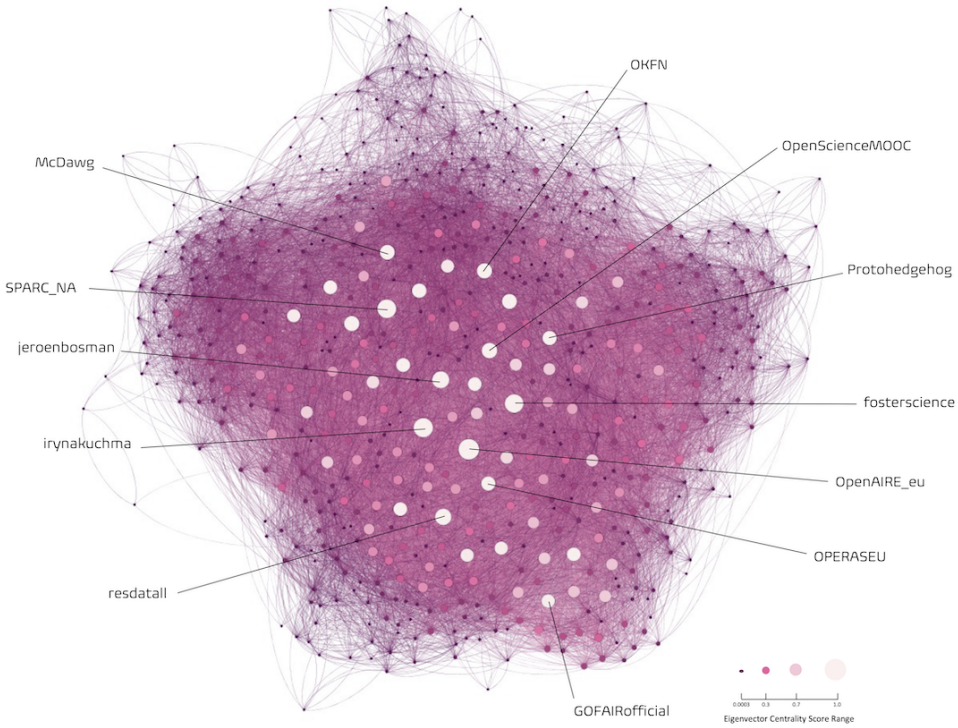


Figure 8.9: Network visualization of Community 4 detected by the algorithm. Node size and colour indicate most influential nodes, where paler, larger nodes are highly influential, and smaller, darker ones are less influential. the most influential nodes are textually labeled.

Community 5 As I mentioned earlier, this sub-community is the smallest out of all the networks found in the modularity solution, with 42 nodes and 171 edges. It has the highest density, transitivity, and reciprocity of all the sub-communities (0.099, 0.46 and 0.77, respectively). Figures 8.11 and 8.12 provide the related visualizations for this community.

This community consists largely of accounts which are related to The Open University (OU), a British public research-focused university, and the largest in the United Kingdom based on number of enrolments. The top three influential ac-

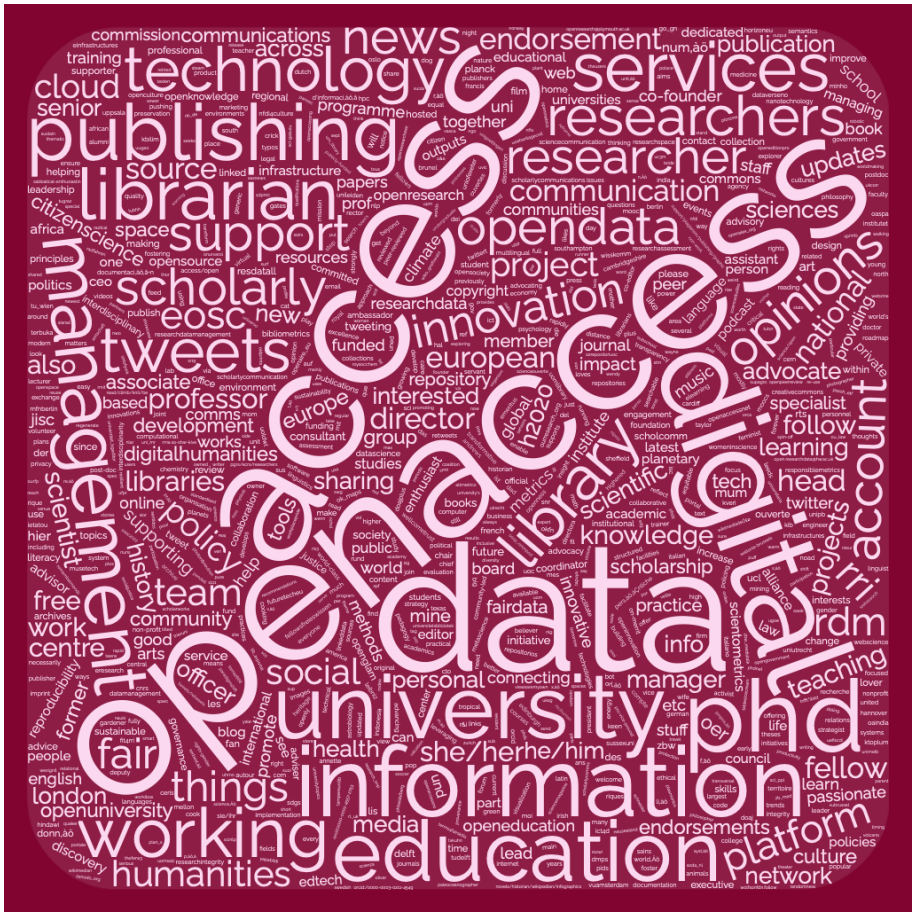


Figure 8.10: Word cloud generated using adjectives listed in Community 4's members' biographies. Larger words indicate higher frequency in the word list.

counts (i.e., those with highest eigenvector centrality ratings) in this network belong to different faculties at the institution, including STEM (Science, Technology, Engineering and Mathematics; @OU_STEM), astrobiology (@Astrobiology_OU) and physics (@OU_SPS). Others belong to researchers who are employed by The Open University.

Although the university's name has the word 'open' in it, this seems not to relate to open science in the sense that I mean it for the context of this study. Instead, it relates to the fact that the university is open in how it approaches university-level study. Its goal is to give "anyone, anywhere the power to learn" (according to its website: `https://www.open.ac.uk/about/main/`). Practically, this means that most of its courses are offered online or in a hybrid capacity, and that it allows flexibility in how students can approach studies. The words in the

word cloud for this network reflect the interests of this small group of accounts – words like *university, learning, teaching* and *openuniversity* indicate this.

Unfortunately, this sub-community seems to be relatively unrelated to the constellation even though it shares edges with several nodes in other sub-communities (which you can see in Figure 8.13). As it is a sub-community relatively unrelated to the reform and open science constellation, I will not attempt to explore it further.

Betweenness Centrality

Brokers – actors that bridge different communities of practice in a constellation – are important for these constellations. As I discussed in the previous two chapters, in Wenger's theory brokers transfer elements across boundaries between practices (in both directions, ideally). Brokers have an active role in this information transferral process, and should be compared with actors who simply hold multimembership and do not necessarily do any active work to bridge community boundaries.

I wanted to explore the possibility of whether brokers could be reasonably identified in the follow network. In the SNA literature, betweenness centrality is used to find nodes which may play the role of broker in a network (Burt, 2002). High betweenness centrality scores denote broker nodes in a graph, as they connect frequently to other nodes and sub-groups in the network (Abbasi, Hossain, & Leydesdorff, 2012). When I first got the calculated betweenness centrality scores for each node in the large network, I saw that nearly all of the major ones were bots and other organizational accounts. As I discussed earlier, such accounts follow a disproportionate number of accounts, and are often followed by large numbers of others themselves. This means that, naturally, their betweenness centrality scores would be high. Although some organizations and groups play a clear brokering role in the constellation, bridging different communities of practice in their activity, many accounts run by bots, or belonging to, for instance, journals, appeared to create noise and obscure meaningful bridge nodes more than anything.

In the interest of revealing which Twitter accounts may play a genuine and clear brokering role in the constellation (on the basis of the quantitative measure of betweenness centrality), I removed a number of accounts from the data set, leaving only human accounts, and accounts that belong to groups, teams or communities in the constellation.[22] I filtered the data set manually, relying on a

[22]Some readers might wonder why I did not just filter the main data set to remove these accounts from the beginning. I considered this, and decided that both kinds of accounts, that is, bots/organizational accounts and human-run accounts, play a role in the reform and constellation. They all have a part in contributing to the discussion of the joint enterprise on Twitter.

Bots can have valuable community-maintenance roles, for instance, because they are efficient in amplifying information and are not selective when they do it (i.e., if they are programmed to

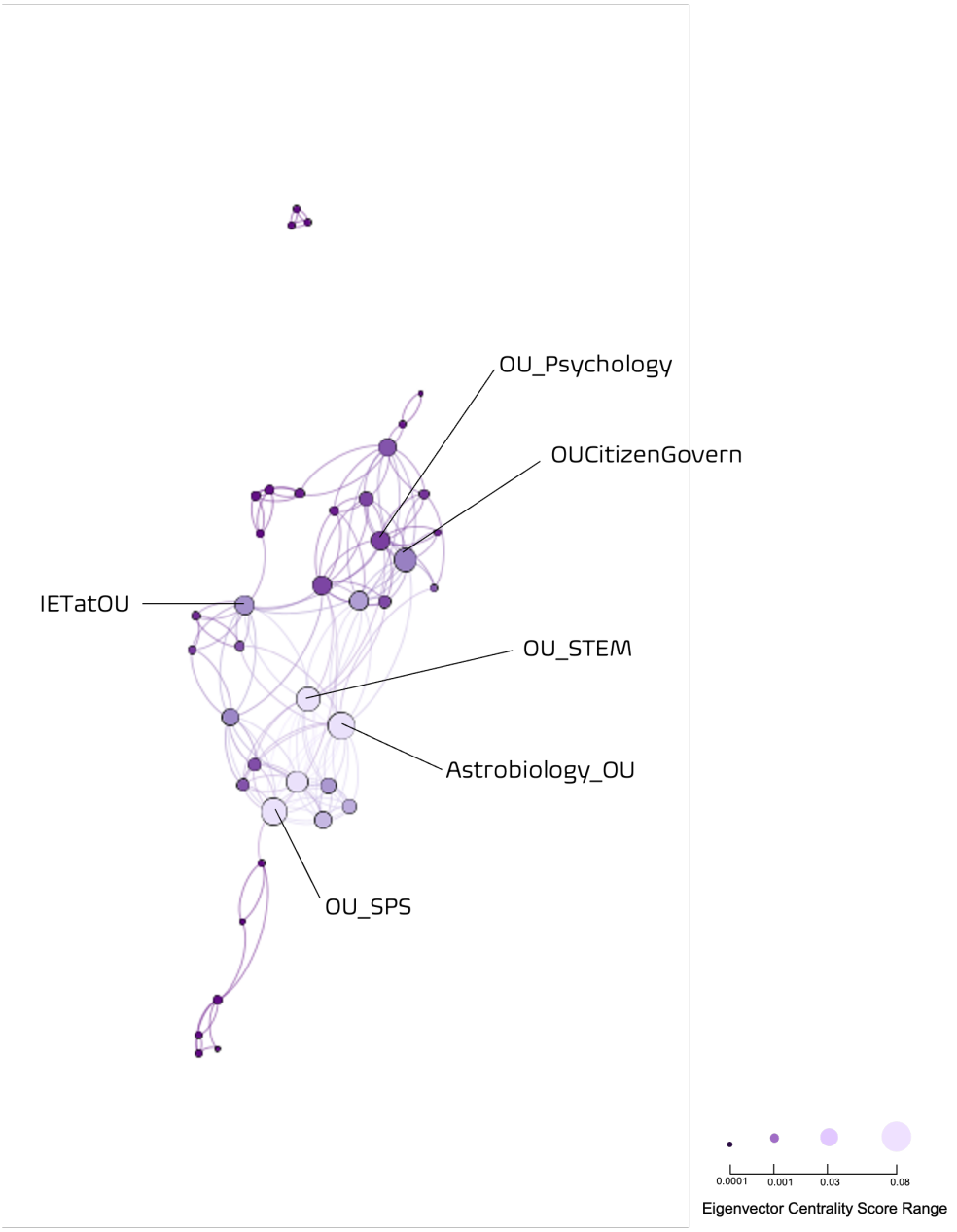


Figure 8.11: Network visualization of Community 5 detected by the algorithm. Node size and colour indicate most influential nodes, where paler, larger nodes are highly influential, and smaller, darker ones are less influential. The most influential nodes are labeled with their Twitter handle.

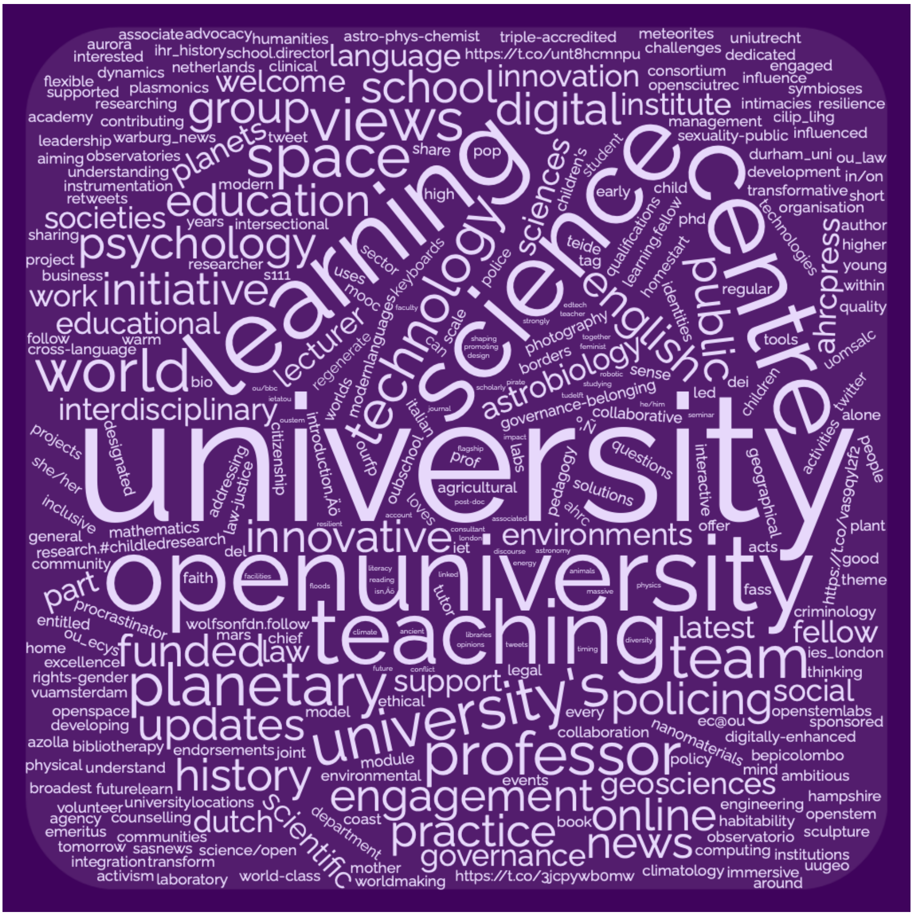


Figure 8.12: Word cloud generated using adjectives listed in Community 5's members' biographies. Larger words indicate higher frequency in the word list.

combination of the text in the users' bios, and the knowledge of the space I have gained over the course of my ethnographic activity. I filtered user accounts out of the data set using the following criteria:

- Automated, or bot accounts
- Accounts belonging to journals and magazines
- Accounts belonging to universities, university departments, or university libraries

retweet anything using #openscience, they will do so indiscriminately). As such, they should be considered an integral part of the main community graph, in my opinion.

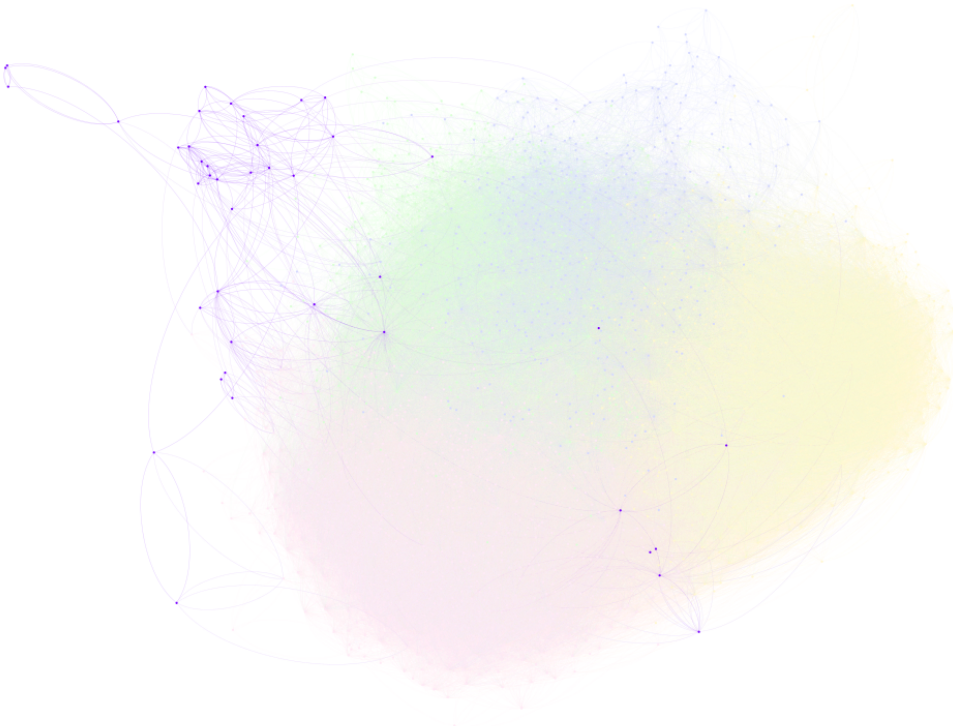


Figure 8.13: Network visualization of the full community, with dark purple nodes and edges indicating Community 5. Node size is consistent, and colour indicates the five communities. The intensity of the node and edge colour is altered to highlight Community 5 and show where its nodes and edges fit within the broader network. To de-emphasize the other four communities, paler colours have been used (they are still consistent with previously used colours: Community 1 blue, 2 green, 3 yellow and 4 magenta).

- Accounts set up for promotion of software, software packages or related tech products
- Accounts associated with repositories, and other products for use in research methods (like Prolific, for instance, which is like Mechanical Turk, except the participants are heavily and carefully curated and screened)
- Accounts for organizations that are only tangentially related to the constellation (like CERN, for instance)

I explicitly kept some organizations and groups in the sample, because they named themselves explicitly as 'open science communities' (or similar things), or because I knew them to be explicitly related to community groups focused on open science and reform. For instance, any of the regional OSCs, the OSF, open- and metascience-focused lab groups and communities that are less clearly labeled by name (but nevertheless are easy to detect based on their Twitter bios) were retained in the sample. SIPS and the UK Reproducibility Network are other examples. If I was unsure of whether or not I should filter a node out, I opted to leave it in.

Data

Description The resulting network contained 1392 nodes, and 53079 edges. The graph and metrics are associated with it are in Figure 8.14 and Table 8.3.

	Network	
Statistic	Follow	Filtered Follow
Nodes	2246	1392
Edges	100128	53079
Ave. Degree	44	38
Diameter	8	8
Ave. Path	2.661	2.57
Density	0.02	0.027
Transitivity	0.22	0.23
Reciprocity	0.49	0.56
Q (communities)	0.323 (5)	0.364 (5)

Table 8.3: Network statistics comparing the original 2246-node follow graph and the filtered version.

The average node had approximately 38 connections. The graph's density was calculated as 0.027, with a transitivity index of 0.23. The network's diameter is 8, and the average path length 2.57. I also recalculated the modularity of the network, which yielded a modularity value of 0.364 with five sub-communities.

Reciprocity for this network is 0.56. The modularity algorithm yielded five sum-communities, and I colour-coded the network graphic to differentiate the sub-communities from one another. Interestingly, the communities the algorithm 'detected' seem consistent in their content as those that were 'detected' in the larger sample which I explored in the previous section, as I will discuss below. This is a kind of validation of the original partitioning the software performed. This network, just as the others, show skewed distributions of the data indicating social power, popularity and influence: Only 17 (1.2%) of the 1392 nodes have an eigenvector centrality score above 0.5, and only four accounts have a score above 0.7.

Comparison with Parent Network This human-only network has marginalized Community 2, identified in the large network as green nodes, as the visualization in Figure 8.14 shows. This is unsurprising, as in the parent network it contained many bots, organization and journal accounts which the filter excluded. Now, this sub-community is influenced by accounts associated with the OSC networks in Europe (meaning that these accounts hold the highest eigenvector centrality scores for this sub-community in the filtered network). The OSC accounts of Utrecht, Leiden, Amsterdam, Rotterdam, Eindhoven, Nijmegen and Groningen all appear in the top 10 accounts ranked by eigenvector centrality. Community 5 has also been marginalized, again, largely because it contained many organization nodes in the parent network. In this refined graph, there are only six deep purple nodes remaining – these are located at the top right of the network, at the periphery. These accounts belong to human users who are working at or students attending the OU.

Communities 1, 3 and 4 (once again, denoted by blue, yellow and magenta respectively) are large in the filtered network, based on node count, which is clear in the network visualization. The top influential nodes in this community reflect the top nodes in the original sample, though now @BrianNosek and @OSFramework occupy the top two positions. Community 4 still dominates in terms of how many of their members occupy influence spots in this network. While Community 3 is represented in the top 20 by seven members, more than half of the list is dominated by Community 4, again (12 nodes).

Although a similar general structure was retained after the network was filtered, leaving mostly single actor-run user accounts, the filtering has not greatly changed important network statistics. Table 8.3 shows these values compared side by side. The filtered follow graph (Figure 8.14) is marginally denser, slightly more locally clustered, and shows a slightly higher tendency for its members to reciprocate follow behavior. It is comparably modular.

Surprisingly, at least to me, filtering out almost 1,000 nodes (some of the most embedded nodes among them), has barely changed the main properties of the network. Intuitively, I would have thought that reciprocity and transitivity

would have increased in the filtered sample, as I would have expected that the filtering would have changed a mixed information and social media network to a mostly social network.

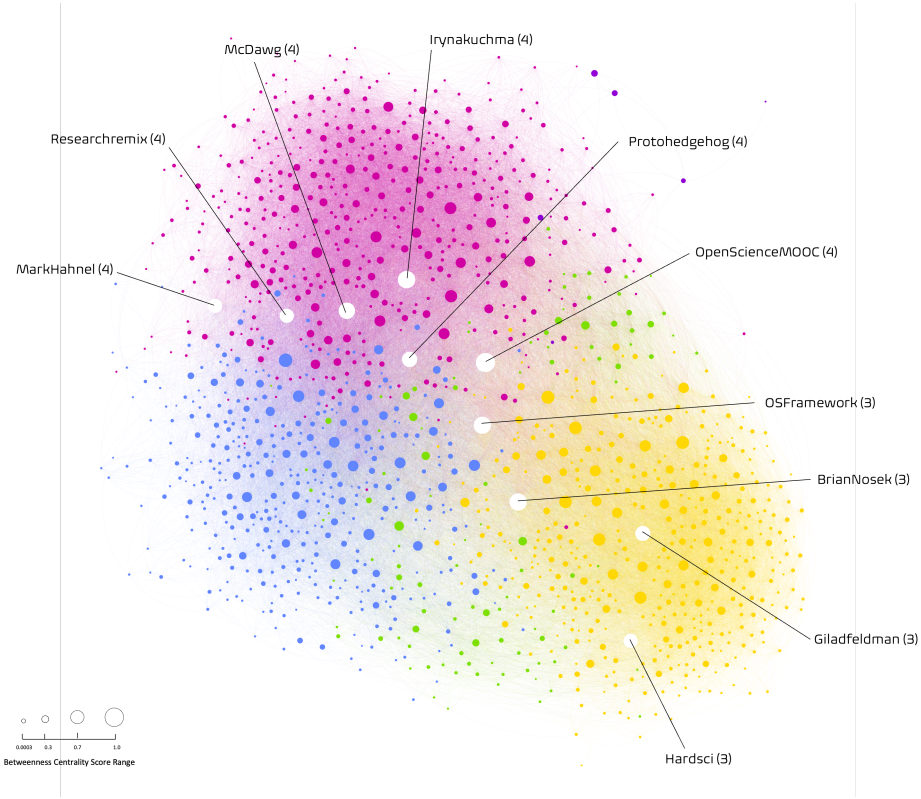


Figure 8.14: Network visualization for the community with bots and organizational accounts removed. Top 10 betweenness centrality-scoring nodes are coloured white for emphasis. Their community is denoted alongside their handle in parentheses. Node and edge color for the rest of the network denote which community they were categorized into (consistent with the previous visualizations of the sub-communities, Community 1 is shown in blue, Community 2 in green, Community 3 in yellow, Community 4 in magenta, and Community 5 in dark purple). Node size denotes betweenness centrality score.

Potential Brokers While it is not possible to identify brokers in a *qualitative* sense using these data, I can attempt to identify people who hold a brokerage position in the structure. Such people, who sit between paths between other nodes, are listed in Table 8.4. They are ordered by their betweenness centrality score. Nosek and the OSF possibly hold brokerage positions in this network, based

on their high betweenness centrality scores. This is not surprising, as the OSF is a very central organization and offers a platform for many different 'kinds' of open scientists and reformers to use.

Rank	Handle	Betweenness	Eigenvector	Community
1	OpenScienceMOOC	172624.41	0.70	4
2	irynakuchma	85496.48	0.60	4
3	BrianNosek	82260.13	0.94	3
4	OSFramework	79439.65	1	3
5	McDawg	65893.04	0.53	4
6	Protohedgehog	51953.79	0.71	4
7	giladfeldman	48704.08	0.41	3
8	MarkHahnel	40443.36	0.51	4
9	researchremix	39879.87	0.59	4
10	hardsci	39794.49	0.51	3

Table 8.4: Top-ten accounts for betweenness centrality in the refined (bot-free) network, ranked by betweenness centrality score, along with their eigenvector centrality scores and the community they were assigned to by the algorithm.

This is consistent with my observations and other qualitative findings. Nosek has a central role in the open and reform constellation, and is active in engaging with many different groups on Twitter on the topic of open science and reform. Here, we can see that it is hard to match the idea of a structural broker with the kind of broker that Wenger describes. I have observed Nosek engaging with many sub-communities in the network, which is reflected in his high ranking in the table, but I am not sure that he performs brokering in terms of bringing practices from one community to another (and back again), for instance. His attempts at engaging with some reform sub-communities are not always answered with interest or receptivity, which would make it difficult for a brokering role to be carried out even if one such as Nosek were inclined to play such a role.

Although this part of the network analysis did not yield what I expected exactly, it is interesting to see evidence of a mismatch between SNA's concept of betweenness centrality in terms of how it defines nodes which play the role of structural broker, and Wenger's concept which defines brokers as people who actively transfer information between different domains and CoPs.

8.6 Streaming Twitter Data

I decided to exploit Twitter's API as much as possible in service of arriving at a descriptive and faithful snapshot of the open science and reform community, from different angles. Searching biographies for certain keywords and selecting actors

in that way, as I did for the network in **Section 7.5** is one method, and streaming using keywords is another one. I use the latter in this section.

During streaming, in which a live connection with the Twitter API is maintained, one can collect data in real time, that is, as they are produced by Twitter users. One can capture tweet and retweet content, and who engages with whom, for instance. Given the nature of this research, I chose to focus on the latter data, and used different keywords (in this case hashtags) to select actors.

#openscience is a commonly-used hashtag in tweets and retweets, especially for information exchange and to signal the content of posts in discussion threads, and so I chose to stream data using this hashtag as a search query. These data form a second potential 'open science network', and will provide a picture of the people engaging in open science-related information exchange and discussion.

Though a very popular open science conference, SIPS, was already over by the time I was collecting data, the AIMOS conference which I mentioned before, was held online in December 2020, providing me an opportunity to tap into those data instead. I used the conference's official hashtag #AIMOS2020 to collect AIMOS-related network data. These data form a third potential network, and should expose a set of people at least clearly interested in OS-, reform- and metascience-related research practices.

I streamed data for both of these hashtags via the Gephi plugin 'Twitter Streaming Importer' developed by Matthieu Totet (2016). The plugin, using Twitter's streaming API, allows an authenticated user to stream tweets in real time, based on the use of keywords, hashtags, or user accounts. It gets the information on the activity between users mentioned in tweets and retweets, and uses Gephi's visualization functionalities to explore the network data obtained.

In each instance, I input the hashtags as strings into the 'words to follow' field and 'added' them to feed them into the plugin as search queries. Applying network logic to the stream in essence applies an algorithm to get the information in a tweet to a network graph in Gephi.

By using the so-called *Bernadamus Projection* logic option, which is similar in function to the *User Network Logic* option (which captures interactions between users based on hashtags used in tweets: Strick, 2020), a tie between users is only created if the tweet, mention or retweet contains certain keywords. For my analysis I chose #openscience or #AIMOS2020 (depending on which network we are referring to). It aims to reveal more the "cluster" of discussions between users and reduce noise, according to Totet (personal communication; December 1, 2020). In the network it generates, nodes are generated when a user tweets using the hashtag specified, and ties are formed when users mention other users in combination with one of the specified hashtags, or retweet their posts (that include the relevant hashtag).

#openscience Stream

As I already mentioned, I collected Twitter user network data using '#openscience' as a search query. These data were collected over a period of two weeks, on each day in that 14-day time-frame, from different times of day and night (given the international nature of the open and reform community) to attempt to get as representative a sample as possible. Data collection windows varied in length from two to several hours in duration. Data collection occurred for a total of 355 hours over 15 days: from Sunday 12pm on December 27, 2020 to Sunday 7am on January 10, 2021.

Although people use #openscience in a number of different contexts, a very common usage of the hashtag in tweets and retweets (as opposed to in their bios) is to share information. This information might pertain, for example, to upcoming conferences, symposia or other events, past talks and presentations, published articles, preprints and blogposts on open science-related topics, and news items (e.g., the introduction and promotion of newly-formed OSC groups or organizations). I was interested in using this hashtag because I surmised that people actively sharing open science-related information would be part of the community to some degree.

Results for Analysis of the #openscience Network

Over the period of data collection for this hashtag, I obtained data for 2392 nodes and 9472 edges, though these data were incomplete: data on friend and follower counts were missing from several hundred accounts collected. This is possibly due to a bug, or because the information on private or 'locked' accounts was not available. I deleted these partial data. The complete data contained 1649 nodes and 6258 edges. The results are given in Table 8.5 and a visualization of the graph in Figure 8.15.

The table indicates that the average user in this network follows fewer than four other users using #openscience in this corpus. The shortest path between the two most distant nodes in this corpus is quite long, at 14. That and a very low density and transitivity of 0.002 and 0.008 respectively, indicate that this network is sparse, and relatively locally disconnected. Reciprocity is very low for this graph, with a ratio of 0.09 meaning that less than ten percent of ties are reciprocated.

This network is smaller than the follow network in terms of nodes, and is a lot sparser, containing fewer ties between users and fewer closed triangles. This is made apparent by a comparison of the density of this network to the follow network (0.002 compared with 0.02, and 0.024 which is the density of the #AIMOS2020 network). Comparing the diameter of the networks is also revealing: the follow graph's diameter is close to that expected for a 'small world' network, where everyone knows everyone, however the diameter of the #open-

science stream network is twice that, at 14. The network is highly modular (Q = 0.827, 139 communities 'detected'), but most of the sub-communities contain fewer than 20 members. A comparatively low reciprocity ratio raises interesting questions about the relations between the set of actors who use #openscience in their tweets. It likely indicates that, contrary to my informal expectations, people who use the hashtag in their tweets are a different subset of people who use 'open science', 'science reform' and 'metascience' and other related terms in their biographies. Certainly, in hindsight I can see that the most active and influential accounts in this network are different to the people who are most active and influential in the open/reform follow graph.

Metric	Value
Nodes	1649
Edges	6258
Ave. Degree	3.795
Diameter	14
Ave. Path	4.295
Density	0.002
Transitivity	0.008
Reciprocity	0.09
Q (communities)	0.827 (139)

Table 8.5: Network metrics for the #openscience stream network.

The network statistics calculated for this graph tell us several possible things about its membership and its structure. The high Q value, low reciprocity and transitivity, as well as high diameter and path length all indicate that the people using this hashtag on Twitter are a highly fragmented group of people. Nodes in this network are not closely connected with one another, which indicates that they may not work closely with one another, or share much with one another in terms of information on this network. Reciprocity is very low also, which further indicates that the users in this network tend to be isolated, and may not know each other, or not participate in larger, more cohesive communities. It is likely that many accounts use #openscience as a way of transmitting information (whether that is to amplify the content others have shared, or to draw attention to their own content), rather than as a means of social connection. This is supported by some literature (see e.g., Grover & Kar, 2020; Hermann, Nehls, Eitel, Barri, & Gammel, 2012).

News or Social Media? From looking at this graph, we can also learn that bots and organizational accounts have interesting and strong influences over the structure of the network. This also feeds into the idea that the use of #openscience

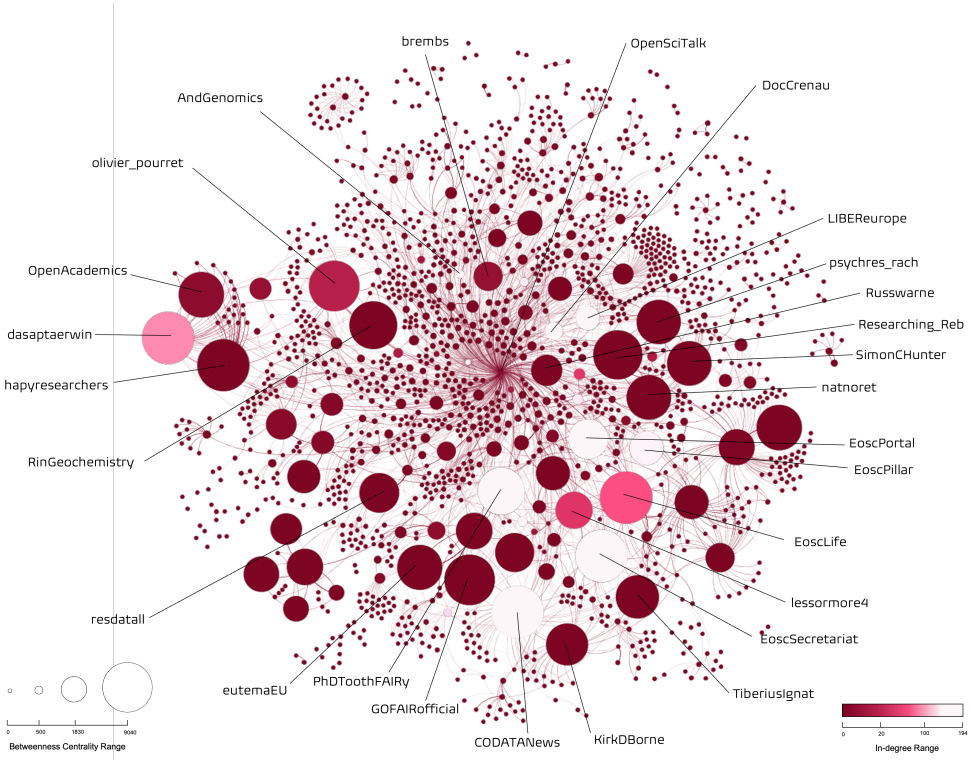


Figure 8.15: Network graphic for the #openscience hashtag graph. Color denotes in-degree range (paler colors indicate nodes with higher in-degree scores), while node size denotes betweenness centrality score range (larger size indicates nodes with higher betweenness centrality score.)

places this platform in the role of an information and news media site, rather than a social media site, a topic explored thoroughly by Kwak, Lee, Park and Moon (2010). Automated and organizational accounts are very active on Twitter, retweeting and following at abnormally high rates. They exert a lot of influence on the network structure, particularly when it comes to centrality and out-degree scores. This is no novel finding, however. Schuchard, Crooks, Stefanidis and Croitoru, for instance, demonstrate this effect of bots on social media network structures in an interesting empirical study (2019), showing that bots, though a minority in the group, strongly influence the network with very high eigenvector centrality scores due to attempts to initiate contact with other users through (for example) retweets at a rate much higher than human users. The exploratory findings for this data set of user accounts support their conclusions, which is evident in the below table which shows the top 20 accounts in this network ranked by eigenvector centrality.

The varying ways in which open/reform actors use Twitter is interesting to consider in the context of the idea of the CoP. In the function of a CoP (or a constellation of them), information transmission is just as important as the formation of ties due to shared interests or familiarity (because these elements contribute to reification and issues of boundaries and negotiation of meaning, for instance). It is interesting to see that bots and organizations tend to be the ones most active in passing on information within the network (while human accounts are perhaps more active in homophily-related, social network activity on this platform).

Rank	Handle	Eigenvector	Betweenness
1	CODATANews*	1	5696.95
2	EoscLife*	0.88	1610.90
3	dasaptaerwin	0.84	1831.29
4	eoscsecretariat*	0.83	3216.21
5	hapyresearchers*	0.80	0
6	GOFAIRofficial*	0.63	0
7	olivier_pourret	0.62	1016.57
8	researching_reb	0.53	43.70
9	ringeochemistry*	0.47	0
10	phdtoothfairy	0.44	3976.02
11	openacademics*	0.39	376.71
12	virusesimmunity*	0.39	0
13	eutemaeu*	0.38	0
14	natnoret	0.37	23.69
15	simonchunter	0.36	20.73
16	psychres_rach	0.36	47.889
17	tiberiusignat	0.33	0
18	kirkdborne	0.31	0
19	eoscportal*	0.28	3567.21
20	eu_h2020*	0.27	0

Table 8.6: Top 20 #openscience users ranked by eigenvector centrality score (EC), with associated betweenness centrality scores listed. Automated (bot) or organizational account denoted by *.

Betweenness Centrality and In-Degree The network graph highlights an interesting contrast between this betweenness centrality (which indicates actors which may have a brokering or bridging role in the network) and in-degree (a measure of popularity). These two metrics tend to be moderately to strongly correlated (see e.g., Valente, Coronges, Lakon, & Costenbader, 2008), but in the case of this network, we find that they are more weakly correlated than expected ($r = 0.33$). The network visualization is consistent with this. The figure shows

both betweenness centrality and in-degree: The former is denoted by node size (larger nodes indicate higher betweenness centrality scores), and the latter, by colour intensity (paler shades indicate higher in-degree). One would expect the lighter nodes to be larger in the case that an increase in one variable is associated with an increase in the other, but this is only true for some of the nodes. For instance, @dasaptaerwin, @eoscsecretariat and @codatanews are all nodes which have high betweenness centrality scores, and high in-degree. Compare this with users such as @andgenomics and @openscitalk, which self-describe as broadcasting bots (i.e., their goal is to amplify and transmit information across and within groups of users on Twitter). The former has the sixth highest in-degree in the network, meaning that it is a popular account in this corpus, but it has a relatively low betweenness centrality score (its score is 11, where the range of scores is between 0 and 9040). This means it does not sit on the paths between other nodes very often and is unlikely to play an kind of brokering role, at least in this hashtag network, which somewhat defeats its purpose as a community broadcaster.

The other account, @openscitalk, shows the opposite effect: It is ranked relatively high on betweenness centrality (with a score of 410), but its in-degree is only 6 (on a range between 0-194). This bot account is not popular in this corpus (though it is active in its following behavior, with the second-highest out-degree of all the nodes), though it sits a position between many nodes, which the network graph shows (it is positioned very centrally, and it is clear to see all of the outgoing edges which emphasize its prolific following behavior). This combination of characteristics indicate that this bot is in a position to be an efficient information transmission account, were it to be targeted by a higher number of accounts in this corpus. Please note that when I say "in this corpus", I mean to delineate this network from the broader network of open science and reform chat on Twitter – this is only a small sample of that possible network, and I can only discuss certain nodes in the context of this particular network, in which they feature[23].

Eigenvector Centrality and Betweenness Centrality The table showing the top 20 most influential users tweeting using #openscience gives us a sense of what users dominate hashtag-driven engagements on Twitter. Note that this does not mean that these accounts dominate the most relevant or meaningful discussions relating to open science and reform, just the discussions which feature the hashtag. Although I have noticed that some of the more important or relevant discussions relating to the community's identity and boundary formation (for instance) use the hashtag, I have observed that it is much more common for the hashtag to be used in posts which share information and news within the network.

[23]That said, this particular example might extend to the broader network. The account @OpenSciTalk is a relatively less well-followed 'open science' broadcaster bot (under 3.2k followers), compared with its siblings @openscience (71.5k followers), or @_open_science_ (12k followers)

Use of hashtags to help amplify posts is a well-known technique, because automated accounts retweet using them as a keyword as they crawl Twitter. Including '#openscience' within a tweet, for example, will prompt bots @openscience and @OpenSciTalk to retweet that post (some accounts' bios explicitly say which hashtags will automate a retweet, for instance, this one's bio includes the text: "Use #OpenScience or #OpenSciTalk for RTs.").

Accounts like @CODATANews and @EoscLife are users which, like the broadcasting bots I mentioned (which are programmed to automate amplification of tweets by retweeting hashtag posts, and follow many accounts, partly with the aim of being followed back), are designed to transmit information across the network. They are accounts which are either automated or partially automated (that is, control over posts and retweets is shared by a human and bot – a bot will automatically share some content, but the bot's posts are supplemented by a human user), and are often associated with an organization of some kind that has a mission or goal to enrich or facilitate community development.

For example, the website associated with @CODATANews has the following text on its homepage: *"As the Committee on Data of the International Science Council (ISC), CODATA helps realise ISC's vision of advancing science as a global public good. CODATA does this by promoting international collaboration to advance Open Science and to improve the availability and usability of data for all areas of research."*. The remit of CODATA is very broad, as its target audience is scientists across the globe, which naturally places it in a role involving brokering, or at least bridging, between existing sub-communities. This node has a high betweenness centrality score, which supports the possibility that succeeds in a kind of bridging function in this hashtag network, and possibly beyond. Account @EoscLife is another interesting example of an influential node. The website linked in the profile says of the organization: *"EOSC-Life brings together the 13 Life Science 'ESFRI' research infrastructures (LS RIs) to create an open, digital and collaborative space for biological and medical research."* This account not only is highly influential in this hashtag network, it is likely to bridge the medical and life sciences communities and the general open science communities, a conclusion supported by the node's high betweenness centrality score.

Global Influences I have already discussed that the open science and reform constellation of communities of practice is somewhat globalized. That said, open science, science reform and metascience efforts do not seem well represented by some groups (An extreme example of this is the continent of Africa, which is greatly under represented in science: North, Hastie, & Hoyer, 2020) and are concentrated in hubs, to some degree. For instance, areas in the Netherlands, Germany and the UK have been heavily involved in the grassroots journal-club initiative 'ReproducibiliTEA', which first began in Oxford, UK and has since spread to 109 different institutions, 78 of which are held in European localities.

The 'Open Science Communities' are another example of Europe-based initiatives. These are independent grassroots communities that have popped up in multiple places, particularly concentrated in the Netherlands. Another hub can be found in Melbourne, Australia, where the influential Melbourne Open Research Network and research group behind the RepliCATS project originated and is based.[24] Open science researchers from Melbourne are also behind AIMOS, the Association for Interdisciplinary Meta-Research and Open Science.

Unsurprisingly, this network reinforces this picture of a globalized constellation of communities. Diving into each of the profiles of the human accounts on the top 20 list, that the most influential and central nodes in this hashtag network represent a range of different countries. This network's most influential human account, @dasaptaerwin, belongs to Dasapta Erwin Irawan, a hydrogeologist based in Indonesia, working on RINarxiv, Indonesia's preprint server. Olivier Pourret, owner of the seventh most influential account in this network, @olivier_pourret, is a geochemist based in France. Rebecca Johnson, Simon Hunter and Rachel Taylor (@researching_reb, @SimonCHunter, and @psychresrach) are all researchers based in Glasgow, Scotland. Esther Plomp (@PhDToothFAIRy) is based in the Netherlands, Nathalie Noret (@natnoret) hails from York, UK. Tiberius Ignat (@TiberiusIgnat) is based in Munich, Germany, and Kirk Borne (@KirkDBorne) in Maryland, USA.

#AIMOS2020 Stream

These data were streamed from Twitter for a period of time surrounding the AIMOS 2020 conference, as mentioned before. In this context, I was interested in users discussing the AIMOS conference using its official hashtag, #AIMOS2020, and followed the same method as described before (using the streaming importer, the same projection method and entering the keyword into the query field).

The conference itself was held on 3rd and 4th December, however, in my experience, discussion online about conferences tends to begin a few days before, with people who are looking forward to the event sharing their excitement, rallying of other members, planning of social events (naturally largely online due to not only the nature of a conference of this size, but also due to COVID-19 restrictions affecting the entire world), and providing information for potential last-minute sign-ups.

The discussion tends to take several days after a conference to come to a close, too, as people like to use Twitter to reflect and reminisce, so it is good to cast a broad net. I began streaming using the official AIMOS hashtag at 11pm on December 1, and allowed the stream connection to stay open until 11pm on December 7, during which period data were continuously streamed from Twitter.

[24]Funded by DARPA, the project represents an enormous crowd-sourced effort to develop and test methods to produce accurate predictions about reproducibility of social science research findings, with 4000 papers ultimately assessed.

I did not experience any WIFI outages during this time (connection issues throw an error message in Gephi, and stop the streaming process so I would be made aware of them), so I do not anticipate any data were lost.

Description

I began streaming from the Twitter API three hours before the conference began, and ended it 24 hours after the conference ended. In this period, I obtained network data involving 147 nodes and 523 edges. A given node in this network will have an average of fewer than 4 interactions with others, and its density is 0.024, which means that this corpus, compared with the following hashtag network, is relatively well-connected. Again, in comparison with the #openscience network, it is close to a 'small world', with a short diameter and average path length. These network properties are listed in Table 8.7, and the network graph in Figure 8.16.

Metric	Value
Nodes	147
Edges	523
Ave. Degree	3.558
Diameter	7
Ave. Path	2.945
Density	0.024
Transitivity	0.18
Reciprocity	0.16
Q (communities)	0.47 (5)

Table 8.7: Network metrics for the #AIMOS2020 stream network.

Analysis

Naturally, this network's graph is dominated by the AIMOS account (@aimos_inc), which was retweeting every post which included the conference hashtag. Other influential nodes in the network are associated with accounts belonging to Fiona Fidler and Hannah Fraser (an Australian researcher with heavy involvement in Fidler's work and initiatives, and a history of working with her). This network is fairly modular, with a *Q* value of 0.47 (the algorithm 'detected' five communities). The sub-communities are indicated in the graph by different colours (see the caption on Figure 8.16), and eigenvector centrality is indicated by node size, where larger nodes have the highest scores.

Brokering Some nodes have a high betweenness centrality score because they are very central nodes (and naturally lie on paths between other nodes), but oth-

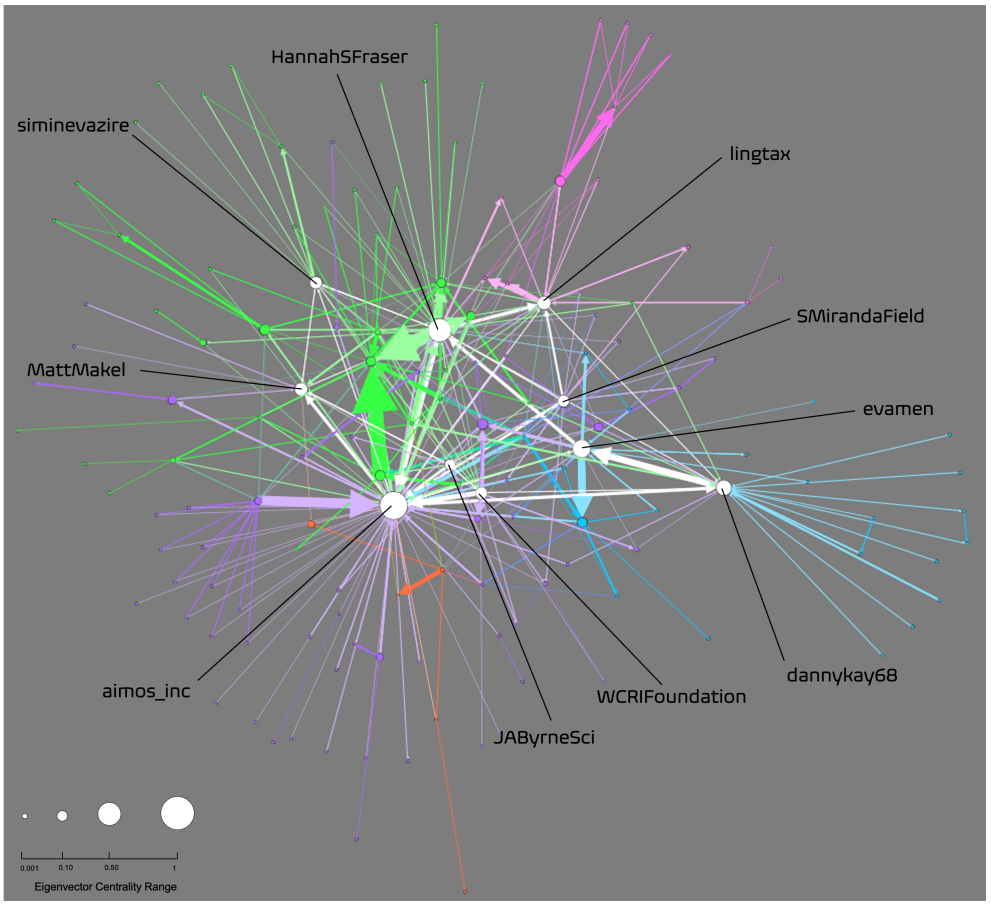


Figure 8.16: Graphic showing #AIMOS2020 stream network. Colour indicates communities (the modularity solution yielded five, denoted in purple, pink, blue, green and orange), and white nodes indicate the ten nodes with the highest betweenness centrality score (the potential structural brokers). These top-ten nodes have also been labeled with their Twitter handle. Node size denotes influence in the network (larger size indicates higher eigenvector centrality score).

ers are in such positions because they have ties to multiple communities (and likely have a brokering function), rather than just sharing connections with other sub-community members. Actors such as @fidlerfm (Fiona Fidler), @evamen (Eva Mendes), @SMirandaField (myself), @lingtax (Matthew Ling), @mattmakel (Matt Makel) and @siminevazire (Simine Vazire) have ties to at least two sub-communities within this graph, which you can see from Figure 8.16. This, along with their high betweenness centrality scores, indicates that they are likely to receive and/or transmit information from and to multiple different sub-communities of people (Abbasi et al., 2012). Whether this means they are brokers in *Wenger's* sense of the word is a conclusion I cannot draw with these data, however.

Psychology, Open and Metascience This network, in my estimation, comes closest to approximating the 'metascience' community members who are active in the in-person events I have attended. Actors like Simine Vazire, Brian Nosek, Fiona Fidler, Matt Makel, Lisa Spitzer, Julia Rohrer, Dorothy Bishop, Chris Hartgerink and Sophia Crüwell are people I have come into close contact with at SIPS and Metascience conferences in the past, and with whom I have had in-person discussions about relevant issues. In this group are many individuals that, over the course of my ethnographic study, have had clear roles in steering early reform activity in the open, reform and metascience community. Their areas of research are in, or closely border, psychology.

I will discuss a few of these actors to illustrate their roles in the community, especially earlier on. I have already discussed Nosek's crucial role in kicking off many large-scale initiatives. Simine Vazire is Editor in Chief of SIPS's flagship journal, *Collabra: Psychology*, and has been central actor in establishing the improving psychology community through SIPS. Fiona Fidler heads AIMOS, the Melbourne reproducibility network MORN, and is the principal investigator of the large-scale project RepliCATS that I have discussed previously. She has also been present at each of the SIPS I have attended (i.e., all barring the first, in 2016). Hartgerink is one of the people behind statcheck, which is software programmed to find inconsistencies in reported statistics in articles. This software has been the center of much discussion in the community, especially soon after its release in 2016, and the sticker advertising it is usually among those stickers seen on the laptops of 'open science people'. Matt Makel is first author of one of the first publications to raise concerns about the fact that replication studies are conducted so rarely in psychology (*Replications in Psychology Research: How Often Do They Really Occur?* Makel, Plucker, & Hegarty, 2012), while actors like Ling, Spitzer, Rohrer and Holcombe are active in Twitter discussions and contribute to the literature on topics surrounding reform in (psychological) science.

Although I think it is likely that this particular network reflects some of the original 'open science' group membership in terms of its composition, it does not have as high a reciprocity statistic as I would expect.

Global Representation As with the #openscience network, this network represents a globally diverse group of people. People in this network represent Australia, New Zealand, the USA, Mexico, Canada, Brazil, Germany, the Netherlands, France, Spain, Portugal, Norway, Czechoslovakia, the UK, China (Hong Kong and Taiwan), Thailand, Indonesia, the Philippines, and India.

8.7 Summary and Conclusions

Quantitative Comparison

I will not dwell too long on comparing the three networks in terms of their quantitative metrics, however a quick look at these metrics side-by-side (given in Table 8.8) provides an interesting picture. Although the open and reform follow graph is somewhat sparse, the #openscience graph is extremely sparse and lacking in cohesion by comparison (densities calculated for each of these networks are 0.02 and 0.002, respectively). With a density of 0.024, the #AIMOS2020 network is the most dense of the three. Once again, network size is an important factor to keep in mind when comparing networks on density and diameter, however, the #openscience hashtag streaming graph is almost a thousand nodes smaller than the follow graph (meaning that, quantitatively speaking, we would probably expect it to be denser than the follow graph, rather than sparser). The #openscience network is the most modular, with 139 communities 'detected'. The other two graphs are similarly modular to one another: The software 'detected' five sub-communities in each.

	Network		
Metric	Follow Graph	#openscience	#AIMOS2020
Nodes	2246	1649	147
Edges	100128	6258	523
Ave. Degree	44.581	3.795	3.558
Diameter	8	14	7
Ave. Path length	2.661	4.295	2.945
Density	0.02	0.002	0.024
Transitivity	0.22	0.008	0.18
Reciprocity	0.49	0.09	0.16
Q (communities)	0.323 (5)	0.827 (139)	0.47 (5)

Table 8.8: Table comparing network statistics for each of the three primary graphs analyzed in this chapter – follow graph, and two streamed hashtag graphs (#openscience and #AIMOS2020).

There is overlap between these networks, though not as much as one might

expect. Only 10 accounts appeared in all three networks. Twelve accounts appeared in both the hashtag networks, 26 appeared in both #AIMOS2020 and follow networks, and 28 accounts appeared in both the follow and #openscience networks.

Concluding Analyses

There are a few different hypotheses and questions which arise from comparing these different graphs, concerning validity of the methodology, and whether we can infer anything from them about their appropriateness as proxies for describing the greater reform and open constellation. One overarching question that is worth considering concerns what the *function* of the conceptual greater reform/open science network might be. Is it primarily an *information network*, in which nodes act to broadcast and amplify news and information about open science, metascience and the reform movement within the community? Or is it, in contrast, a *social network*, where friends and colleagues connect with one another to provide support, belonging and to share identity-affirming content? The skewed distributions of the degree data for each network supports the former, say Myers and colleagues (2014), as does the sparse connectivity of the #openscience network.

The latter function (i.e., of the social network) is supported by the shorter diameters and higher densities of the follow graph and the #AIMOS202 hashtag network. I argue that the most plausible explanation is that *both* functions for tie formation – news and social interests – are at work, as earlier articles have found (Kwak et al., 2010). People on Twitter use it for various reasons, and there is no expectation or requirement for it to be *either* an information medium *or* a social medium. For some reformers and open science people, Twitter facilitates mobilization and activism for the movement. These people can rally and increase alignment with their purpose by broadcasting information and news on a wide scale. At the same time, people connect with their friends and colleagues on Twitter, and become friends with Twitter acquaintances as they discover people they share interests and life and work philosophies with. People discover life partners, co-authors and jobs, and expose themselves to perspective shifts as well as being polarized in their 'bubbles'. These are all functions of social networks.

Coming back to an earlier point I raised, how suitable is the use of bio keywords when it comes to identifying members of the group? This question comes to the fore when we compare these networks, too. Many people readily self-identify, but at the same time, not everyone who self-identifies as a 'reformer' or open science advocate truly acts in these capacities, participating in their community along with other members who similarly self-identify. Many people choose against self-identifying, or do not think much of using adjectives like "science reform" and "reproducibility" in their Twitter bios to do so. I argue that both mechanisms are at play for this constellation, and it means that defining the group

by these keywords is likely a method too flawed to be genuinely useful, at least if used on its own. That said, the use of these keywords yielded a valid set of sub-communities, which indicates that it is at least one sound approach.

Similarly, does using the hashtags #openscience and #AIMOS2020 produce a faithful snapshot of one or more facets of the group structure or its membership? The number of nodes in the #AIMOS2020 network is close to the number of people who participated in the first day of the conference (i.e., 150 people), according to the conference organizers. Even if this number is not exactly accurate, I consider that, combined with my observation and experience during my ethnographic work, a sign that the hashtag network approximates that sub-community of conference attendees (and, thus, a true sub-community within the greater reform/open community) relatively well. If so, does this hashtag network mean that its corresponding sub-community in the constellation is a small-world network, where everyone knows everyone? Where information is transmitted throughout the network quickly and easily, and where people work and communicate closely with one another? The @openscience hashtag network is a different story. This hashtag is heavily used, but is it used by actors who are true participants in the constellation? Is this fragmented, sparse, widely dispersed engagement network a reflection of the true nature of the reform and open science constellation?

I argue that both possibilities are likely. In the light of the findings of the previous chapter, I argue that if one attempts to consider the reform, open and metascience constellation as a single community, one will discover a group of people that is highly fragmented and sparse, with a lack of evidence of people working together on a single thing. If, instead, one conceives of this group as a constellation of communities of practice, we will see a much different picture. We will see that although the constellation ultimately unites in their shared enterprise of 'scientific improvement', it is more fruitful to emphasize the dense communities of practice that stud the landscape like clusters of stars in the sky. Communities like the one I captured using the 2020 AIMOS conference hashtag, for instance. Communities like the five the algorithm 'detected' when it was run on the large follow graph. I will explore this concept in more detail in this dissertation's conclusion (**Chapter 9**), and consider the possibilities opened up by the introduction of the concept of a constellation.

9

Conclusion

9.1 Summary of Findings

The research 'problem' I entered this doctoral topic of study with was broad, complex, nebulous and, at times, intractable. It demanded that I keep an open mind, and remain neutral and receptive to multiple and varied streams of information and perspectives. It was well-suited to an exploratory, mixed-methods approach, and indeed, that is the nature of the research I present in this book. As such, this 'discussion' section does not contain any mention of predictions or expectations. In this final chapter, I summarize my ethnographic and network analysis findings, then convert some of my interpretations into recommendations. In doing so, I consciously shift my stance from a participant-observer to reform activist, though I try not to completely leave neutrality behind.

Ethnography

Constellation of Communities of Practice

Before undertaking the ethnographic analyses reported in **Chapter 7**, I had long referred to a community of practice, or an 'open science' or 'reform' community. Such descriptions attributed a homogeneity to the group that does not exist. That description belied the complexity and diversity that these people represent in how they perceive their enterprise, and how they approach their practices. Undoubtedly, the most salient conclusion I made in the light of the ethnographic material I explored was that the open science, metascience and reform collective is best described by Wenger's 1998 concept of a 'constellation of communities of practice'.

Recall, from **Chapter 7**, that the communities of practice (CoPs) within a constellation share an enterprise (or have related enterprises, if not the same one), and share a common cause, as well as some members. The constellation provides a space which allows for multiple and competing discourses to exist and influence one another. The ethnographic findings certainly expose each of these

elements in the open and reform group. Their shared enterprise – the betterment of science – is pursued in varying ways across the constellation, and contributorship is defined separately for each CoP, depending on their own competencies and priorities. For instance, some CoPs in the constellation prioritize diversity, inclusion and tone, while others focus on improving methodology and statistics. Some of the CoPs pursue their version of the enterprise using elements of quantitative practice, while others work from their roots in qualitative traditions. The landscape of this constellation is heterogeneous and complex.

Some communities such as those associated with the Psychological Science Accelerator, for instance, are putting a great deal of time and effort into exploring the 'team science' approach on a global scale. Others in philosophy of science focus on deep meta-issues like when and whether and how a research result can be treated as evidence for a scientific claim. Yet others who study scientometrics consider the role of academic achievement indicators (such as the h-index and impact factor) in academia, and explore how scientometric evaluation impacts practice in science. The work of these CoPs varies in terms of how applied or fundamental the approach is. Two examples that are close to me and my own work (current and future) give a sense of what I mean here. The Nanobubbles project (spanning France and the Netherlands),[1] funded by the European Research Council seeks to understand how the process of error correction of science works, and when it fails. The project combines the natural sciences, engineering and the humanities and social sciences in its interdisciplinary approach to the problem. Another collaboration, the Responsible Research Project,[2] takes a practical approach to developing integrity and robust research competency across multiple scientific disciplines. Led by researchers in the Netherlands and the UK, this group aims to establish a community of practice across Europe and the UK, centering on responsible research practices.

At the same time, but each in very different ways, the various CoPs of the reform and open space pull the movement toward better science forward. They contribute differently and uniquely to the common goal of reform, and their competencies overlap and complement each other even as they struggle in disagreements on Twitter.

Power and Dominance, and the Collective Identity

I observed that power is unevenly distributed across the constellation. Some CoPs claim dominance over others, with their priorities weighing more heavily in the equation of the collective identity than others. Historically, we have seen an emphasis on positivist and methods-focused traditions in crisis discourse. Consider the strong focus on replication and preregistration, and other quantitative open

[1] See the group's blog here: `https://nanobubbles.hypotheses.org`.

[2] See `https://www.universiteitleiden.nl/en/news/2021/10/a-community-of-practice-in-responsible-research-across-europe`.

science practices, for instance. The distinction between 'good' and 'bad' science is often made, and the definition of 'good' is often very restricted. The foregrounding of these elements of reform practice has overshadowed and diminished others, such as diversity and equality, and discussion on tone.

Despite this, the more marginalized CoPs are starting to assert their own agendas into the collective focus. The idea of an 'open science buffet' has gained momentum, and the focus has turned to making science more accessible to those who have been sidelined in the past (e.g., disabled researchers, and those based in developing and poorer countries). There is clear tension in the reform and open science space, and it is clear that the collective identity is in the process of negotiation between member CoPs in the constellation.

Pulling at the Moorings of Tradition

Member communities of practice continually work at shaping their own identities, as well as the collective enterprise. They push and pull amongst themselves, all the while trying to break free from (and, in some cases, completely dismantle) the structures of traditional academia, and wrest control from *it*. We see elements of this part of the collective enterprise everywhere, from the establishment of novel publication platforms like ResearchEquals (which completely turns the traditional publishing system on its head by instantiating a 'pay to close' model)[3], and Octopus, to the introduction of new article submission types like registered reports (which are accepted at 300 journals world-wide, as of April 18, 2022) which explicitly build in peer review of the planned research methodology which is conducted before the data are collected and analysed. We see open research methodology being taught to undergrads and master students at universities all over the world, and citizen and crowd-sourced science initiatives are gaining more funding and exposure all the time.

Network Exploration

Bruckman argued that embracing the idea of communities with fuzzy boundaries, and delineating groups by their members' characteristics is the best way to approach the issue of online community definition. The network analysis I conducted was in the spirit of her recommendation. I collected Twitter follow and friend data based on searches made using terms that the group used themselves in their profiles and in their posts on the platform. I focused on users who included adjectives such as 'open science', 'metascience' and 'science reform' in their biographies, and those who used certain relevant terms in their posts. Comparing

[3]Whereby authors have to pay a fee if they want a more restrictive license on their work. For a completely open licence – such as a CC0, meaning that no rights are reserved and the author places their work entirely in the public domain – publishing is free. For an all rights reserved license, however, authors must pay ResearchEquals 549.99 Euro.

the networks produced by these different search parameters formed the basis of my network analysis.

Comparison by Network Metrics

When comparing different networks on the metrics I obtained through Gephi and NetworkX, several things became salient. First, and most important is that the network analysis supports the idea of the online open science and reform network being, in fact, a constellation of sub-networks. The follow network, a large, broad network of 2246 nodes (the users within it were obtained via a search of biography terms, see **Chapter 8**), was relatively sparse and modular. Analysis revealed a possible structure to this network, which included five sub-communities (of which four were large and well-defined). Two of these four resembled 'small-world' networks. In other words, the analysis indicates the existence of communities of practice within a broader constellation defined by certain important keywords. That said, the four sub-networks were relatively modular themselves, each with at least four communities subsumed in *their* structure. This may indicate that there are at least twenty separate communities of practice within the constellation, which can be broadly grouped into scientific discipline, or by some other grouping variable.

News Medium or Social Network?

Although Twitter is often called a social networking site, some research (see e.g., Kwak et al., 2010) has advanced the idea that people and other entities use it just as often as a news and information-broadcasting medium, and that it is frequently used as both. Although I did not formally explore this hypothesis, my network analysis supports the idea that the open and reform constellation makes use of Twitter in both its possible capacities. Automated and organizational accounts tend to use Twitter to broadcast information and amplify the content of others, depending on the text of posts, while human users tend to follow people they know in some capacity (friends, family and coworkers), or people who post content that interests or amuses them. This was indicated by low reciprocity and dominance of automated accounts in the #openscience network, compared with higher reciprocity and dominance of human accounts in the algorithm-detected sub-communities (communities 1, 3 and 4 in particular showed characteristics of a social network usage).

Globalization

The #AIMOS2020 and #openscience networks I captured using separate live streams gave evidence of a somewhat globalized open science and reform network (though, some countries like the UK and USA do dominate). The 20 most

influential user accounts in the #openscience network alone represented 7 different countries, while the #AIMOS2020 network represented nineteen different countries. As many researchers have emphasized (see, for instance, Merle, Reese & Drews for a study on global identity and the online social world: 2019), Twitter is highly globalized. Merle and colleagues say that Twitter and sites like it have "become channels to express political opinions, and communicate with people across the globe." (p. 2019, *ibid*) Since the start of the pandemic, it is likely that this is even more the case, as scientific societies and institutions have had to turn to hybrid or completely virtual conferences and workshops. The AIMOS, SIPS and Metascience conferences in 2021 all had the listings for their sessions and talks in multiple time zones to facilitate the variety of different people who had signed up.

The somewhat global character of the open and reform science constellation is a reflection of the broader academic context out of which it has emerged. On one hand, it represents the fact that science is a global enterprise: Many benefit from scientific advancement, and people from many countries participate in its practice. That said, the constellation reflects the under representation of some groups in science. Consider the continent of Africa, for instance, which is greatly under represented in scientific practice and scholarship, despite having the largest population by continent in the world (North et al., 2020).

Study Limitations

Another salient finding revealed by the network analysis is that the whole open science and reform constellation is difficult to capture using this method. This is likely because there is no formal community, and no single reliably- or even widely-used descriptor which one can search Twitter's user data bank. Hashtags like #openscience, and nouns like 'transparency' and 'reproducibility' allow for the collection of a decent proxy network (or networks), but it is still the case that these networks are hazy, distorted reflections of the *true* online open and reform constellation (if there were to be such a thing). Necessarily, some people are missed. Some people, as I mentioned in **Chapter 8**, do not self-describe using any of the adjectives or nouns I queried. Many rarely use hashtags like #openscience, and some not at all.

This is undoubtedly a limitation of any quantitative approach to the problem of delineating an informal community. Unless a group of people can be reliably categorized by some kind of quantifiable descriptor, this issue will remain unsolved.

The quantitative approach I have used in this dissertation to explore the constellation of open, metascience and reform communities of practice is not the only one with limitations. Entire communities of practice seem to be overlooked even after three and a half years of ethnographic research. Take the example of forensic science reform. As Duke University School of Law professor Dr. Brandon

Garrett's fascinating book *Autopsy of a Crime Scene* (2022) details, the forensic science reform community is a large and active one. Their movement, which centers around reforming forensic science methodology, and challenging the criminal justice system's unhealthy reliance on it, predates the one I have been studying by *at least* two decades. Their efforts have not come to my attention during my ethnographic study. I only discovered this community through randomly stumbling upon the narrated version of Garrett's book in Audible, an audio-book app, despite members of the reform community such as Jason Chin and Lindsay Malloy being in or adjacent to such topics.

9.2 Concluding Reflections

The findings of this doctoral work revealed that the collective of people who study, practice and advocate for scientific reform are not a single cohesive, homogeneous community with a coordinated practice in service of achieving a joint enterprise. I have presented arguments driven by qualitative and quantitative evidence that *this group are better categorized as a constellation of multiple, diverse communities of practice*, with shifting, mutable boundaries, in which a collective identity and enterprise is being continually negotiated. Each CoP in this constellation, though united by the broader joint enterprise of improving science, coordinate their own varied approaches which are characterized by different competencies and academic traditions, and driven by a diverse set of interests, priorities and needs.

A constellation is a complex social structure, certainly, how could a group of interconnected but diverse communities not be? Wenger says that "as communities of practice differentiate themselves and also interlock with each other, they constitute a complex social landscape of shared practices, boundaries, peripheries, overlaps, connections and encounters" (1998; p. 118). I argue that the open and reform space is especially so. For one thing, the joint enterprise is a constantly evolving thing by nature, because that enterprise *is* reform. That means that power within the constellation is dynamic; which CoPs are dominant is subject to constant change.

Additionally, 'locality' (Wenger speaks of individual CoPs as being local; the wider configuration of a constellation can be thought of as global) is complex. It is likely that most CoPs in the constellation are widely distributed in terms of geography (though certainly, some CoPs such as the UK reproducibility network for instance, find cohesion in being geographically close), despite ties of academic tradition or discipline, or links between one another due to participation in virtual platforms like Twitter. Another factor contributing to complexity is that these CoPs are not only connected within the constellation; they also exist within the constellation which is a component of the wider academic community. As such, the constellation's member CoPs are sensitive and subject to academia's social

structures, tensions and demands in addition to those of the constellation.

Current Challenges and Recommendations

This complexity brings with it challenges and roadblocks to the success of the joint enterprise. I consider some of these in this section, and provide recommendations for how the various CoPs might shift their perspectives and practices to accommodate the unique needs of the diverse constellation, with the ultimate goal of propelling the reform movement forward. Over the course of my ethnographic study, I have recognized a few key areas for improvement for this – my – field. The literature is also rich with examples of problems that this movement faces. I have identified a lack of true discussion among CoPs. There's plenty of talking, plenty of *discourse*, but little genuine engagement or listening. I, and others I have spoken with have seen a lack of self-reflection and humility. There has been a lack of engagement about the issue of diversity and inclusion from many CoPs, which some CoPs have pointed out.

Talk Versus Discussion

"The good news is that everyone is talking...". Koehler (p. 5; 2010) was referring to progress in forensic science reform. Although my ethnography revealed much *talking*, I often failed to see real evidence of *listening* among different CoPs in the constellation. Koehler's 2010 article communicates an (albeit cautious) optimism about major structural reform in the forensic sciences, but I admit I am pessimistic about whether just talking will lead to progress, as Koehler suggests it can. A fruitful conversation requires coordinated communication, with receptivity at both ends, rather than just talking loudly at cross-purposes (which is, at times, what it seems like). Listening in the face of contrasting perspectives and backgrounds requires a true time and effort investment, even when it is just between two people. What a monumental effort fruitful communication would require between multiple, overlapping, diverse CoPs!

Despite these barriers to effective communication between CoPs, many people occupy brokering positions within the constellation. People like Richard Morey, and Lisa DeBruine, for instance, who seem to be able to seamlessly interface with multiple CoPs, talking about R packages and statistics with modellers and programmers, while at the same time championing and supporting diversity and good tone in scientific debate. As the reform movement continues to grow and gain adopters, it is likely that more people take up the role of brokers, and perhaps it is possible that over time, talk gives way to conversation.

Diversification and Inclusivity

WEIRD Reform Diversification is a painful topic for many researchers who are working using few resources to improve science, because despite the international

nature of the constellation I highlighted earlier, many actors and groups are concerned about a US- and Euro-centric reform movement. For instance, there has been much concern about the fact that the SIPS conferences have so far only been held in the continental US and Western Europe (2016-2019), and that two of the next are to be held in Canada and Italy. SIPS has responded with a call for proposals for where to host the 2024 SIPS conference. Included in the call's text is the following passage: "When major conferences are held in only one or two geographic regions, less financially secure scholars are systematically excluded, which limits the exposure of their perspectives to an international community." SIPS's global engagement task force (put together to respond to concerns about a lack of inclusion and accessibility after the 2019 Rotterdam conference) recommended that "SIPS hold the annual conference in geographically diverse regions, including those traditionally labeled as 'Global South', 'Low and Middle Income Countries'."

SIPS and other formal societies within the constellation have a long way to go, this gesture notwithstanding. Although it is a good step forward, the call for proposals is disappointing, in my opinion, because it places quite a burden on people with existing barriers to participation. The proposals need a great deal of detailed information, and for most regular people (with limited experience with event management), would take some time to put together. Alongside other information (such as a discussion of why the author of the proposal is able to be a suitable host of the conference), proposals need to include information on where the conference would take place in the requested host country, whether the proposed facilities are adequate for a certain number of participants, whether the site has spaces that facilitate networking and informal exchanges, and is close to eateries, lodging, and transportation. The proposal must also include an approximation of the cost required to reserve the site for the conference. Even more information is needed which addresses potential advantages of the conference location such as accessibility, funding sources which can be applied for, and SIPS groups in the area. Proposals must provide a discussion of potential limitations of the host location, such as laws or local ordinances that target marginalized members of the SIPS community and potential visa restrictions.

Attendees of past meetings in the US and Europe have not been required to do the legwork needed to fill in one of these proposals – SIPS did all of that for them. It is doubtless difficult to access some information about a potential travel location if you are not a local there yourself, but to me, it seems inequitable for people in disadvantaged positions to have to do the hard work if they want SIPS to be held somewhere that they can easily attend. To me, this is an attempt at accessibility and inclusivity that falls dreadfully short of its supposed goal. I have heard SIPS attendees talk about the problems of WEIRD research (i.e., research that is focused on and conducted largely by western, educated, industrialized, rich and democratic peoples), yet it is difficult for this society to break an existing pattern of WEIRD conferences.

Infrastructures Many in the reform movement aim to reform as much of science as possible, not just positivist scientific traditions for whom replication is most applicable (though as I have previously argued, positivist traditions dominate, and much of the discussion is led by those in positivist disciplines). That means that the open and reform constellation's discourse needs to open and adapt so that learning can take place about how best to serve a range of scientific traditions, and how to meet other research goals and needs. This brings in the issue of infrastructures. What infrastructures exist for supporting open and reform goals tend to fail to recognize that not all research material can and should be digitized (and, for that matter, that some processes surrounding research material production and storage cannot be easily *digitalized*). Additionally, existing infrastructures need to be improved to facilitate as many different kinds of researchers as possible, including qualitative researchers, those with disadvantaged backgrounds, and older scientists. Openness on its own does not bring the joint enterprise forward. Training and education, useful and usable infrastructures, and administrative support must accompany sharing of data, materials and code.

Self-criticism and Humility

It appears to me (and others), ironically, that many CoPs within the open and reform constellation are lacking in self-reflection and humility. New initiatives and practices are being produced and adopted at a fast rate, and they generate clear and direct benefits for both science and scientists. At the same time, the scientific reform movement is still in its infancy and the long-term effects of many interventions have not been evaluated. The situation is made even more difficult by the fact that academic science is a complex system, and the downstream consequences of modifying complex systems are notoriously difficult to anticipate.

Many have expressed concerns about these issues. Devezer, Navarro, Vandekerckhove and Buzbas (2020), for instance, argue that reform policies have "little evidentiary backing" and are "based on methods which are suggested with no framework for assessing their validity or evaluating their efficacy" (p. 1). Ioannidis (2014) and Tiokhin and colleagues (2021) have shared similar considerations, particularly regarding the potential for unintended consequences of otherwise well-intentioned incentives and changes.

9.3 Toward a Richer Reform Identity

As I have already mentioned, different CoPs in the constellation have different philosophies, priorities and needs, and specialize in different academic competencies. This means that their contributions to the movement will vary greatly from one another, and that work has to be done for the movement's collective identity to reflect these diverse contributions.

Discussion

With all of this said, there is cause for hope. The reform movement is yet in its infancy. As more people adopt the reform mindset and practices, and more inter-CoP collaborations are born, it is likely that gaps between CoPs will become narrower. One key practice which will be invaluable to the process of advancing this positive change, and developing greater cohesion within the constellation is reflexivity.

As Finlay (2002) puts it, reflexivity is a process in which the researcher explicitly, continually engages in an analysis of personal influences on the research process. As myself and Derksen have explained in our 2021 article, reflexivity assists the researcher to question and adapt their interpretations of the research material, based on issues that arise throughout the study. In our article, we consider the idea of establishing a "new culture of reflexivity" in science that can play a role in increasing the quality of research findings, which complements an existing and growing interest in transparency. At present, researchers in empirical fields tend not to include reflexivity explicitly in their research practices, or include reflections and positionality statements in their written reports. Were they to do so, and those reflections were to be read by peers, they could be used as information sources. Replicating authors, for instance, could use the information to guide their replication studies.

Reflexivity can also be valuable as a means of remaining humble. Although many reform initiatives confer great benefits to science, they are still borne of people with their own agendas and flaws, and who are still embedded in an academic system that the joint reform enterprise has explicitly labeled as systemically troubled. Interrogating the movement's reform practices is part of remaining humble and self-reflective, and will no doubt serve the progress of the momentum well.

A self-reflective and humble reform movement will be able to recognize and remedy the issue of overgeneralizing standards, practices and principles. Practical reforms and support for them must be field- or CoP-specific. Reformers must encourage and foster richer, qualitative assessment and evaluation, with value-laden criteria alongside existing quantitative approaches. Long-term strategies for reform, including infrastructures must be incentivized and made accessible to as wide a user-base as possible, involving stakeholders from diverse disciplines and traditions.

Conclusion

Science has come a long way since the shocking details of Diederik Stapel's misconduct were first laid bare for all to see. In 11 short years, we have gone from simple, clumsy preregistration templates, controversial special issues such as 'Replicability in Psychological Science: A Crisis of Confidence?' in *Perspectives in Psychological Science* (Issue 6, November, 2012) and novel and exciting calls for

replication studies (such as the *Journal of Experimental Psychology: General*'s of 2014) to countless global grass-roots reproducibility networks, international conferences and sophisticated platforms and software which present a real challenge to existing academic structures and traditions. Although the reform movement has had its dark, unpleasant periods, it remains remarkable in the momentum it has maintained for more than a decade. I am optimistic that change in science is yet to come from the activism of reformers, metascientists and open science advocates, that will enact massive and persistent shifts in philosophy, practice and perceptions. I have hope in this movement – in the kind of change it has the potential to make in science – and will be right there alongside my peers, working to establish the change I want to see in science.

References

Abbasi, A., Hossain, L., & Leydesdorff, L. (2012). Betweenness centrality as a driver of preferential attachment in the evolution of research collaboration networks. *Journal of Informetrics*, *6*(3), 403–412.

Abraham, A., & Hassanien, A.-E. (2012). *Computational social networks: Tools, perspectives and applications*. Springer Science & Business Media.

Achenbach, J. (2015, March 1). Why do many reasonable people doubt science? *National Geographic*.

Agar, M. H. (1982). Toward an ethnographic language. *American Anthropologist*, *84*(4), 779–795.

Akar, E., & Mardikyan, S. (2018). User roles and contribution patterns in online communities: A managerial perspective. *Sage Open*, *8*(3), 1–19.

Allen, C., & Mehler, D. M. (2019). Open science challenges, benefits and tips in early career and beyond. *PLoS Biology*, *17*(5), e3000246.

Alvarez, R. M. (2014, September 28). *The pros and cons of research preregistration.* Retrieved from `http://blog.oup.com/2014/09/pro-con-research-preregistration/`

Antusch, S., Bradshaw, A., Carson, J., Constantino, S., Horder, J., Kousta, S., ... Sutherland, M. E. (2022). Five years of nature human behaviour. *Nature Human Behaviour*, *6*(1), 11–14.

Armeni, K., Brinkman, L., Carlsson, R., Eerland, A., Fijten, R., Fondberg, R., ... Zurita-Milla, R. (2021). Towards wide-scale adoption of open science practices: The role of open science communities. *Science and Public Policy*, *48*(5), 605–611.

Attride-Sterling, J. (2001). Thematic networks: An analytic tool for qualitative research. *Qualitative Research*, *1*(3), 385–405.

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, *533*(7604), 452–454.

Banks, G. C., Field, J. G., Oswald, F. L., O'Boyle, E. H., Landis, R. S., Rupp, D. E., & Rogelberg, S. G. (2019). Answers to 18 questions about open science practices. *Journal of Business and Psychology*, *34*(3), 257–270.

Barbour, V. (2015). Publish or perish culture encourages scientists to cut corners. *The Conversation*.

Barrick, M. R., Stewart, G. L., Neubert, M. J., & Mount, M. K. (1998). Relating member ability and personality to work-team processes and team effectiveness. *Journal of Applied Psychology*, *83*(3), 377–391.

Bartlett, T. (2018, September 11). 'I want to burn things to the ground'. *The Chronicle of Higher Education*.

Bastian, H. (2016, December 5). *Reproducibility crisis timeline: Milestones in tackling research reliability.* Plos Blogs.

Bastian, H. (2017, August 29). *Bias in open science advocacy: The case of article badges for data sharing.* Plos Blogs.

Bauernfeind, R. H. (1968). The need for replication in educational research. *The Phi Delta Kappan*, *50*(2), 126–128.

Beaulieu, A. (2004). Mediating ethnography: Objectivity and the making of ethnographies of the internet. *Social Epistemology*, *18*(2-3), 139–163.

Beaulieu, A. (2010). Research note: From co-location to co-presence: Shifts in the use of ethnography for the study of knowledge. *Social Studies of Science*, *40*(3), 453–470.

Begley, S. (2012, March 28). *In cancer science, many "discoveries" don't hold up.*

Berger, J. O., & Wolpert, R. L. (1988). *The likelihood principle*. Institute of Mathematical Statistics.

Bergmann, C. [@chbergma]. (2019, March 9). *So glad the buffet metaphor is catching on, there are so many solutions out there. Don't try to stuff yourself on everything, select what works for this study and let's steadily improve our fields... #openscience.* [Tweet]. Twitter. Retrieved from `https://twitter.com/chbergma/status/1104338904646385665`

Bergmann, C. [@chbergma]. (2021, November 18). *Seems a good moment to remind people of the Buffet approach to open science. Not everything works for everyone, and it is ok to pick practices.* [Tweet]. Twitter. Retrieved from `https://twitter.com/chbergma/status/1461199447531085829`

Björk, B.-C. (2017). Gold, green, and black open access. *Learned Publishing*, *30*(2), 173–175.

Bliemel, M. J., McCarthy, I. P., & Maine, E. M. (2014). An integrated approach to studying multiplexity in entrepreneurial networks. *Entrepreneurship Research Journal*, *4*(4), 367–402.

Blommaert, J. [@drjulie_b]. (2017, June 4). *Wow Open Science^{TM} seems like such an accepting and kind group of people #sarcasm.* [Tweet]. Twitter. Retrieved from `https://twitter.com/drjulie_b/status/871268434750959617`

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(10), P10008.

Bochner, A., & Ellis, C. (2016). *Evocative autoethnography: Writing lives and telling stories*. Routledge.

Bolderston, A. (2012). Conducting a research interview. *Journal of Medical Imaging and Radiation Sciences*, *43*(1), 66–76.

Borgatti, S. P., Everett, M. G., & Johnson, J. C. (2018). *Analyzing social networks*. SAGE Publications Sage CA: Los Angeles, CA.

Bourdieu, P. (1986). The forms of capital. handbook of theory and research for the sociology of education. In J. G. Richardson & J. S. Richardson (Eds.), *Handbook of theory and research for the sociology of education*. Greenwood Press, New York.

Bourdieu, P. (1990). *The logic of practice*. Stanford University Press.

Bourdieu, P. (1996). Understanding. *Theory, Culture & Society*, *13*(2), 17–37.

Bourdieu, P. (2013). *Distinction: A social critique of the judgement of taste*. Routledge.

Brownlie, D., Hewer, P., & Tadajewski, M. (2009). Thinking 'communities of academic practice': On space, enterprise and governance in marketing academia. *Journal of Marketing Management*, *25*(7–8), 635–642.

Bruckman, A. (2005). Learning in online communities. In *The Cambridge Handbook of the Learning Sciences*. Cambridge University Press.

Burgelman, J.-C., Pascu, C., Szkuta, K., Von Schomberg, R., Karalopoulos, A., Repanas, K., & Schouppe, M. (2019). Open science, open data, and open scholarship: European policies to make science fit for the twenty-first century. *Frontiers in Big Data*, *2*, 1–6.

Burt, R. S. (2002). The social capital of structural holes. *The New Economic Sociology: Developments in an Emerging Field*, *148*(90), 122.

Butler, D. (2017). Gates foundation announces open-access publishing venture. *Nature News*, *543*(7647), 599.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Rand McNally, Chicago, USA.

Carney, D. R., Cuddy, A. J., & Yap, A. J. (2010). Power posing: Brief nonverbal displays affect neuroendocrine levels and risk tolerance. *Psychological Science*, *21*(10), 1363–1368.

Carp, J. (2013). Better living through transparency: Improving the reproducibility of fmri results through comprehensive methods reporting. *Cognitive, Affective, & Behavioral Neuroscience*, *13*(3), 660–666.

Cesario, J. (2014). Priming, replication, and the hardest science. *Perspectives on Psychological Science*, *9*(1), 40–48.

Chambers, C. D. (2013). Registered Reports: A new publishing initiative at Cortex. *Cortex*, *49*(3), 609–610.

Chambers, C. D. (2014, May 20). Psychology's 'registration revolution'. *The Guardian*. Retrieved from `https://www.theguardian.com/science/head-quarters/2014/may/20/psychology-registration-revolution`

Chambers, C. D. (2019). *The seven deadly sins of psychology*. Princeton University Press.

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

Clancy, M. (2013). Is reflexivity the key to minimising problems of interpretation in phenomenological research? *Nurse Researcher*, *20*(6), 12–16.

Clarke, V., & Braun, V. (2019). Feminist qualitative methods and methodologies in psychology: A review and reflection. *Psychology of Women and Equalities Section Review*, *2*(1), 13–28.

Clear, J. (2013, October 17). How to be confident and reduce stress in 2 minutes per day. *James Clear*.

Clemens, M. A. (2017). The meaning of failed replications: A review and proposal. *Journal of Economic Surveys*, *31*(1), 326–342.

Clifton, R. A. (1999). The education of university students: A social capital perspective. *College Teaching*, *47*(3), 114–118.

Connolly, P., & Healy, J. (2004). Symbolic violence, locality and social class: The educational and career aspirations of 10-11-year-old boys in belfast. *Pedagogy, Culture and Society*, *12*(1), 15–33.

Corker, K. [@katiecorker]. (2019, September 7). *Alternative explanation: not just self serving bias but identity politics. People don't want to be seen as a "replicators." I see the same pattern with other practices. A: are you an open science person? B: no way! A: do you share your data, etc? B: of course!* [Tweet]. Twitter. Retrieved from `https://twitter.com/katiecorker/status/1170373620649971712`

Crang, M., & Cook, I. (2007). *Doing ethnographies*. Sage.

Crüwell, S. [@cruwelli]. (2018, October 21). *We should call people out, but individually (as has been done before). Blanket statements dismissing the entire movement and its practices are just as unhelpful as the bros.* [Tweet]. Twitter. Retrieved from `https://twitter.com/cruwelli/status/1053781162450190336`

Csiszar, A. (2016). Peer review: Troubled from the start. *Nature*, *532*(7599), 306–308.

Cuddy, A. (2012a, October 28). Act powerful, be powerful. *CNN*.

Cuddy, A. (2012b, June). *Your body language may shape who you are [video file]*. Retrieved from: `https://www.ted.com/talks/amy_cuddy_your_body_language_may_shape_who_you_are`.

Danvers, A. [@Alex_Danvers]. (2021, March 21). *Today I changed my profile to indicate I'm interested in science reform, not open science. It's been hard for me to come to grips with the idea that the community I originally found under the open science label (in 2012) is different from the way the label is being used now.* [Tweet]. Twitter. Retrieved from `https://twitter.com/Alex_Danvers/status/1373736173152079872`

Darcy, O. (2019, March 22). How Twitter's algorithm is amplifying extreme political rhetoric. *CNN Business*.

DeBruine, L. [@LisaDeBruine]. (2018, September 23). *Tonight I'm feeling weirdly excluded from the open science community, not by people in it, but by people outside insisting it's all "bros", which is so far from my experience that I have to assume they must be excluding me (and tons of awesome women and men) from the community.* [Tweet]. Twitter. Retrieved from `https://twitter.com/LisaDeBruine/status/1043964752534425602`

Derksen, M. (2019). Open psychology: Transparency and reproducibility. *Psicología, Conocimiento y Sociedad, 9 (2), 318-330*.

Derksen, M., & Field, S. (2021). The tone debate: Knowledge, self, and social order. *Review of General Psychology*, 1–12.

de Ruiter, J. P. (2018). The meaning of a claim is its reproducibility. *Behavioral and Brain Sciences*, *41*.

Devezer, B., Navarro, D. J., Vandekerckhove, J., & Buzbas, E. O. (2020). The case for formal methodology in scientific reform. *Royal Society Open Science*, *8*(3), 200805.

Devine, S., Bautista-Perpinya, M., Delrue, V., Gaillard, S., Jorna, T., van der Meer, M., ... Visser, J. (2020). Science fails. Let's publish. *Journal of Trial and Error*, *1*(1), 1–5.

Diener, E., & Biswas-Diener, R. (2018). The replication crisis in psychology. In *KU PSYC2020* (p. 1).

Dominus, S. (2017, October 18). *When the revolution came for amy cuddy* (Vol. 29).

Dow, B. J., & Condit, C. M. (2005). The state of the art in feminist scholarship in communication. *Journal of Communication*, *55*(3), 448–478.

Dunlap, K. (1925). The experimental methods of psychology. *The Journal of Genetic Psychology*, *32*, 502–522.

Easterbrook, S. (2009, December 8). *Software quality in climate research*.

Ebersole, C. R., Axt, J. R., & Nosek, B. A. (2016). Scientists' reputations are based on getting it right, not being right. *PLoS Biology*, *14*(5), e1002460.

Eckert, P. (2006). Communities of practice. In R. E. Asher & J. M. Y. Simpson (Eds.), *Encyclopedia of language and linguistics*. Elsevier Science & Technology.

Ellis, C., Adams, T. E., & Bochner, A. P. (2011). Autoethnography: An overview. *Historical Social Research*, *36*(4), 273–290.

Elms, A. C. (1975). The crisis of confidence in social psychology. *American Psychologist*, *30*(10), 967–976.

Engber, D. (2017, June 7). Daryl bem proved esp is real: Which means science is broken. *Slate*.

Enserink, M. (2012, November 28). Final report: Stapel affair points to bigger problems in social psychology. *ScienceInsider*.

Erlandson, D. A., Harris, E. L., Skipper, B. L., & Allen, S. D. (1993). *Doing naturalistic inquiry: A guide to methods*. Sage.

Fane, B., Ayris, P., Hahnel, M., Hrynaszkiewicz, I., Baynes, G., & Farrell, E. (2019). *The state of open data report 2019*. Digital Science. Retrieved from `https://doi.org/10.6084/m9.figshare.9980783.v2`

Fanelli, D. (2009). How many scientists fabricate and falsify research? a systematic review and meta-analysis of survey data. *PloS one*, *4*(5), e5738.

Fanelli, D. (2018). Opinion: Is science really facing a reproducibility crisis, and do we need it to? *Proceedings of the National Academy of Sciences*, *115*(11), 2628–2631.

Faye, C. (2012). American social psychology: Examining the contours of the 1970s crisis. *Studies in History and Philosophy of Science Part C: Studies in*

History and Philosophy of Biological and Biomedical Sciences, *43*(2), 514–521.

Fecher, B., & Friesike, S. (2014). Open science: One term, five schools of thought. In S. Bartling & S. Friesike (Eds.), *Opening science: The evolving guide on how the internet is changing research, collaboration and scholarly publishing.* Springer Nature.

Feld, S. L., & Carter, W. C. (2002). Detecting measurement bias in respondent reports of personal networks. *Social Networks*, *24*(4), 365–383.

Felps, W., Mitchell, T. R., & Byington, E. (2006). How, when, and why bad apples spoil the barrel: Negative group members and dysfunctional groups. *Research in Organizational Behavior*, *27*, 175–222.

Ferguson, C. J., & Heene, M. (2012). A vast graveyard of undead theories: Publication bias and psychological science's aversion to the null. *Perspectives on Psychological Science*, *7*(6), 555–561.

Fetterman, D. M. (2019). *Ethnography: Step-by-step*. SAGE Publications Sage CA: Los Angeles, CA.

Field, S. M. (2020, April 9). On the beauty of publishing an ugly registered report. *Bayesian Spectacles*. Retrieved from: https://www.bayesianspectacles.org/on-the-beauty-of-publishing-an-ugly-registered-report/.

Field, S. M., & Derksen, M. (2021). Experimenter as automaton; experimenter as human: Exploring the position of the researcher in scientific research. *European Journal for Philosophy of Science*, *11*(1), 1–21.

Field, S. M., Hoekstra, R., Bringmann, L., & van Ravenzwaaij, D. (2019). When and why to replicate: As easy as 1, 2, 3? *Collabra: Psychology*, *5*, 1–15.

Field, S. M., Wagenmakers, E.-J., Kiers, H. A., Hoekstra, R., Ernst, A. F., & van Ravenzwaaij, D. (2020). The effect of preregistration on trust in empirical research findings: Results of a registered report. *Royal Society Open Science*, *7*, 181351.

Field, S. M., Wagenmakers, E.-J., Newell, B. R., Zeelenberg, R., & van Ravenzwaaij, D. (2016). Two Bayesian tests of the GLOMO^{sys} model. *Journal of Experimental Psychology: General*, *145*, e81.

Fiesler, C., & Proferes, N. (2018). "Participant" perceptions of Twitter research ethics. *Social Media + Society*, *4*(1), 2056305118763366.

Finlay, L. (2002). "Outing" the researcher: The provenance, process, and practice of reflexivity. *Qualitative Health Research*, *12*(4), 531–545.

Fox, S. (2000). Communities of practice, foucault and actor-network therory. *Journal of Management Studies*, *37*(6), 853–868.

Fraser, H., Parker, T., Nakagawa, S., Barnett, A., & Fidler, F. (2018). Questionable research practices in ecology and evolution. *PLoS one*, *13*(7), 1-16.

Freeman, L. (2004). The development of social network analysis. *A Study in the Sociology of Science*, *1*(687), 159–167.

Fried, E. (2018, June 28). *Sips18 collected resources, and reflections of a sips virgin*. Retrieved from: https://eiko-fried.com/sips2018-collected-resources/.

Gagliardi, D., Cox, D., & Li, Y. (2015). Institutional inertia and barriers to the adoption of open science. In E. Primeri & E. Reale (Eds.), *The Transformation of University Institutional and Organizational Boundaries* (pp. 107–133). Springer Nature.

Garrett, B. (2022). *Autopsy of a crime lab: Exposing the flaws in forensics*. University of California Press. Retrieved from https://books.google.nl/books?id=99OnzgEACAAJ

Gaskó, N., Bota, F., Suciu, M. A., & Lung, R. I. (2020). A game theoretical analysis of academic writing co-authorship networks. *Journal of Scientometric Research*, *9*(3), 319–325.

Gee, J. P. (2005). Semiotic social spaces and affinity spaces. In D. Barton & K. Tusting (Eds.), *Beyond communities of practice; language power and social context*.

Geertz, C. (1973). *The interpretation of cultures*. Basic Books.

Gelman, A. (2016, September 21). What has happened down here is the winds have changed. *Statistical Modeling, Causal Inference, and Social Science*.

Gergen, K. J. (1973). Social psychology as history. *Journal of Personality and Social Psychology*, *26*(2), 309–320.

Gergen, K. J. (1985). The social constructionist movement in modern psychology. *American Psychologist*, *40*(3), 266–275.

Gernsbacher, M. A. (2018). Writing empirical articles: Transparency, reproducibility, clarity, and memorability. *Advances in Methods and Practices in Psychological Science*, *1*(3), 403–414.

Gervais, W. M. (2021). Practical methodological reform needs good theory. *Perspectives on Psychological Science*, *16*(4), 827–843.

Gherardi, S., & Nicolini, D. (2002). Learning in a constellation of interconnected practices: Canon or dissonance? *Journal of Management Studies*, *39*(4), 419–436.

Gigerenzer, G. (2018). Statistical rituals: The replication delusion and how we got there. *Advances in Methods and Practices in Psychological Science*, *1*(2), 198–218.

Gleave, E., Welser, H. T., Lento, T. M., & Smith, M. A. (2009). A conceptual and operational definition of 'social role' in online community. In *2009 42nd hawaii international conference on system sciences* (pp. 1–11).

Godlee, F., Gale, C. R., & Martyn, C. N. (1998). Effect on the quality of peer review of blinding reviewers and asking them to sign their reports: A randomized controlled trial. *Journal of the American Medical Association*, *280*(3), 237–240.

Goffman, E. (1978). *The presentation of self in everyday life*. Harmondsworth, London.

Golbeck, J. (2013). *Analyzing the social web*. Newnes.

Gold, R. L. (1958). Roles in sociological field observations. In *Social Forces* (Vol. 36, pp. 217–223). Routledge.

Goodman, S., & Greenland, S. (2007). Why most published research findings are false: Problems in the analysis. *PLoS Medicine*, *4*(4), e168.

Grandjean, M. (2016). A social network analysis of Twitter: Mapping the digital humanities community. *Cogent Arts & Humanities*, *3*(1), 1171458.

Grant, W. J., Moon, B., & Busby Grant, J. (2010). Digital dialogue? australian politicians' use of the social network tool Twitter. *Australian Journal of Political Science*, *45*(4), 579–604.

Grbich, C. (2004). *New approaches in social research*. Sage.

Gross, M. T. (1997). The need for replication studies – is it really a done deal? *Journal of Orthopaedic & Sports Physical Therapy*, *25*(3), 161–162.

Grover, P., & Kar, A. K. (2020). "# Government" – Understanding Dissemination, Transparency, Participation, Collaboration and Engagement on Twitter for Citizens. In *International working conference on transfer and diffusion of it* (pp. 365–376).

Hammersley, M., & Atkinson, P. (1995). Recording and organizing data. In *Ethnography: Principles in practice* (pp. 175–204). Psychology Press.

Hastrup, K. (1992). Writing ethnography. In J. Okely & H. Callaway (Eds.), *Anthropology and autobiography* (Vol. 29). Routledge.

Hermann, T., Nehls, A. V., Eitel, F., Barri, T., & Gammel, M. (2012). Tweetscapes-real-time sonification of Twitter data streams for radio broadcasting. In *Proceedings of the international conference on auditory displays.*

Hertz, R. (1997). *Reflexivity & voice*. Sage Publications.

Hilgard, J., & Jamieson, K. H. (2017). Science as "broken" versus science as "self-correcting". In K. H. Jamieson, D. M. Kahan, & D. A. Scheufele (Eds.), *The Oxford Handbook of the Science of Science Communication.* Oxford University Press.

Hillery, G. A. (1955). Definition of community. *Rural Sociology*, *20*, 111–123.

Hogg, M. A., & Williams, K. D. (2000). From i to we: Social identity and the collective self. *Group dynamics: Theory, Research, and Practice*, *4*(1), 81–97.

Holcombe, A. [@ceptional]. (2018, October 21). *No, I wasnt' saying that. I do think it should be called out. Just that occasional twitter users get the wrong impression of bullying being more representative, due to the algorithms.* [Tweet]. Twitter. Retrieved from `https://twitter.com/ceptional/status/1053785225002594304`

Hong, J. F., & O, F. K. (2009). Conflicting identities and power between communities of practice: The case of it outsourcing. *Management Learning*, *40*(3), 311–326.

Hornsey, M. J. (2008). Social identity theory and self-categorization theory: A historical review. *Social and Personality Psychology Compass*, *2*(1), 204–222.

Horton, R. (2015). Offline: What is medicine's 5 sigma? *The Lancet*, *385*(9976), 1380.

Howard, A. [@DrAndreaHoward]. (2021, March 20). *In a setting where to many, open science already feels closed off to only a select group of Super Science insiders, the last thing we need is another outlet for self-congratulatory posturing. Just... do the science in as open a way as you can, when you can.* [Tweet]. Twitter. Retrieved from `https://twitter.com/DrAndreaHoward/status/1373257197812133889`

Huston, P., Edge, V., & Bernier, E. (2019). Open science/open data: Reaping the benefits of open data in public health. *Canada Communicable Disease Report*, *45*(11), 252–256.

Huszár, F., Ktena, S. I., O'Brien, C., Belli, L., Schlaikjer, A., & Hardt, M. (2022). Algorithmic amplification of politics on Twitter. *Proceedings of the National Academy of Sciences*, *119*(1), e2119924119.

Ioannidis, J. P. (2014). How to make more published research true. *PLoS Medicine*, *11*(10), e1001747.

Ioannidis, J. P., Allison, D. B., Ball, C. A., Coulibaly, I., Cui, X., Culhane, A. C., ... van Noort, V. (2009). Repeatability of published microarray gene expression analyses. *Nature Genetics*, *41*(2), 149–155.

Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). Forceatlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PloS one*, *9*(6), e98679.

Jakobson, R., & Sebeok, T. A. (1960). Closing statement: Linguistics and poetics. In R. E. Innis (Ed.), *Semiotics: An introductory anthology* (pp. 147–175). Indiana University Press Bloomington, IN.

Jarrett, C. (2016, September 16). Ten famous psychology findings that it's been difficult to replicate. *Research Digest*.

Java, A., Song, X., Finin, T., & Tseng, B. (2007). Why we Twitter: Understanding microblogging usage and communities. In *Proceedings of the 9th webkdd and 1st sna-kdd 2007 workshop on web mining and social network analysis* (pp. 56–65).

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, *23*(5), 524-532.

Johnson, P. R., & Yang, S. (2009). Uses and gratifications of Twitter: An examination of user motives and satisfaction of Twitter use. In *Communication Technology Division of the annual convention of the Association for Education in Journalism and Mass Communication in Boston, MA* (Vol. 54).

Kahneman, D. (2012, September 26). *A proposal to deal with questions about priming effects.*

Kahneman, D. (2014). A new etiquette for replication. *Social Psychology*, *45*(4), 310–311.

Kathawalla, U.-K., Silverstein, P., & Syed, M. (2021). Easing into open science: A

guide for graduate students and their advisors. *Collabra: Psychology*, *7*(1), 18684.

Kawulich, B. B. (2005). Participant observation as a data collection method. In *Forum qualitative sozialforschung/forum: Qualitative social research* (Vol. 6).

Kerr, N. L. (1998). Harking: Hypothesizing after the results are known. *Personality and Social Psychology Review*, *2*(3), 196–217.

Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L.-S., ... Nosek, B. A. (2016). Badges to acknowledge open practices: A simple, low-cost, effective method for increasing transparency. *PLoS Biology*, *14*(5), e1002456.

Kislov, R. (2014). Boundary discontinuity in a constellation of interconnected practices. *Public Administration*, *92*(2), 307–323.

Kjerholt, M., Wagner, L., Delmar, C., Clemensen, J., & Lindhardt, T. (2014). Continuity in care trajectories of older chronically ill patients in a battlefield of competing rationales. *International Journal of Older People Nursing*, *9*(4), 277–288.

Klein, R., Ratliff, K., Vianello, M., Adams Jr, R., Bahník, S., Bernstein, M., ... Nosek, B. A. (2014). Data from investigating variation in replicability: A "many labs" replication project. *Journal of Open Psychology Data*, *2*(1), e4.

Kleinberg, J. (2000). The small-world phenomenon: An algorithmic perspective. In *Proceedings of the thirty-second annual acm symposium on theory of computing* (pp. 163–170).

Koehler, J. J. (2010). Forensic science reform in the 21st century: A major conference, a blockbuster report and reasons to be pessimistic. *Law, Probability & Risk*, *9*(1), 1–6.

Kohana, M., Okamoto, S., & Kaneko, M. (2013). A clustering algorithm using Twitter user biography. In *2013 16th International Conference on Network-Based Information Systems* (pp. 432–435).

Kuhn, T. S. (1962). *The structure of scientific revolutions*. University of Chicago Press.

Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a social network or a news media? In *Proceedings of the 19th international conference on world wide web* (pp. 591–600).

Larivière, V., & Sugimoto, C. R. (2018). Do authors comply when funders enforce open access to research? *Nature*, *562*, 483–486.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge University Press.

LeCompte, M. D., & Goetz, J. P. (1982). Problems of reliability and validity in ethnographic research. *Review of Educational Research*, *52*(1), 31–60.

Levelt Committee, Noort Committee, & Drenth Committee. (2012). *Flawed science: The fraudulent research practices of social psychologist Diederik Stapel* (Tech. Rep.). Retrieved from `https://www.rug.nl/about-ug/`

latest-news/news/archief2012/nieuwsberichten/stapel-eindrapport-eng.pdf

Levin, N., & Leonelli, S. (2017). How does one "open" science? questions of value in biological research. *Science, Technology, & Human Values*, *42*(2), 280–305.

Levin, N., Leonelli, S., Weckowska, D., Castle, D., & Dupré, J. (2016). How do scientists define openness? Exploring the relationship between open science policies and research practice. *Bulletin of Science, Technology & Society*, *36*(2), 128–141.

Li, L. C., Grimshaw, J. M., Nielsen, C., Judd, M., Coyte, P. C., & Graham, I. D. (2009). Evolution of wenger's concept of community of practice. *Implementation Science*, *4*(1), 1–8.

Liamputtong, P. (2007). *Researching the vulnerable: A guide to sensitive research methods*. Sage.

Liamputtong, P. (2019). Qualitative inquiry. In P. Liamputtong (Ed.), *Handbook of research methods in health social sciences* (pp. 9–25). Springer Singapore.

Lincoln, Y. S., & Denzin, N. K. (2003). *Turning points in qualitative research: Tying knots in a handkerchief*. Rowman Altamira.

Lovejoy, K., & Saxton, G. D. (2012). Information, community, and action: How nonprofit organizations use social media. *Journal of Computer-Mediated Communication*, *17*(3), 337–353.

Lowe, D. (2016, January 22). Attack of the research parasites. *Science Commentary*.

Macià, M., & Garcia, I. (2017). Properties of teacher networks in Twitter: Are they related to community-based peer production? *International Review of Research in Open and Distributed Learning*, *18*(1), 110–140.

Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science*, *7*(6), 537–542.

Malinowski, B. (2020). *A diary in the strict sense of the term*. Routledge.

Mann, B., & Keller, J. (2013). Why a feminist volume on pluralism? *Philosophical Topics*, *41*(2), 1–11.

Marcus, A., & Oransky, I. (2018). The data thugs. *Science*, *359*(6377), 730-732.

Marsden, P. V. (1993). The reliability of network density and composition measures. *Social Networks*, *15*(4), 399–421.

Masuzzo, Paola C. [@pcmasuzzo]. (2021, August 2). *"Open science is not just about improving the way we share data and methods; it is also about improving the way we think, work and interact with each other. It's about technology enabling social infrastructure that can promote inclusivity to create kinder science."* >. [Tweet]. Twitter. Retrieved from https://twitter.com/pcmasuzzo/status/1422097805003345920

Maton, K. (2003). Reflexivity, relationism, & research: Pierre bourdieu and the epistemic conditions of social scientific knowledge. *Space and Culture*, *6*(1),

52–65.

Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? what does "failure to replicate" really mean? *American Psychologist*, *70*(6), 487–498.

Maykut, P., & Morehouse, R. (2002). *Beginning qualitative research: A philosophical and practical guide*. Routledge.

Mayring, J. (2014). Qualitative content analysis. In *Forum qualitative sozialforschung/forum: Qualitative social research* (Vol. 2, pp. 1–18).

McCarthy, R. (2016, June 23). *The average Twitter user now has 707 followers.*

McRobbie, A. (2002). A mixed bag of misfortunes? bourdieu's weight of the world. *Theory, culture & society*, *19*(3), 129–138.

Mering, M., & Hoeve, C. D. (2020). A brief history to the future of open access. *Serials Review*, *46*(4), 300–304.

Merle, M., Reese, G., & Drews, S. (2019). #Globalcitizen: An explorative Twitter analysis of global identity and sustainability communication. *Sustainability*, *11*(12), 3472.

Merriam, S. B. (1988). *Case study research in education: A qualitative approach*. Jossey-Bass.

Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigations*. University of Chicago Press.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology: General*, *90*(2), 227–234.

Meyer, M., & Chabris, C. (2014, July 31). Why psychologists' food fight matters. *Slate*.

Mitchum, R. (2015, May 18). What academia can learn from open source. *Computation Institute*.

Monahan, T., & Fisher, J. A. (2010). Benefits of 'observer effects': Lessons from the field. *Qualitative Research*, *10*(3), 357–376.

Mons, B., Neylon, C., Velterop, J., Dumontier, M., da Silva Santos, L. O. B., & Wilkinson, M. D. (2017). Cloudy, increasingly fair; revisiting the fair data guiding principles for the european open science cloud. *Information Services & Use*, *37*(1), 49–56.

Moonesinghe, R., Khoury, M. J., & Janssens, A. C. J. W. (2007). Most published research findings are false – but a little replication goes a long way. *PLoS Medicine*, *4*(2), e28.

Moran-Ellis, J. (1996). Close to home: The experience of researching child sexual abuse. In M. Hester, L. Kelly, & J. Radford (Eds.), *Women, violence and male power: Feminist activism, research and practice.* Open University Press, Buckingham.

Munafò, M. R., Nosek, B. A., Bishop, D. V., Button, K. S., Chambers, C. D., Percie du Sert, N., … Ioannidis, J. P. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*(1), 1–9.

Musante-DeWalt, K., & DeWalt, B. R. (2010). *Participant observation: A guide for fieldworkers*. Rowman Altamira.

Myers, S. A., Sharma, A., Gupta, P., & Lin, J. (2014). Information network or social network? the structure of the Twitter follow graph. In *Proceedings of the 23rd international conference on world wide web* (pp. 493–498).

Naaman, M., Boase, J., & Lai, C.-H. (2010). Is it really about me? message content in social awareness streams. In *Proceedings of the 2010 acm conference on computer supported cooperative work* (pp. 189–192).

Nadai, E., & Maeder, C. (2005). Fuzzy fields. multi-sited ethnography in sociological research. *Forum Qualitative Research, On-line Journal*, *6*(3), 1–10.

Nau, D. (1995). Mixing methodologies: Can bimodal research be a viable post-positivist tool. *The Qualitative Report*, *2*(3), 1–5.

Newman, M. E., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E*, *69*(2), 026113.

Ngiam, W. [@will_ngiam]. (2021, March 19). *Advocating for transparent and open science involves empowering researchers to engage in reproducible practices and lowering barriers to do so, such as by creating accessible tools, collaboration and guides – not to bully and shame them into it.* [Tweet]. Twitter. Retrieved from `https://twitter.com/will_ngiam/status/1372710832740696064`

North, M. A., Hastie, W. W., & Hoyer, L. (2020). Out of Africa: The underrepresentation of African authors in high-impact geoscience literature. *Earth-Science Reviews*, *208*, 103262.

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S., Breckler, S., ... Christensen, G. (2015). Promoting an open research culture. *Science*, *348*(6242), 1422–1425.

Nosek, B. [@BrianNosek]. (2019, August 15). *The behavior is corrosive and counterproductive. It isn't obvious to me whether there is anything to do but ignore it – give it no extra oxygen.* [Tweet]. Twitter. Retrieved from `https://twitter.com/BrianNosek/status/1162117670780776449`

Nosek, B. [@BrianNosek]. (2021a, November 14). *Everyone owns the risks. We exacerbate those risks when we use group-based language (eg, open scientists, SIPS, social psychologists) in criticism of behaviors by individuals. Inevitably, the groups are more heterogeneous than the basis of the criticism, widening the conflict.* [Tweet]. Twitter. Retrieved from `https://twitter.com/BrianNosek/status/1459887751218749444`

Nosek, B. [@BrianNosek]. (2021b, March 2). *So, if we derive some identity from open science as a movement, where does it get attached? It can be orgs and others when available, but in the murky realm, when bad actions occur in the name of open science, then we are forced to ask ourselves "Am I part of that?!".* [Tweet]. Twitter. Retrieved from `https://twitter.com/BrianNosek/status/1374050375477641217`

Nosek, Brian. [@BrianNosek]. (2018, June 23). *If you care about research in-*

tegrity, open science, and reproducibility, then follow #SIPS2018 for the next few days. You won't regret it. And, next year, whatever your field, you might decide to go to the conference in Rotterdam in July: http://improvingpsych.org. [Tweet]. Twitter. Retrieved from https://twitter.com/BrianNosek/status/1010378383039983616

Oentaryo, R. J., Low, J.-W., & Lim, E.-P. (2015). Chalk and cheese in Twitter: Discriminating personal and organization accounts. In *European conference on information retrieval* (pp. 465–476).

Open Science Collaboration. (2012). An open, large-scale, collaborative effort to estimate the reproducibility of psychological science. *Perspectives on Psychological Science*, *7*(6), 657–660.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), 947–951.

Oransky, I. (2021, April 2). *25,000: That's how many retractions are now in the retraction watch database.* Retrieved from https://retractionwatch.com/2021/04/02/25000-thats-how-many-retractions-are-now-in-the-retraction-watch-database/

Pappas, S. (2016, October 14). The new science of willpower: Can self-control really get used up? *Live Science*.

Parand, F.-A., Rahimi, H., & Gorzin, M. (2016). Combining fuzzy logic and eigenvector centrality measure in social network analysis. *Physica A: Statistical Mechanics and its Applications*, *459*, 24–31.

Pashler, H., & Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science*, *7*(6), 531–536.

Pashler, H., & Wagenmakers, E.-J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science*, *7*(6), 528–530.

Perlis, M. M. (2012, November 4). Assume the position: Posing your way to personal power and excellence. *Forbes*.

Peters, Gjalt-Jorn. [@matherion]. (2020a, January 6). *It leaves more room for diversification, live and let live, etc, where being in any single community doesn't matter so much any more. People can just be as 'communal' as they feel comfortable with, and each community can include/exclude as they deem consistent with their values.* [Tweet]. Twitter. Retrieved from https://twitter.com/matherion/status/1214108524864823298

Peters, Gjalt-Jorn. [@matherion]. (2020b, January 6). *True. And I'm increasingly feeling that if a community is to some degree about values/ideals, there are inevitably people who disagree. We don't need one community with everybody; we can have lots of communities, too; and community-less/-fluid people, too.* [Tweet]. Twitter. Retrieved from https://twitter.com/matherion/status/1214108524864823298

Peterson, D., & Panofsky, A. (2020). *Metascience as a scientific social movement.* SocArXiv.

Peterson, D., & Panofsky, A. (2021). Self-correction in science: The diagnostic and integrative motives for replication. *Social Studies of Science*, *51*(4), 583–605.

Pickering, M. J., & Ferreira, V. S. (2008). Structural priming: A critical review. *Psychological Bulletin*, *134*(3), 427–459.

Pillow, W. (2003). Confession, catharsis, or cure? Rethinking the uses of reflexivity as methodological power in qualitative research. *International Journal of Qualitative Studies in Education*, *16*(2), 175–196.

Pink, S. (2016). Digital ethnography. In S. Kubitschko & A. Kaun (Eds.), *Innovative methods in media and communication research* (pp. 161–165). Palgrave Macmillan Cham.

Popper, K. (2005). *The logic of scientific discovery*. Routledge.

Prinz, F., Schlange, T., & Asadullah, K. (2011). Believe it or not: How much can we rely on published data on potential drug targets? *Nature Reviews Drug Discovery*, *10*(9), 712–712.

Proferes, N. (2017). Information flow solipsism in an exploratory study of beliefs about Twitter. *Social Media + Society*, *3*(1), 1–17.

Qu, S. Q., & Dumay, J. (2011). The qualitative research interview. *Qualitative Research in Accounting & Management*, *8*, 238–264.

Rai, T. [@tage_rai]. (2020, September 14). *I had to deal w/a lot of trolls today in my feed. But I feel worst for people who replied and then had to put up w/the random racist, sexist trolling that has become standard fare in the "open" science community. Just look at the progression of this shit. Ya'll got a problem.* [Tweet]. Twitter. Retrieved from `https://twitter.com/tage_rai/status/1305296983301775361`

Ramsey, D., Annis, B., & Everitt, J. (2002). Definitions and boundaries of community: The case of rural health focus group analysis in southwestern manitoba. *Prairie Perspectives: Geographical Essays*, *5*, 189–203.

Ranehill, E., Dreber, A., Johannesson, M., Leiberg, S., Sul, S., & Weber, R. A. (2015). Assessing the robustness of power posing: No effect on hormones and risk tolerance in a large sample of men and women. *Psychological Science*, *26*(5), 653–656.

Reavey, P. (2012). *Visual methods in psychology: Using and interpreting images in qualitative research*. Routledge.

Rehm, M., & Notten, A. (2016). Twitter as an informal learning space for teachers!? The role of social capital in Twitter conversations among teachers. *Teaching and Teacher Education*, *60*, 215–223.

Reinhardt, W., Ebner, M., Beham, G., & Costa, C. (2009). Creativity and innovation competencies on the web. In *Edumedia conference, Salzburg*.

Resnick, B. (2016, March 25). What psychology's crisis means for the future of science. *Vox*.

Rheingold, H. (1993). A slice of life in my virtual community. In L. M. Harasim (Ed.), *Global networks: Computers and international communication* (pp.

57–80).

Rhodes, E. (2015, January). Speaking out on the 'replication crisis'. *The Psychologist*, *28*, 8–13.

Robson, S. G., Baum, M. A., Beaudry, J. L., Beitner, J., Brohmer, H., Chin, J. M., ... Thomas, A. (2021). Promoting open science: A holistic approach to changing behaviour. *Collabra: Psychology*, *7*(1), 30137.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641.

Ross-Hellauer, T. (2017). What is open peer review? a systematic review. *F1000Research*, *6*(588).

Ross-Hellauer, T., Deppe, A., & Schmidt, B. (2017). Survey on open peer review: Attitudes and experience amongst editors, authors and reviewers. *PloS one*, *12*(12), e0189311.

Rubin, M. (2017). When does harking hurt? identifying when different types of undisclosed post hoc hypothesizing harm scientific progress. *Review of General Psychology*, *21*(4), 308–320.

Rubin, M. (2020). The costs of HARKing. *The British Journal for the Philosophy of Science*.

Sandelowski, M. (2000). Whatever happened to qualitative description? *Research in Nursing & Health*, *23*(4), 334–340.

Sarno, D. (2009, February 18). *Twitter creator jack dorsey illuminates the site's founding document. part i.* Retrieved from: `https://latimesblogs.latimes.com/technology/2009/02/twitter-creator.html`.

Scheffer, T. (2007). Event and process: An exercise in analytical ethnography. *Human Studies*, *30*(3), 167–197.

Schensul, S. L., Schensul, J. J., & LeCompte, M. D. (1999). *Essential ethnographic methods: Observations, interviews, and questionnaires* (Vol. 2; J. J. Schensul & M. D. LeCompte, Eds.). Rowman Altamira.

Scherer, Laura. [@ldscherer]. (2021, November 18). *Just one final thought: In spite of this experience, open science and science reform are really important to me. And it won't stop being important to me because some people who have an "open science" identity have bad takes, or are even total assholes, on twitter.* [Tweet]. Twitter. Retrieved from `https://twitter.com/ldscherer/status/1461402895061069842`

Schmidt, S. (2009). Shall we really do it again? the powerful concept of replication is neglected in the social sciences. *Review of General Psychology*, *13*(2), 90–100.

Schneider, J., Rosman, T., Kelava, A., & Merk, S. (2021). (re) building trust? journals' open science badges influence trust in scientists. *PsyArXiv*.

Schuchard, R., Crooks, A. T., Stefanidis, A., & Croitoru, A. (2019). Bot stamina: Examining the influence and staying power of bots in online social networks. *Applied Network Science*, *4*(1), 1–23.

Schulz, H.-J., & Hurter, C. (2013). Grooming the hairball – how to tidy up network visualizations? In *Infovis 2013, ieee information visualization conference.*

Scott, S. (2013, July 25). *Pre-registration would put science in chains.*

Shapin, S., & Schaffer, S. (2011). *Leviathan and the air-pump: Hobbes, boyle, and the experimental life*. Princeton University Press.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366.

Simons, D. J. (2014). The value of direct replication. *Perspectives on Psychological Science*, *9*(1), 76–80.

Singleton, R., & Straits, B. C. (2010). *Approaches to social research*. Oxford University Press.

Smith, H. W. (1975). *Strategies of social research: the methodological imagination*. Prentice-Hall, NJ.

Spellman, B. A., Gilbert, E. A., & Corker, K. S. (2018). Open science. *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience*, *5*, 1–47.

Srivastava, P., & Hopwood, N. (2009). A practical iterative framework for qualitative data analysis. *International Journal of Qualitative Methods*, *8*(1), 76–84.

Star, S. L., & Griesemer, J. R. (1989). Institutional ecology, translations and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science*, *19*(3), 387–420.

Stark, B., Stegmann, D., Magin, M., & Jürgens, P. (2020). Are algorithms a threat to democracy? the rise of intermediaries: A challenge for public discourse. *Berlin: AlgorithmWatch*, *30*, 1–69.

Statista Research Department. (2022, April 12). *Twitter – Statistics & facts.* Retrieved from: `https://www.statista.com/topics/737/twitter`.

Steckler, T. (2015). Preclinical data reproducibility for R&D – the challenge for neuroscience. *SpringerPlus*, *4*(1), 1–4.

Steele, C., Butler, L., & Kingsley, D. (2006). The publishing imperative: The pervasive influence of publication metrics. *Learned Publishing*, *19*(4), 277–290.

Stephansen, H. C., & Couldry, N. (2014). Understanding micro-processes of community building and mutual learning on Twitter: A 'small data' approach. *Information, Communication & Society*, *17*(10), 1212–1227.

Stephens, M., & Poorthuis, A. (2015). Follow thy neighbor: Connecting the social and the spatial networks on Twitter. *Computers, Environment and Urban Systems*, *53*, 87–95.

Strick, B. (2020, May 27). How I scrape and analyse Twitter networks [case study]. [Web log post]. <*Ben*>. Retrieved from: `https://benjaminstrick.com/how-i-scrape-and-analyse-twitter-networks/`.

Stricker, J., & Günther, A. (2019). Scientific misconduct in psychology. *Zeitschrift für Psychologie*.

Strickland, L. H., Aboud, F., Gergen, K. J., Jahoda, G., & Tajfel, H. (1976). General theory in social psychology. *Personality and Social Psychology Bulletin*, *2*(2), 148–153.

Stroebe, W., Postmes, T., & Spears, R. (2012). Scientific misconduct and the myth of self-correction in science. *Perspectives on Psychological Science*, *7*(6), 670–688.

Stroebe, W., & Strack, F. (2014). The alleged crisis and the illusion of exact replication. *Perspectives on Psychological Science*, *9*(1), 59–71.

Takes, F. W., & Kosters, W. A. (2011). Determining the diameter of small world networks. In *Proceedings of the 20th acm international conference on information and knowledge management* (pp. 1191–1196).

Tennant, J. P., Dugan, J. M., Graziotin, D., Jacques, D. C., Waldner, F., Mietchen, D., ... Colomb, J. (2017). A multi-disciplinary perspective on emergent and future innovations in peer review. *F1000Research*, *6*(1151).

The Editors of Encyclopaedia Britannica. (2022, April 14). *Twitter*. Retrieved from: `https://www.britannica.com/topic/Twitter`.

Tinati, R., Halford, S., Carr, L., & Pope, C. (2014). Big data: Methodological challenges and approaches for sociological analysis. *Sociology*, *48*(4), 663–681.

Tiokhin, L., Panchanathan, K., Lakens, D., Vazire, S., Morgan, T., & Zollman, K. (2021). Honest signaling in academic publishing. *PloS one*, *16*(2), e0246675.

Totet, M. (2016, April 25). *Twitter streaming importer: Naoyun as a Gephi plugin*.

Traag, V. A., Waltman, L., & Van Eck, N. J. (2019). From louvain to leiden: Guaranteeing well-connected communities. *Scientific Reports*, *9*(1), 1–12.

United States National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1979, April 18). *The belmont report: Ethical principles and guidelines for the protection of human subjects of research*. Department of Health, Education, and Welfare.

Valente, T. W., Coronges, K., Lakon, C., & Costenbader, E. (2008). How correlated are network centrality measures? *Connections (Toronto, Ont.)*, *28*(1), 16–26.

Van Maanen, J. (2011). *Tales of the field: On writing ethnography*. University of Chicago Press.

Van Rooyen, S., Godlee, F., Evans, S., Black, N., & Smith, R. (1999). Effect of open peer review on quality of reviews and on reviewers' recommendations: A randomised trial. *British Medical Journal*, *318*(7175), 23–27.

Vasilevsky, N. A., Brush, M. H., Paddock, H., Ponting, L., Tripathy, S. J., LaRocca, G. M., & Haendel, M. A. (2013). On the reproducibility of science: Unique identification of research resources in the biomedical literature. *PeerJ*, *1*, e148.

Vazire, S. (2018). Implications of the credibility revolution for productivity, creativity, and progress. *Perspectives on Psychological Science*, *13*(4), 411–417.

Wacquant, L. J. (2002). Scrutinizing the street: Poverty, morality, and the pitfalls of urban ethnography. *American Journal of Sociology*, *107*(6), 1468–1532.

Wacquant, L. J., & Bourdieu, P. (1992). *An invitation to reflexive sociology*. Polity Cambridge.

Wagenmakers, E.-J. (2012). A year of horrors. *De Psychonoom*, *27*, 12–13.

Wagenmakers, E.-J., & Dutilh, G. (2016, October 31). Seven selfish reasons for preregistration. *APS Observer*, *29*.

Wagenmakers, E.-J., Gronau, Q. F., & Vandekerckhove, J. (2019, Mar). *Five bayesian intuitions for the stopping rule principle.* PsyArXiv. Retrieved from psyarxiv.com/5ntkd doi: 10.31234/osf.io/5ntkd

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H. L., & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science*, *7*(6), 632–638.

Walsh, E., Rooney, M., Appleby, L., & Wilkinson, G. (2000). Open peer review: A randomised controlled trial. *The British Journal of Psychiatry*, *176*(1), 47–51.

Walters, W. P. (2020). Code sharing in the open science era. *Journal of Chemical Information and Modeling*, *60*(10), 4417–4420.

Ware, M. (2008). Peer review in scholarly journals: Perspective of the scholarly community – results from an international study. *Information Services & Use*, *28*(2), 109–112.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge University Press.

Wellman, B. (1979). The community question: The intimate networks of East Yorkers. *American Journal of Sociology*, *84*(5), 1201–1231.

Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge university press.

Wenger, E. (2000). Communities of practice: The key to knowledge strategy. In E. L. Lesser, M. A. Fontaine, & J. A. Slusher (Eds.), *Knowledge and communities* (pp. 3–21). Routledge.

Wenger, E., McDermott, R. A., & Snyder, W. (2002). *Cultivating communities of practice: A guide to managing knowledge*. Harvard Business Press.

Wenger, E., Trayner, B., & De Laat, M. (2011). *Promoting and assessing value creation in communities and networks: A conceptual framework.* Ruud de Moor Centrum, the Netherlands.

Wenger, E. C., & Snyder, W. M. (2000). Communities of practice: The organizational frontier. *Harvard Business Review*, *78*(1), 139–146.

Whitaker, K., & Guest, O. (2020). #bropenscience is broken science: Kirstie whitaker and olivia guest ask how open 'open science' really is. *The Psychologist*, *33*, 34–37.

Williams, J., & Chowdhury, R. (2021, April 14). Introducing our responsible machine learning initiative. *Twitter Company Blog*.

Williams, M. L., Burnap, P., & Sloan, L. (2017). Towards an ethical framework for publishing Twitter data in social research: Taking into account users' views, online context and algorithmic estimation. *Sociology*, *51*(6), 1149–1168.

Wojcik, S., & Hughes, A. (2019, April 24). *Sizing up Twitter users.* Retrieved from: `https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/`.

Wolfram, D., Wang, P., Hembree, A., & Park, H. (2020). Open peer review: Promoting transparency in open science. *Scientometrics*, *125*(2), 1033–1051.

Yap, P.-T., Fan, Y., Chen, Y., Gilmore, J. H., Lin, W., & Shen, D. (2011). Development trends of white matter connectivity in the first years of life. *PloS one*, *6*(9), e24678.

Yong, E. (2012a). Bad copy. *Nature*, *485*(7398), 298-300.

Yong, E. (2012b, March 10). A failed replication draws a scathing personal attack from a psychology professor. *National Geographic*.

Yong, E. (2015, August 27). How reliable are psychology studies? *The Atlantic*.

Yoon, H. Y., & Park, H. W. (2014). Strategies affecting Twitter-based networking pattern of south korean politicians: Social network analysis and exponential random graph model. *Quality & Quantity*, *48*(1), 409–423.

Zappavigna, M. (2011). Ambient affiliation: A linguistic perspective on Twitter. *New Media & Society*, *13*(5), 788–806.

Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2018). Making replication mainstream. *Behavioral and Brain Sciences*, *41*, e120.

Samenvatting

Gebeurtenissen in de jaren 2011 en 2012 legden de basis van wat een sociale beweging in de wetenschap zou worden. Het begon met problemen rond het bestuderen en rapporteren van psychologische fenomenen, schokkende fraudezaken waarbij gegevens werden vervalst, en berichten over de verrassend hoge schattingen van de prevalentie van wetenschappelijk wangedrag. De bezorgdheid over de integriteit, de geldigheid en de betrouwbaarheid van het onderzoek breidde zich uit en trof een groot deel van de wetenschappelijke gemeenschap; een kettingreactie die in gang zette wat nu bekend staat als de wetenschappelijke hervormingsbeweging.

Deze beweging heeft tot doel de wetenschap te verbeteren. Zij streeft naar verbetering van wetenschappelijke methoden, statistische technieken, theorieontwikkeling, verslaglegging, het academische publicatiesysteem, en inclusiviteit in de wetenschappelijke gemeenschap. Om deze doelstellingen te bereiken, richten mensen en organisaties binnen de beweging zich op bewustmaking van de problemen, bieden zij onderwijs en steun bij het overnemen van hervormingsfilosofieën en -praktijken, en werken zij aan de totstandbrenging van nieuwe onderzoeksinfrastructuren ter ondersteuning van de grotere transparantie en diversiteit die de beweging is begonnen te bevorderen.

Hoewel de hervormingsbeweging en haar doelstellingen bewonderenswaardig en veelbelovend lijken, gaat het toch om een groep mensen, met alle onvermijdelijke intermenselijke uitdagingen van collectieve actie van dien. Pesterijen en wij-tegen-hen mentaliteiten, samen met de heterogeniteit die de hervormingsbeweging heeft geërfd van haar moedersysteem, de academische wereld, werken in een complexe samenloop. De spanningen en ontevredenheid die zij veroorzaken hebben geleid tot scheuren binnen het geheel, waardoor het collectief uiteen is gevallen en misschien zelfs is verwaterd.

In deze dissertatie doe ik verslag van het onderzoek dat ik bij deze groep mensen heb uitgevoerd. Vragen over de cultuur en de structuur van de gemeenschap waren mijn drijfveer. Ik wilde weten uit wat voor soort mensen de gemeen-

schap bestaat, en aan wat voor soort onderwerpen en praktijken zij prioriteit geven bij het nastreven van hun gemeenschappelijke doel. Hoe behandelen ze kwesties van lidmaatschap en identiteit in de context van hun gedeelde doelen, en hoe wordt omgegaan met problematische actoren? Hoe gebruikt de groep het platform Twitter om deel te nemen aan hun activiteiten als een gemeenschap? Wat is de algemene structuur van de gemeenschap? Is het een samenhangend geheel, of is het meer complex?

Ik gebruik een mix van methoden om deze vragen te onderzoeken, en mijn bevindingen te interpreteren. Etnografische methoden zorgden voor een rijkdom en complexiteit in mijn beschrijving van de hervormingsconstellatie, terwijl een sociaal netwerkperspectief me in staat stelde om vanuit een kwantitatieve invalshoek een idee te krijgen van de mogelijk modulaire structuur van de groep.

Een 'Community of Practice' (CoP) is een groep die een gemeenschappelijk doel of zorg deelt, en die zich als een gemeenschap verenigt om dat doel te bereiken. In deze dissertatie pas ik het CoP-concept van Etienne Wenger toe op de wetenschapshervormingsgroep. Ik beschouw hoe de groep onderhandelt over identiteit en betekenis, en hoe zij samen optrekken om hun gezamenlijke onderneming van wetenschapshervorming uit te oefenen.

Ik beschouw het verwante concept van een constellatie van CoP's, en ik beargumenteer dat hoewel de mensen die de wetenschapshervorming aansturen zinvol kunnen worden beschouwd als een CoP, het misschien nuttiger is om hen te beschouwen als een constellatie van praktijkgemeenschappen. Een 'constellatiestructuur', waarbij meerdere subgroepen bestaan binnen de bredere hervormingsgemeenschap, houdt rekening met de pluraliteit van de identiteit van de hervormingsgroep: Zij biedt ruimte voor de verschillen in instrumenten, prioriteiten, modi operandi, en regels en verwachtingen inzake betrokkenheid tussen de bestaande clusters, maar ook voor het duidelijke overkoepelende doel van wetenschappelijke verbetering, dat ongetwijfeld centraal staat in elke subgroep.

Acknowledgments

For my sources...

First, I want to thank the participants and sources that have worked with me to co-produce this research. In a way, I feel as though I am sharing a part of their – our – story as I share this dissertation with the world. This dissertation is a story of science and togetherness. It is a story of the similarities the reform constellation share, and the differences that make the collective so powerful and complex. Almost as much as it is a story of belonging and identity, it is a story of personal isolation and being on the periphery. It is a story about collective disillusionment with science. Not all of the material I collected made its way into this dissertation, but I value every one of the narratives you shared with me.

Being a story-teller is a sacred task and it comes with much responsibility. I have felt the weight of that throughout the process of conducting this research, but it is a role I chose, and one which I have been humbled to carry out. To the people who have been the basis of this research: Thank you for trusting me to tell some of your science story. I hope you find this research interesting and illuminating, and faithful to the different realities and truths represented by the reform movement.

For my family...

Thank you Don, for so much. For everything, really. For being the first to teach me about metascience, for helping me tune my academic moral compass, and for talking shop around the house. Thanks for being an amazing co-author. Thank you for demonstrating every single day that you value my time and work just as much as your own. Thank you for the countless cups of tea you brought to me in my home office on the long nights I spent alone in front of my screen, and for helping my carry my stress and frustration. Thank you to my babies, for teaching me that life is just as much about the little things as about the big things, and

for teaching me how multi-faceted and boundless love can be. Thank you for reminding me to value self-care and hot coffee.

Thank you mum, for being the first to teach me that I should be brave and go for what I want. Thank you for teaching me to love myself, not only other people. Thanks for the flair for words, and for sparking an early love for audiobooks. They have been an escape for me during the stress of pulling together this dissertation while co-parenting two tinies during a pandemic. Thanks dad, for teaching me my way around a toolbox, practically and metaphorically. That's come in handy for this dissertation more than you'd think! Thanks Steven for the weird books. Books about ancient Egyptian embalming practices, the black plague, the great fire of London, and Pompeii. Books by Marx, Descartes and Meister Eckhart. These books taught me about people; how they think and act in different contexts, and how to think about them. You always said I should be an anthropologist. You weren't too far off. I told you at Paul and Fee's wedding reception that I'd be the next PhD in the family. Well, I did it. Thanks Uncy Paul, for so much. If it wasn't for you forcing me to call uni admin and get myself re-enrolled in my undergrad, I would never have written this thesis. You and Fee were there that day to guide me to what would be one of the best choices I have ever made, and you were there through some of the lowest points in my life to help me through. Your unwavering love means more to me than you know.

Dankje Mam, Papje, Lydia, Broertje en Sanne. Jullie verwelkomde me met open armen en open harten in de familie van Ravenzwaaij-de Boer. Ik had me geen lievere schoonfamilie kunnen wensen. Bedankt voor het geven van liefde, troost en steun aan Don, de kinderen en ik. Bedankt dat jullie altijd vraagt hoe het met mijn studie gaat en dat jullie altijd nieuwsgierig zijn naar mijn werk. Tot slot, bedankt dat jullie me mijn lieve man hebben gegeven.

For my friends...

Thank you Maarten. You have given this project meaning for me, as much as you have given it guidance. You have made it fun, interesting and (mostly) easy. You have always made time for me; your door was always open, no matter what time of day (or night!) I had a question or a concern. You have never faltered in your support of me, or in your trust in my ability, and that has taught me to trust in myself and what I can do. You're my supervisor and my favourite co-author, but you're also my friend. I've loved our silly chats and serious talks, and it's hard to imagine them ending when this dissertation gets defended. I guess I'll just have to think up a new project for us to work on.

Thank you Henk, for teaching me the value of going a bit slower, and being a bit more careful than is in my nature to be. Thank you for being patient with me and for weathering the storms. Finally, thank you for taking so much time to go so carefully through my work. You have added value and grounding to this

project, even though it hasn't always been easy, and I am glad you have been part of it.

Thank you E.-J., for being one of the first in my academic 'upbringing' to see promise in me. Thank you for working with me on my first publications, and my first PhD funding proposal. Thanks for always putting in so much work on my drafts. You'll always be one of my academic heroes.

Thank you Merle, Joyce, Jasmine, Max, Ymkje-Anna and Nina. It has been fun to share a lab with you all. I appreciate equally the support and silly chats the weekly get-togethers have provided. I am proud to call you my academic peers and my personal friends. Thank you Tassos, for being my first friend in Groningen, and for always just saying it how it is. Thanks for agreeing to be one of my paranympyhs on my academic wedding day. Thank you Anja, Martijn, Frank, and Hanneke, for being my friends. I appreciate you all in different ways, but the common thread is that I love you all for your kindness and friendship. You have each added something to my academic journey in Groningen, throughout both the ReMa and PhD.

Thank you Lynne and Kelly. I don't see either of you nearly as much as I would like but I love you both so very much. Lynne, my warm-weather holiday buddy and Dutch-language oracle. I was so happy when you suggested we get coffee together that night after Dutch classes ended. Maybe when R's a little older we can plan Malta? I really miss the hours of sun, and resting, and walking and talking together. Kelly, mate. I'm sorry I'm so shit at replying to emails and texts. You have always been part of my academic dreams, and I wish you could be part of my academic reality. Australia's just too damn far. I miss you.

Thank you Chris, Ivan, Charlie, and Berna for meaningful and/or fun chats and for being some of my scientific heroes. You have shaped my thinking about academia in general, and about reform in particular, and have influenced the content of this dissertation.

Thank you Marieke, for your guidance, encouragement and for your mentorship. It was good to know your door was always open, and that you always made time to talk, despite your incredible workload. I always felt like you 'had my back', and would help me make things right.

Last, but certainly not least, thank you to my buddies at JOTE. Thank you for giving me a little community. I love how kind and warm our group is, and I am grateful for you all. Thank you for giving me a platform for my reform activism, and for helping me reify my ideas.