Standardization in science: effects and issues of guidelines for biomedical reporting

Schniedermann, A.

Chapter 1

Historical background and social contexts

This is an article with a mug of beer in its hand.

Greg Myers (1991, p. 59) about the writing style of Francis Crick's review "Split Genes and RNA Splicing" from 1978.

1.1 <>

In the previous introduction, we got an idea about criticisms of scientific literature that turned into the perception of a global science crisis. We have also got a glimpse of how the biomedical community attempted to solve one of the issues by issuing and implementing new reporting standards for scientific authors. But before we turn towards this particular attempt at science reform and the details of how it came about and became pervasive, we might be tempted to ask, "What happened to the biomedical research literature?" Now, we will dive deeper into the social and historical background of my research project to address this question.

First, I will provide the fundamental role of formal publishing for the globally interconnected research system that is modern biomedicine. I will briefly introduce several aspects of traditional and modern research reporting. These aspects are as fundamental as how scientific practice and observations in the laboratory become written language and published facts. Furthermore, recurring patterns in the writing and structuring of texts form literary types or genres, which are not only characteristic for the discourse but also unfold social capabilities and performative powers in scholarly communication. Furthermore, I will show how biomedical publications can influence healthcare policymaking, medical regulation, or medical practice. These domains, however, are driven by their own logic, interests, and requirements, which, in turn, influence and shape scholarly communication. I will also explain how factors such as

authorship, journal editing, and research evaluation transformed scholarly communication throughout the 20th century.

Some of these prehistories can be dated back to the very origins of modern science, i.e., long before the 20th century. However, the historical remarks and intellectual antecedents that I will provide should not be considered as a chronological and rigorous historical investigation of the advent and fall of modern publishing. Rather than writing a profound and exact history of publishing, the presented story will cover certain aspects of scientific communication in a more conceptual and selective way that combines fundamental aspects of scientific writing, the role of medical knowledge in society (or, more precisely, the role of society for medical knowledge), and the peculiarities of the biomedical research domain. I hope that such a more panoramic type of historical background supports our understanding of the complexities, causes, and plausibility of the case, i.e., a situation that some experts consider a crisis of scholarly communication. Taken together, this chapter will provide the necessary background for understanding why there are reporting guidelines and why they are important. It will also provide the background for a deeper and more thorough understanding of the later chapters, which themselves feature some historical remarks and may be read independently.

Readers who are well aware of the discourse may skip this chapter, which became the longest one in this book. Several aspects I review in this chapter are established and more fundamental knowledge within the science studies or elsewhere. For example, my discussion on literary genres might be well-known to literary scholars, as is my history of randomized controlled trials to medical historians. However, it is not the reframing of any particular one of these strands but rather their combination into a single story that I consider as my unique and novel contribution. In this regard, I want to mention my background in the science studies, which is not too deeply rooted in any of its constituting fields of research but rather starts with an interdisciplinary curriculum. Additionally, I aim to present my research in a manner that requires minimal prerequisites and background knowledge, which will serve as the starting point for this chapter.

1.2 Writing knowledge

Formal scholarly communication can be understood as the functional backbone of modern science. Today, biomedical research is carried out across almost every part of the world, spanning different nations, languages, and cultures. Only when we write down, describe, and publish biomedical insights and ideas do they become part of a global stream of knowledge (R. A. Day, 1989). Other researchers can adopt published

insights, discuss them, compare them to other ideas and results, reject them, or accept them as firm knowledge. Over the course of time, some ideas become outdated, forgotten, and eventually rediscovered. But successful ideas and insights can spark subsequent research and evolve into bigger research trends that drive scientific progress. Some ideas even leave the realm of the biomedical research sector and culminate in new medical therapies or pharmaceuticals, which then become part of healthcare systems. In that sense, biomedical knowledge can have a very straightforward impact on the lives of ordinary people.

This depiction of medical research and its scholarly communication might sound overly positive and idealistic. But what are the social and technical anchor points that construct and shape the realities of the medical sciences? Scholars from various backgrounds in the science studies have shed light on several different dimensions and phenomena around medical publishing. In fact, historical reflections on medicine and the communication of medical knowledge existed for centuries and established their own scholarly tradition. But an important moment when biomedicine received substantial attention by ethnographers and sociologists of science was the so-called *laboratory studies*. While this stream of research focused predominantly on the events and practices performed by researchers in the lab, they also drew attention to the relationship between doing and writing science (Knorr & Knorr, 1978; Knorr-Cetina, 1981; Latour, 1987). Others unveiled the rhetoric formations and "literary technologies" of science (Shapin & Schaffer, 1985/2011, p. 25). These form the technical language of scientific articles and represent the fundamental principles of formal scholarly communication (see also Myers, 1991). Scientometrics and an adjacent branch in the sociology of science have similarly focused on the reception of scientific writing. Here, scholars used metadata from scientific publications to create large-scale insights about the social configurations and dynamics of scholarly communication (Gingras, 2016; Petersohn & Heinze, 2018; Wouters, 1999).

Literary scholars also analyzed how these textual formations became repeated patterns and standardized text forms in scientific writing. They studied how such genres get constructed and communicated and how they spur likewise repeated patterns in reception and standard behaviors among audiences (Bazerman, 1988; C. Miller, 1984; Swales, 2004). They showed us how texts can unfold performative capabilities and do things in the world and in science too (Bazerman, 1994; V. K. Bhatia, 2016; see also R. Day, 2021). Thus, medical trials and reports thereof have become testimonials of the efficacy of drugs that impact the well-being and safety of patients and the balance sheets of pharmaceutical companies alike (Carpenter, 2010; Marks, 1997; Michaels, 2008; Sismondo, 2009). Coming from a different background, medical sociologists also

shed light on such performative capabilities of texts when they studied how protocols and medical treatment guidelines shape the practices of doctors and nurses in evidence-based medicine (Keating & Cambrosio, 2012; Knaapen, 2013; Stepp, 1999; Timmermans & Berg, 2003).

Many of these perspectives shed light on how texts and documents create links between their authors and their audiences. Of course, a common question is, "What makes a text valuable for its readers?" Is it the knowledge it reports? Is it the way it is formulated? While the value of science might be one of the most traditional questions in the philosophy of science or traditional science studies (K. C. Elliott, 2017; Merton, 1942/1973), this question has been approached by those who want to govern and steer science more recently. Their answer to this question had a profound impact on scientific practices, among them writing and publishing, as we will see below. Because of this impact, recent studies in the sociology of science drew their attention to the effects and consequences of evaluating science based on writing and publishing. They showed how the current evaluatory regimes fundamentally changed the behavior of medical researchers and the face of the published record (Butler, 2003; de Rijcke et al., 2016; Gläser et al., 2002; Woelert, 2013). Up to a culture of *publish or perish*[5], medical researchers are incentivized to publish more and more units of texts, allegedly at the cost of other quality goals such as relevance, robustness, or validity (Rushforth & de Rijcke, 2015; Van Dalen & Henkens, 2012; Weingart, 2005).

The relationship between laboratory practice and reports thereof, the technical rhetorics of scientific texts, the performative powers of medical guidelines, and the role of published articles in science evaluation—writing and publishing research—is central to all these perspectives. On the one hand, we can see how writing and publishing have become so crucial for a globally interconnected biomedical research system. But on the other hand, this brief overview of the different perspectives already offers a set of phenomena that seem to be distractions from any pure scientific pursuit. Why would biomedical researchers need rhetoric when they could just communicate the facts? What does it mean for the biomedical literature if medical industries, who are driven by private interests, depend on its contents and types of reporting? What are the implications of researchers being evaluated and promoted predominantly based on their writing and publishing? Such questions imply a multitude of expectations and pressures on biomedical writing, which turns it into a tug of war between different interests and actors.

[5] See also Moskovkin on the origin of the phrase (2024).

I will argue that some of these pressures are inherent to the craft and philosophically indispensable for acts such as scientific writing. Others, however, have led to criticisms and scandalizations, proclamations of crises, and finally led to interventions such as reporting guidelines. The following three sections will trace some of the above-mentioned aspects in more detail to better understand the different pressures on the biomedical research system and how these pressures influence each other. In chapter 1.3, I will introduce some fundamental aspects of codifying scientific knowledge in the context of medical research. Afterwards, I will explain how biomedical texts have performative powers within research and even beyond, and how these powers become engraved in literary genres and document types. Afterwards, in chapters 1.5. and 1.6, I will focus on the massive growth in scientific publishing, the emergence of quantitative research evaluation during the 20th century, and its implications on the wider research sector.

1.3 Codifying medical knowledge

1.3.1 The medical research sector

When researchers talk about science, they usually speak of a particular scientific field they were trained in. These stories create popular histories of science, which are often populated by successful lighthouse projects and historical turning points that give us impressions of big science and major breakthroughs (de Solla Price, 1963; e.g. Launius, 2007). But cultures and practices in research differ substantially between physics, chemistry, or social science, and even between medical subfields such as surgery, nursing science, or epidemiology. Researchers and science communicators from any field discuss key moments of discovery and breakthrough and create popular stories of their field.

In biomedical research, the development and proliferation of the medical trial, or randomized controlled trial has become such a story. Many discoveries and advancements in the field result from the continuous performance of medical trials, so that its emergence and importance has become firm part of the field's "conventional history" (Podolsky, 2010, p. 355). However, although any sense of *the medical trial* may have become a boundary object for the field, we must keep in mind that it cannot be pinned to a particular place in the world or moment in the history of the field (Marks, 1997).

When asked about the role of medical trials in the history of medicine, medical research, or drug regulation, many scholars in the field would likely reference Sir Austin Bradford

Hill's 1948 study on streptomycin. For instance, an editor of the New England Journal of Medicine concludes that "the discovery and use of antimicrobials is doubtless one of the greatest accomplishments of medical science in the millennium just past." (NEJM, 2000, p. 48). The trial proved the effectiveness of streptomycin as an antibiotic treatment for tuberculosis. Popular histories point towards its invention and execution of a rigorous trial methodology, which involves control groups and randomization, that turned the randomized trial into a paradigm of proving medical efficacy or effectiveness (Crofton, 2006; Epstein, 2007).

But more sophisticated accounts of its origins argued that it wasn't just the methodological innovation but rather the social and political context, in which healthcare reformers and academic statisticians banded together, that made the method thrive (Bothwell et al., 2016; Marks, 1997, 2000; Podolsky, 2010). Although it was not the first time that modern trial methods were applied, their rigorous execution and discussion in reflexive reports put them in a new light. The success of the trial's design became a starting point for the further development of the methodology (Junod, 2008). Initially, it was especially the tuberculosis research community that became accustomed to this way of doing clinical research, which quickly spurred further refinements in their fight against one of the most severe diseases at that time (Crofton, 2006; Marks, 2000; see allso Podolsky, 2010).

Besides the importance of the social and political contexts that paved their way, randomized controlled trials can be understood as a culmination point of different intellectual strands within science. Some have argued that the emergence of modern psychology fueled the rise of experimental control group designs in medicine. The young field sought out "marketable methods" (Danziger, 1998, p. 101) and, although originally focused on the assessment of educational techniques, spearheaded a culture of collaborating with other disciplines and non-academic actors (Orr, 2018). But most importantly, sophisticated statistical methods became a key part of modern research studies and especially controlled trials.

Only with proper statistics could quantitative observations and measurements be accumulated and transformed into an overall study result such as a correlation coefficient or an effect size. It has been argued that modern statistics as a professional field co-evolved with the growing attention given to experimental trials in medicine and psychology (Marks, 1997; Valier & Timmermann, 2008). For example, hypothesis testing was originally developed during the 1930s by Ronald Fisher and Karl Pearson but is still considered as a standard tool in modern statistics (I. Chalmers et al., 2002). As we will see throughout this research, statistics and statistical reporting play a crucial role in the

development of other research designs, such as meta-analyses and systematic reviews. But medical statistics have also been at the center of critiques of the state of biomedical literature, with many critics identifying statistical reporting as a major factor for the alleged publication crisis the field is experiencing today, as we will see further down below. However, it wasn't just intellectual developments such as statistics and experimentalism that shaped the face of a novel way to do medical research during the early 20th century.

As already mentioned, the success of the streptomycin trial and randomized-controlled trials in general depended very much on the social and political contexts of postwar medical research. As such, the very idea of a biomedical sector that comprises academic and private research institutions, hospitals and doctors, pharmaceutical companies, and even patient organizations has been identified as the fertile ground for innovative study designs (A. E. Clarke et al., 2010).[6] So while the different research practices were not particularly novel, the trial was performed in a new form of cooperative investigation which used epistemic narratives for political purposes, i.e. about the superiority of statistical reasoning, but especially benefited from powerful socio-political arrangements (Marks, 2000). For example, during the Second World War, the British government had reorganized medical actors in order to keep the medical system capable when Britain became involved in the war. The British Medical Council established links and collaborations between different medical faculties and hospitals, originally for the sake of crisis response. This led those actors to get used to a type of collaboration that connected medical research with medical treatment, so that knowledge could be translated from the "bench-side" to the "bedside",[7] and paved the way for more straightforward processes between the development and testing of therapies (Curry, 2008; Meldrum, 2000; Roth, 2022). In addition, the reorganization enabled a diverse set of actors to get used to each other and streamline collaborative efforts. Research faculties, hospitals, funders and regulators, and also the pharmaceutical industry came with different perspectives and interests with regards to modern medicine (Valier & Timmermann, 2008).

Many of these social and organizational arrangements prevail until today. Modern medical trials are multicentered research projects that are dispersed over multiple organizations of various types. They can span organizations from different countries and

[6] See Roth on a dedicated history of that term (2022).

[7] "Benchside" refers to the workbench in the laboratory, while "bedside" refers to the treatment of medical patients. Sometimes, this phrase is also used to describe the connection between basic and applied research, i.e., clinical trials.

thus regulatory frameworks for pursuing research and managing research funding (Carney, 2000). Additionally, some collaborative multicentered trials can last for many years or even decades, necessitating measures or structures to withstand changes in personnel and funding (see also Cambrosio et al., 2009; A. E. Clarke et al., 2010; Cosgrove et al., 2016; Keating & Cambrosio, 2012; Roth, 2022; Sismondo & Greene, 2015; Solomon, 2015; Timmermans & Berg, 2003). In a more recent trend, such complex networks of different organizations, motivations, or practices and cultures of doing research are turned into internationally operating research consortia (Catney & Lerner, 2009). Such consortia can be created to work on a specific disease or medical condition and develop an array of interventions such as drugs, therapeutical processes and treatment guidelines, or even healthcare measures (Cambrosio et al., 2009; Keating & Cambrosio, 2012). Historical studies have shown that collaborative efforts and new interdisciplinary configurations can further stabilize and support the development of new research infrastructures or fields. In the wake of such developments, new communities are formed, additional actors are enrolled, and new tasks and specializations allow for novel career paths that solidify the future of the community (Fujimura, 1988; see also Galison, 2010).

The ways and functions of how locally and temporally dispersed researchers in biomedicine form a global research community have gained substantial attention in the science studies. As indicated above, a central factor that makes researchers find themselves in a discipline or community is their research practices, which they might share with peers from the opposite side of the globe as much as with their colleagues next door. In that sense, a community can be understood as being made up of individuals who ask similar research questions, choose the same study designs, methods, or apparatus, or draw conclusions in a similar way (Abbott, 1988; Bowker & Star, 1999; Whitley, 1985/2000). Another crucial aspect is how researchers present, discuss, and reflect on the various practices and procedures that take place in their laboratories. Especially by exchanging written texts, researchers form discursive communities that bridge physical, political, or cultural borders (Swales, 1990). In such communities, scientific publications are carriers of information that can be retrieved anywhere in the world at any time. They can be read and consumed independent of the moment when the experiments were performed or the reports were drafted (Shapin & Schaffer, 1985/2011). It might become apparent, then, that writing and distributing texts are the critical factors in turning individual experience into collective knowledge. But how are those bridges built? How can writing about medical research (co-)create a research community?

1.3.2 Stabilization of concepts, descriptions, and genres

A crucial aspect in medical research is the reification and standardization of epistemic objects, which comes with the establishment of classifications, standards, and nomenclature—community-defined concepts and descriptions of things and practices (Bowker & Star, 1999; Rheinberger, 1997). If an author uses a concept in her medical research paper, for example, "tuberculosis" as the name for a disease, she agrees to a whole history of coining, debating, and redefining that concept. Being coined and discussed by previous medical researchers, the concept emerges from established disciplinary knowledge where it is used and interrelated in always similar and recurring ways, so that "even a single word can represent a complex theory" (Fleck, 1935/1979, p. 42; see also Chang, 2012).

Especially medical practice and the medical sciences have created impressive catalogues of standardized concepts and classifications that allow for identifying, sorting, and filtering things such as diseases, drugs, chemicals, specimens, treatments, technical equipment, and even humans (Bowker & Star, 1999; Lampland & Star, 2009; Timmermans & Berg, 2003). Scholars in the science studies draw a lot of attention to the social lives of such classification systems and discuss how definitions such as *the 70-kilogram adult white male* as the *standard human* can be inherently problematic. Such standards not only neglect diversity and promote exclusionary characterizations of the world but can have dangerous implications. An effective dosage of a drug for a standard human may be a lethal overdose for a non-standard one (Epstein, 2009). But besides performative and actionable moments that emerge from classifications, their representation in the form of concepts is essential for disciplinary communication. But beyond assigning names to things, conceptual homogeneity also means to interrelate concepts, depict dynamic phenomena, and describe the research procedure itself.

The emergence of new scientific methodologies and research designs not only comes with the application of novel investigative techniques but also with certain ways to report their application and discuss their benefits and drawbacks. As such, the standardization of medical research also implies that the field grew a reflective literature that is concerned with forms rather than contents. For example, it specifies and discusses research designs (such as randomized controlled trials), techniques of data aggregation and statistical tests, tools for manipulation and research management, and practices of disciplinary self-regulation (Abbott, 1988; Fujimura, 1987, 1988; Meldrum, 2000; Stephan, 2015). For instance, Sir Bradford Hill, the project leader of the famous streptomycin trial, simultaneously published a handbook on the method used in the trial (Crofton, 2006). In a similar vein, Ronald Fisher's publications *Statistical methods for research workers* (1925) and *The design of experiments* (1935) paved the way for statistics to become a part of the medical research discipline (Valier & Timmermann,

2008). Later, medical statisticians were one of the first communities that criticized issues in applied statistics on a field-wide level and promoted checklists for assessing statistical reporting (e.g., Gardner et al., 1986; see also Peterson & Panofsky, 2021). Such works not only contributed to finding a gold standard for doing trials but also suggested what information is crucial to understanding statistical analyses and thus reports of medical research projects.

The forms, concepts, and parts of biomedical research reports relate to the various things and phenomena that a research community is concerned about. Concepts and vocabulary originate from what a community of researchers consented upon and form that community's knowledge (Knorr-Cetina, 1981). In contrast, descriptions of phenomena and observations allow readers to retrace the experiences of the investigators, although the forms of descriptions again represent common writing styles and patterns of scientific storytelling. This allows readers and scientific peers to gain insight into the study, even if they weren't present during its execution. Based on the concepts and descriptions used, they become able to connect the trial with their own experience and domain knowledge (see also Kitcher, 1993). Moreover, the report allows them to judge if the trial was performed in accordance with the disciplinary standards and decide whether it was a legitimate scientific activity. Thus, readers of scientific reports become "virtual witnesses" (Shapin & Schaffer, 1985/2011, p. 60), who envision the sequence of events and replicate the performed experiment "in the laboratory of the mind and the mind's eye" (Shapin & Schaffer, 1985/2011, p. 60).

The application of standardized vocabulary and nomenclature to explain the sequence of events in a research project must be understood as part of the core set of literary techniques that characterize the experimental report. Only if the author describes the apparatus, the manipulations performed by the experimenter, and the resulting phenomena may the reader follow the sequence of events. In that sense, experimental reports live up to the epistemic values and quality goals of experimental research, for instance, objectivity and validity, by the proper execution of procedures (Amsterdamska, 2005; Douglas, 2004). Although Steven Shapin and Simon Schaffer developed their history of experimentalism based on the case of Robert Boyle's air pump experiments from the 17th century, we can find that the methodological peculiarities of randomized controlled trials in medicine are informed by the very same principles.[8] In this regard, David Moher and Jesse Berlin, two important metascientists,

[8] Interestingly, Harry Mark's history of the clinical trial does recognize Shapin and Schaffer's account only superficially, supposedly because Marks draw a more distinct

wrote "The ultimate test of systematic reviews, as with randomized controlled trials, is whether the reader has the confidence that the report is an accurate reflection of what went on during its various stages" (Moher & Berlin, 1997, p. 267). While we will turn towards these two study types again later, this quote illustrates that codification is not just about concepts and background knowledge.

Effective reading and virtual witnessing require standardized descriptions of scientific practices and procedures. Similar to how the scientific concepts used in journal articles connect to the domain knowledge of their readers, the descriptions of laboratory practices must resonate with the reader's own work experiences. Only then can the reader "go along" (Latour, 1987, p. 60) with the paper and become advocates who fuel its dissemination, defend its ideas, or even build further research on it.

The stabilization of concepts, descriptions, and related definitions supports researchers in their communicative efforts and makes them efficient and effective. Rather than individually define and describe their work, they can subscribe to dominant definitions and reuse them for reporting and presenting results (Bowker & Star, 1999; Xylander & Nordmann, 2022). The uniform usage of concepts allows for community members to talk about the *same thing* without requiring prolonged explanations and "virtuosities of communication" (Bazerman, 1988, p. 23). Moreover, researchers can express their findings in the form of new conceptual linkages or definitions, for example, by creating a relationship between a chemical agent and the function of an organ in the human body.

Concepts and descriptions become pointers towards the codified and tacit knowledge of the audience and its disciplinary identity (Bazerman, 1981, 1988; Collins, 2012; Kitcher, 1993). For example, *acid* as a concept not only has an intellectual history as a chemical component that was discovered and described. Moreover, if it is communicated to a chemist, it prompts their professional knowledge around those concepts, such as its chemical formula, the way it reacts with other agents, or required safety measures when dealing with it in the laboratory. For the researcher, a single concept can determine what actions should be taken, what questions might be appropriate to ask, or what knowledge is required to understand what is going on (see also Chang, 2012). In principle, a single word can become the key to a library of

boundary between natural, laboratory-based science and the clinical sciences he focused on (for example, see Marks, 1997, p. 8 for reference).

knowledge.[9] In the context of scientific literature, such links are not only established by concepts but also by formal references to other texts (S. Wyatt et al., 2017).

On a broader scale, establishing concepts and standard descriptions sets the foundation of the community. In that sense, concepts and descriptions can be seen as building blocks and socio-cognitive devices of a field's intellectual discourse. Theoretical accounts from the sociology of knowledge stress that these devices are part of the phenomenal structure of the discourse, which highlights their meanings, relationships, causal schemes, and normative implications (Keller, 2006). Their ontologies, i.e., ways how they represent material reality, which are central to philosophy of science (Dummett, 1973), are not of primary concern. Thus, the emergence of new theories, trends, or scientific communities has an intricate relationship to the creation and solidification of new scientific concepts and descriptions, irrespective of the materialities to which these concepts may refer (Fujimura, 1992).

But occasionally, concepts, descriptions, and their links to disciplinary knowledge become outdated and must be reformed. New observations or theoretical perspectives can challenge the dominant interpretations of a concept or their links to other concepts. In such cases, the concept or description becomes uncertain to some extent, and community members redefine its meaning or decide to abandon the concept altogether (Dupré & Leonelli, 2022; Fujimura, 1988; Whitley, 1985/2000). Although especially the conceptual stability makes communication efficient, a certain fuzziness, incompleteness, or "science friction" (Edwards et al., 2011, p. 667) remains. Conceptual uncertainty can be reduced by active communication. Otherwise, it leaves room for creative interpretation and spurs new ideas or serves as an appropriate playground for

[9] In contrast to the natural sciences, the social sciences and humanities employ standardized sets of concepts and concept-knowledge relationships to a far lesser extent. In this regard, it has been argued that there are disciplinary differences in the degree of codification of knowledge (Zuckerman & Merton, 1972). Often, the relationship between concepts and their explanation is itself the main object of the intellectual debate in the humanities. Traditionally, the explanation for the difference refers to the epistemic role of language in the different fields. Under empiricist epistemology, researchers understand language as their way to access the structure and logic that is inherent to nature, so that knowledge just has to be discovered (see also Bazerman, 1981). In contrast, constructivist epistemologies emphasize how the logics and structures in the world are rather created by human observers, where the logic and structure of the language play a fundamental role in this construction process. Against this background, scholars shifted their attention from products to practices or processes, which is why some speak of a communicative turn in the science studies (Keller, 2011).

different perspectives.[10] It allows researchers to creatively “distrust scientific language” (Bazerman, 1988, p. 284). In another case, new insights or perspectives can exceed the scope of the dominant concepts and interpretations in an irreconcilable way. In such cases, all what is new is bound together in new theory-method packages or even whole new research disciplines. In turn, those are represented by new concepts, for example “systematic reviewing” (I. Chalmers et al., 2002), “synthetic biology” (Blümel, 2021), or “biomedicine” (Roth, 2022). Therefore, redefinition and abandonment of established concepts, as well as the emergence of new ones can be seen as vital moments of scientific debates, even if they fuel bigger controversies (Schaffer, 1989; Kuhn, 1962/1996).

1.3.3 Between writing, reading, and understanding

So far, scholars in the science studies have explored the fundamental principles of (formal) scholarly communication. I have briefly sketched the process of transforming practices into descriptions thereof and vice versa. We have seen how research communities create standardized vocabularies that enable stable communication across locations and time. However, these standardized vocabularies also inhibit the fuzziness and uncertainty that allow progress and reform. As previously mentioned, this description of the basic principles of scholarly communication can be viewed as part of an idealization of biomedical research. This idealization has been criticized for telling only parts of the whole story and exaggerating those aspects that seem to be rather rational and plausible, while leaving out much of what really happens in the labs around the world (Csiszar, 2018; Golinski, 1998, p. 201; Knorr-Cetina, 1981). Scholars shed light on the contingent, fuzzy, and non-standard aspects and behaviors observed in the day-to-day practices of biomedical researchers and research authors. Further, they challenged prominent histories of science that argued how rationality, objectivity, or survival of the best idea fueled the historical pathway towards science as we know it today.[11]

---

[10] The STS community can provide an illustrative anecdote in this regard. Scholars explore the ambivalences and ambiguities of individual concepts in creative ways. Often, whole conference themes are developed along a single or a few individual concepts, for example „Circulations“ (Lippert et al., 2023), “Leakage” (https://stsing.org/leakage-inaugural-stsing-conference-tu-dresden). "(In)Sensibilities" (https://4sonline.org/4s_2017_boston.php), or "Innovations, Interruptions, Regenerations" (https://4sonline.org/4s_2019_new_orleans.php)

[11] For example, historical accounts about the works of Galileo Galilei, Louis Pasteur, or Robert Boyle have emphasized the uphill battles these scholars often had to fight within

In terms of reading and writing as components of scholarly communication, it is often necessary to shed light on the role of rhetoric and persuasion. Scientific concepts, explanations, definitions, or styles in describing scientific experiments have been conceptualized as powerful “literary technologies” (Shapin & Schaffer, 1985/2011, p. 25), or persuasive devices that equip authors with the capabilities to bind readers and make them believe what is written in the report. Such devices might be certain descriptions of events and procedures, disciplinary jargon and standards, tables and graphs as forms of presenting data as immutable facts, references to previous literature as support, or the aesthetics of language and presentation. Further, literary technologies and rhetoric strategies exert performative powers on their readers, making them (re)act in certain ways. At the most basic level, authors try to convince readers of their own and their team members' credibility and proficiency as researchers (Bazerman, 1988; R. A. Day, 1989; Golinski, 1998; Knorr & Knorr, 1978; Latour, 1987; Myers, 1991). In this regard, Greg Myers (1991) compared two review articles in biomedicine and showed how a formal and technical writing style persuades readers of established facts, while a more informal and even artistic writing style can be utilized to discuss uncertainties, questions, or speculations in a confident and powerful way, as nicely illustrated by the quote in the beginning of this chapter.

It follows from this perspective that there is not only a necessary mismatch between the contingency and complexity of scientific practices and the *tidiness* or *thinness* of written and published texts (Bazerman, 1994; Kaplan, 2022; see also Knorr-Cetina, 1981). Rather than providing a mechanistically reproducible report that deterministically follows from the performed research process, scientific texts must present a story of the performed experiment that is plausible and thus believable (Knorr & Knorr, 1978; Shapin & Schaffer, 1985/2011). Reports are not merely derivatives of real events, but they have their own politics, requirements, and environments. Consequently, aspects of the experiments performed may be deemed implausible or superfluous to the story in the report, and vice versa. Therefore, there is a discrepancy between the information provided in a scientific publication and the experiences, observations, or information available to the original investigator and author.

The gap between research practices and descriptions thereof provides a room of uncertainty that is constructive and destructive at the same time, as we will see throughout the rest of this book. This gap is affected by the many tasks and processes of knowledge production as well as the rules and conventions of scholarly

---

their academic community (Finocchiaro, 2014; Latour, 1988; Shapin & Schaffer, 1985/2011).

communication, as explained above. The gap is usually managed and handled, used, or even misused by authors, either creatively, strategically, intentionally, or unintentionally. [12] Particularly in modern biomedicine, many different actors and contexts formulate expectations of scientific authors. Most notably, medical and pharmaceutical industries pressure knowledge production processes to meet their financial and regulatory interests. For example, while publishing usually involves different practices, actors, and contexts, which introduce a lot of contingency into the whole process, pharmaceutical companies "ghost manage" the whole pipeline in their favor (Sismondo, 2009). But healthcare policymakers and medical practitioners also exert expectations from knowledge production, which can mean an additional source of pressure on this gap, as I will explain below. However, texts often have an impact on the world because audiences may react in certain ways.

Shedding light on the performative power of texts, however, paved the way for a so-called "linguistic turn" in the science studies, which understands language as a fundamental building block that shapes thoughts, beliefs, and practice (Keller, 2011, p. 56f). In this regard, Charles Bazerman concluded:

> *The emergence of certain patterns of written communication give generic qualities not only to texts, but to the way the texts are used in situations, and even to the character of the situations themselves. Writing is social action. Regularized forms of writing are social institutions, interacting with other social institutions.*
>
> *(Bazerman, 1988, p. 22)*

The plausibility of this thinking is appealing. If we read the results from a microscopic experiment that was performed a couple of years ago, we can access neither the microscope nor the specimen nor the practices of observation and interpretation directly. The only thing at hand are the concepts and their connections laid down in the report of the study. Any sense of retracing the experimental steps or even doing a similar

[12] Although the science reform movement vocally criticizes bad intentions, misbehavior, and fraud, philosophers and sociologists of science have stressed the role of unintentional uncertainty and unavoidable variability. For example, early microscopists had to draw what they saw. These drawings not only varied by lighting and setup but also by how microscopists perceive and interpret basic shapes when observing the same biological component (A. F. Chalmers, 2013).

experiment is always bound to its description, which usually comes in textual form, rather than the experiment itself. Therefore, scientific texts are not just representations of real-world events that act as gateways to previous experiments. Instead, the texts are the events themselves whenever they identify a recipient (Hull, 2012; Keller, 2011; Shankar et al., 2017).

Taken together, we can see that the biomedical research literature is not just a logical consequence of the research practices in the lab. Instead, it is heavily influenced by the background knowledge of researchers, conventions in naming things and processes, as well as interpretations by authors and audiences alike. In addition, authors can even try to utilize the persuasive powers of rhetoric to convince their audiences. But many of the perspectives just discussed highlight a fundamental problem of writing. We investigated how information is translated from real-world situations into the virtual spheres of written (and spoken) language and vice versa. Both spheres are utterly complex and contingent so that we must understand the gap between conduct and reporting, i.e., the process of translation and codification, as a fuzzy and seamless space (see also Bazerman, 1988). This space provides substantial degrees of freedom for authors to choose concepts and descriptions and write their texts (Wicherts et al., 2016).

The degrees of freedom of writers can also lead to conflicting quality judgements within a community. Scientific peers may perceive texts as insufficiently reported or even concealing, as they did so vocally in the discourse about transparent reporting in biomedicine. For example, metascientists such as Doug Altman (1994) prominently elucidated how the expectations of pharmaceutical companies may pressure authors towards certain conclusions in their studies, as we will see in detail further down below. However, rather than explaining *why* medical researchers find themselves in a state of crisis as outlined above, the gap between conducting and reporting and its practical manifestation is more helpful in explaining *how* published studies may diverge from agreed-upon standards and quality expectations. The role of concepts and descriptions of procedure is crucial in this regard and must be elaborated briefly.

Individual concepts can be interpreted differently, but differences sometimes cross a threshold of what others regard as acceptable. Some will deem the variance too great and consider the interpretations to be misinterpretations or even misuses of the concept. Criticisms of the reporting of the randomization process in medical research provide an illustrative example in this regard. Randomization means that study subjects are drawn randomly from the general population so that the sample represents the variety and diversity of the assumed population and that any confounding factors that may result from subgroup-related peculiarities are mitigated (Epstein, 2007). It has

become an integral part of the epistemic promises that are associated with clinical trials. Randomizing is so essential for the epistemic value of trials that the technique was defined as the gold standard in evaluating treatment interventions and even became part of the genre identity "randomized controlled trial" (Timmermans & Berg, 2003). The concept of "randomization," however, has been troublesome. Some researchers seem to interpret it differently than others, so that many types or techniques of randomization are reported in medical studies.[13] Some of these reported techniques have been criticized as less rigorous or even wrong and invaluable (Rennie, 1995; Schulz et al., 1995). Because of the importance of randomization for the validity of medical knowledge, research experts grant authors only a low degree of freedom in interpreting the concept, so that the gap between conduct and reporting is very little in this case (e.g., Altman & Doré, 1990; Weßling, 2011). This example shows how conceptual negotiations and alternative interpretations of concepts are not always moments of potentially fruitful epistemic diversity as they can lead to immediate rejections and criticism in certain cases.

Beyond concepts and their interpretation, explanations of procedures can be even more ambiguous and fuzzy. The gap between a series of laboratory practices and explanations thereof has been subject to debates since the origin of the experimental article (Csiszar, 2018; Shapin & Schaffer, 1985/2011) and remained an important perspective for understanding experimental practices or pathways of whole scientific communities, such as gravitational wave physics (Collins, 1985). However, the insufficiency of textual representations for the factual or virtual reproduction of performed experiments has become especially prominent with the replication crisis in psychology and biomedicine that came to a head in the early 2000s (Field & Derksen, 2021; Nelson, 2024).

Naturally, practical complexities and local contingencies of experiments get lost in the process of abstraction and codification (Bazerman, 1983; Knorr & Knorr, 1978). In fact, making communication efficient means that codifications necessarily require truncation and reduction. Such reductions mean that observers must evaluate and rank information and communicate only what they find most important. In contrast, any full account of information, if it can exist at all, would just mean flooding the audience with

[13] A different example is the calculation of "sensitivity" by medical researchers versus information scientists who use the same formulas to calculate "recall". Since mathematical procedures span many different fields, the same research practices, i.e. mathematical formulas, may be referenced differently. A more abstract conception of this problem is the distinction between meaning and reference which is fundamental to the philosophy of language (Dummett, 1973).

all available information. For example, even highly technical and desk-based study designs such as meta-analyses involve a multitude of decisions about thresholds and parameters (Moreira, 2007; Stegenga, 2011). Although such decisions can be reported straightforwardly, each author must weigh their importance to not flood the final report with information. As a result, a story of discovery in a scientific paper is logical rather than accurate (Howitt & Wilson, 2014). But since reading and comprehension are individual tasks, how could authors even know what information any reader would require?

The recent discourse of the replication crisis features both the problematization of the description of procedures and the interpretation of concepts. Many attempts to systematically replicate previous studies have failed. One reason is the aforementioned information gap between conduct and reporting that has been so richly described by historians and sociologists (Bazerman, 1983; Knorr & Knorr, 1978; Shapin & Schaffer, 1985/2011). But because this gap is more fundamental and maybe unsolvable, many scholars reject idealistic depictions of direct replication in which the *same* research questions are tested by following the *same* study protocol. Rather, a less algorithmic idea of replication considers variance and contingencies as epistemic benefits, because they can prove the robustness of concepts, research questions, and descriptions (Derksen et al., 2024; Guttinger, 2019; Haig, 2022). Therefore, it is not the gap between practical choices and their descriptions that does render reporting insufficient for the purpose of replication.

The gap between conducting research and the descriptions of that research must be regarded as fundamental and inevitable. But if it is basic and has been part of technical writing's history, it can't be the cause of the recent uproar and crisis claims. Rather, the gap paves the way for the crisis in a technical sense. The gap between conduct and reporting offers researchers degrees of freedom in selecting, weighing, and describing their observations. The gap must be understood as a necessary condition for a reporting crisis to exist because it reminds us how scientific publications are not just inevitable and formal machine-like outputs of research projects. Rather, they are carefully constructed tools or devices that are required for multiple, different, and sometimes conflicting trajectories, as I will sketch out in the coming sections. In this regard, Nobel Laureate Immunologist Peter Medawar referred to the fundamental laws of formal scholarly communication instead of misconduct when asking, “Is the scientific paper a fraud?” already in 1963 (Howitt & Wilson, 2014, p. 481).

So far, we have explored some aspects of the relationship between knowing research facts and writing them down. We have seen how there might be a fundamental gap

between conducting and reporting research, which cannot be overcome at all. To illustrate this point, we examined scientific concepts that can be regarded as the smallest components of scientific texts. I have shown briefly how even these small bits of scholarly communication are entangled in a complex social network consisting of research practices, communicative conventions and standards, or disciplinary idiosyncrasies.

There is much more to say when we examine process descriptions, tables and graphs, or other forms of visualizing scientific facts. Current debates on medical imaging, such as western blotting, show how these technologies might play an even bigger role when it comes to persuasion or deceit in modern science (Beck, 2016; Bik et al., 2016; see also T. M. Porter, 1992). However, I will discuss the importance of checklists and flowcharts for the standardization and stabilization of literary genres in biomedical reporting, such as systematic reviews, instead of focusing on whether and how these are persuasive communication devices.

But speaking of containers or genres, we must also ask for the form and not just the content.[14] This requires us to look for the contexts of production and reception of biomedical publications, i.e., the epistemic promises, norms of production, and social expectations that are tied to genres. I will approach these topics in the next section. Further, we will explore how these issues are related to the gap in conducting and reporting research and see what role they play in biomedicine's publishing crisis. In particular, we will travel along some of the traditional tensions in biomedical research that emerged throughout the 20th century, such as the emergence and professionalization of clinical trials and drug regulation, as well as the resulting problem of an ever-growing stock of medical knowledge.

1.4 Medical research for the greater good

1.4.1 From publication to medical practice

[14] It should be noted that the concept of genre can be much broader than I have presented here. Our focus has been on recurring literary technologies and their manifestations in the form of document types. However, the concept can predominantly be used to describe and categorize content and topics. For example, it is used colloquially to describe the differences between science fiction and fantasy literature. By contrast, when describing and categorizing content and topics in science, we tend to speak of disciplines or fields.

Unlike other research fields, biomedicine can be characterized by its close connection between the *bench and the bedside*, or the generation of knowledge and its application (Roth, 2022). Every day, medical doctors make diagnoses and treatments that decide over life and death, long-term health conditions, or human well-being. Historically, the medical profession emerged as the advocate for the human body and turned the curing of diseases into its moral responsibility (Temkin, 1977). For this moral responsibility, the medical profession gained high social status and became a key pillar of almost any modern civilization. But responsibility also implies high stakes and moral expectations. We want doctors to perform their craft properly and to the best of their knowledge. We want policymakers to create efficient and accountable healthcare systems that benefit society at acceptable costs (Tanenbaum, 1994).

As if such high moral requirements call for a tragedy, industrial interests have always heavily influenced modern biomedical research. As the previous section indicated, many methodological and intellectual developments in the early biomedical sector culminated in the emergence of randomized controlled trials. But the performance of medical trials and knowledge translation does not only involve interactions between research labs and hospitals. Rather, the push behind many medical trials often comes from pharmaceutical companies who want to test and license a new drug or treatment. Although medical industries have become essential building blocks of the modern biomedical sector, their role is also problematic. Companies have private interests that can be at odds with the safety and interests of patients or the moral standards of the medical profession. As a result, the back-and-forth between medical industries, medical implications, and political reforms permeates the history of medicine and medical research, which is why we must look at how medical research and scholarly communication in biomedicine were and are shaped by nonscientific actors and their interests.

After a noteworthy history of hazardous drug disasters at the beginning of the 20th century, the United States Food and Drug Administration (FDA) introduced the requirement to prove drug safety prior to market licensing in 1938 (Marks, 1997). Before the implementation of that regulation, the inventors of these drugs either ignored or downplayed their harmful side effects to avoid market bans and save money. But even after these drug safety regulations, many doctors and policymakers remained skeptical of the industry and had substantial “mistrust in the marketplace” (Marks, 1997, p. 343). One of the reasons was an industry-wide tendency to advertise pharmaceutical products with exaggerated promises lacking any proof. So, although the drugs on the market might have been less harmful because of safety regulations, there was little to no proof that they were effective at all.

Subsequently, a new FDA amendment introduced the requirement to prove the efficacy of drugs, which can be done by successfully performing a controlled trial (Junod, 2008). But even if cases such as the streptomycin trial became influential stories of scientific successes, the overall reception of controlled trials was mixed. In fact, many early attempts to prove drug efficiency with randomized controlled trials yielded disappointing results (Meldrum, 2000) and made medical fields adapt to the new standards at different paces, as Archie Cochrane famously criticized during the 1970s (Cochrane, 1972/1999/1972). Furthermore, despite their efforts to find methods for proving the efficacy of interventions, the new medical standard did not resonate well in other fields of research, either (Gueron & Rolston, 2013; Rossi & Wright, 1984).

Some historians argue that the fast-paced development towards randomized controlled trials was not because of any noble quest for medical efficacy. Rather, with study designs and methodological requirements to prove efficacy, policymakers and health professionals found a way to push back the power of the pharmaceutical industry.[15] But others argue that the medical field was indeed ripe for some methodological improvement. Before the 20th century, systematic knowledge was scarce and medical practice and treatment results varied so much that one might hardly speak of a profession at all. Often, effects were of *blockbuster size*, so that positive or negative effects became visible, even without any systematic evaluation or sophisticated statistical analysis (Baron, 2018).

However, the triumphant advance of the randomized controlled trial continued in the early postwar era. The initial proof of efficacy matured into the formal requirement for a full-scale randomized controlled trial in the 1960s (see Barron & Bukantz, 1967). From that point, no drug could be licensed and sold in the United States without performing a randomized controlled trial, and the FDA even issued the reexamination of thousands of formerly licensed drugs (Jasanoff, 2006). This regulation became what Scott Podolsky called "the ultimate substantiation of the controlled clinical trial as the arbiter of therapeutic efficacy in the United States" (Podolsky, 2010, p. 365). Turning randomized controlled trials into a necessary requirement, the already initiated realignment of research professionals, hospitals, and industrial partners soon evolved into a "clinical trial industry" (Meldrum, 2000, p. 755). This industry established the required settings

[15] During official hearings and discussions, industrial lobbyists tried to make their case by pointing to their companies' high investments in the development and testing of new drugs. This shed light on the actual research procedures employed in the industrial labs, which turned out to be neither costly nor scientifically reliable (Junod, 2008).

and procedures to straightforwardly plan and perform immense numbers of clinical trials.

Bearing in mind that the clinical trials and especially the reports thereof are required to get drugs licensed or healthcare policies justified, one can see how published scientific texts have substantial social power. Clinical trial reports are not only commonly produced within a network of different actors. They also turned their reception within different societal domains into a regular or even standardized phenomenon. In that sense, they can be understood as boundary objects between medical research, pharmaceutical industries, and regulatory bodies (Jasanoff, 1990; Star & Griesemer, 1989; see also Astrom et al., 2017). Therefore, as much as a clinical trial proves drug efficacy, its report, as for example published in a scientific journal, can unfold the very same social powers if played out in the right context. Consequently, publications become arenas where struggles and debates around research unfold. The way scientific communities experience and handle research fraud or misconduct can be interpreted in this way.

Cases of research fraud or misconduct often revolve around journal publications. Often, the reported results of an authoritative paper take center stage, providing a case that attracts a lot of attention. For example, papers get criticized for their lack of detail or the implausibility of the description of events. In such cases, the paper's audience—the scientific community—calls the story of the paper into question. Subsequently, the scientific community might pressure the publishers of the journal to retract the publication so that retractions become an important and visible verdict to correct the scientific record (Hesselmann et al., 2021). Although the biomedical research community originally shunned the retraction of papers, controversies and scandalizations around publications that prove a link between child vaccination and autism (A. Moore, 2006) or using hydroxychloroquine as a COVID-19 treatment (O'Grady, 2024) have increased the community's use of retractions. This also fueled the recent surge in retractions (Freijedo-Farinas et al., 2024).

Publications took center stage in criticisms of the influence of the pharmaceutical industry on biomedical research in a similar vein. For instance, scientists have quantified the funding effect by analyzing scientific publications (Als-Nielsen et al., 2003; Carter et al., 2006). Further analyses have uncovered how pharmaceutical companies ghostwrite reports or manage whole publication pipelines to put trial results in a positive light and get the intervention patented, licensed, or marketed to doctors and the general public (Gaudillière, 2013; Sismondo, 2008). Major scandals of ineffective and even harmful drugs can be found throughout the 20th century. A famous example is study 329, which

aimed to prove the efficacy of the antidepressant paroxetine for teenagers during the late 1990s. The trials did not find positive effects in comparison to placebos but instead uncovered harmful side effects. Since the drug was already sold and used for adults, the manufacturer feared a potential general backlash against the drug. Therefore, the official report on the trials mentioned the general effectiveness of the drug while concealing the actual study results and the harmful side effects in its conclusions (Wilson, 2015; see also McHenry & Jureidini, 2008).

Although several regulatory requirements followed these scandals, industrially funded science has been problematized throughout the history of postwar medicine. In 1987, Robert Newcombe wrote that "attention was drawn to it as early as 1963, and it has been 'rediscovered' several times since." (1987, p. 657). However, it was especially during the 1980s when more systematic assessments of this problem were published. Predominantly studies on publication bias found a positive correlation between privately funded trials and positive trial results (Davidson, 1986; Gøtzsche, 1987; Hemminki, 1980). Subsequently, the first calls for a clinical trials register were uttered already during the 1980s (Simes, 1986). While there are different causes and mechanisms that lead to these results (see Sismondo, 2008), it must be noted that industrial interests and economic principles are fundamentally different from epistemic values and principles, as prominently put in the case of the normative structure of science by Robert K. Merton (Merton, 1942/1973; see also Resnik, 2007). Besides pressuring knowledge generation processes so that regulatory requirements are met and drugs become licensable, the industry also attempted to benefit from the institutional role and reputation of the public research sector, such as by misusing references to researchers or universities for marketing purposes (Rasmussen, 2004).

Companies use references to highly reputable academic sciences as a marketing strategy. Beyond getting a drug licensed, companies must disseminate their drugs, integrate them into healthcare plans, and convince doctors to prescribe them. Doctors and healthcare policymakers alike can leverage or diminish the economic impact of any drug or treatment and are thus important recipients in establishing a competitive advantage. The logic of the clinical trial industry does not end with fulfilling the requirements of the regulators but also involves convincing the individual doctor and policymaker, who are both medical experts, about the superiority of their product. As one example, industry experts have been found to be heavily involved in the formation of medical treatment guidelines, which provide great opportunities to make their product become part of the standard intervention (Moreira, 2005). More generally, however, pharmaceutical companies developed a science-based type of marketing that makes use of data-heavy brochures and scientific papers to advertise a drug to doctors

or policymakers (Gaudillière & Thoms, 2013). Therefore, the industrial demand for favorable research does not end with licensing processes and regulatory science but also involves a multitude of soft strategies that make use of the academic sector and the scholarly communication system. Especially scientific publications have been seen as a powerful tool to construct the current state of knowledge about a certain topic.

1.4.2 (Systematic) reviews, meta-analyses, and EBM

The social powers of scientific texts can manifest themselves in the form of different document types or publication genres. While drug regulation is an illustrative example, it is usually a highly bureaucratic act that can involve piles of documents and reports if printed (see also Bazerman, 1994). Although each sheet is a part of the larger puzzle of a drug licensing process, the exact lines of its impacts might be impossible to draw. Further, expert advice, workgroups, and committees with their formal and informal negotiations play important roles as well (Jasanoff, 1990). However, scientific documents also play important roles in other forms of medical translation. Two examples will be introduced next. First, medical treatment guidelines that inform medical practice or healthcare policies. Second, systematic literature reviews that are considered as manifestations of scientific consensus and are, in turn, used to inform treatment guidelines. Thus, both have performative capabilities in the biomedical domain.

The high moral expectation towards medical practice implies that doctors should act to the best of their knowledge, which ideally means rigorously gathered and proven scientific knowledge (Castañeda, 2019; Tanenbaum, 1994). The relationship between medical science and doctoral practice became most visible during the 1990s. Under the term *Evidence-based Medicine* (EBM),[16] experts such as Gordon Guyatt defined new goals for medical education in which future generations of doctors should be taught to better incorporate scientific knowledge into their practice (Raspe, 2018; Sur & Dahm, 2011; Weßling, 2011). Soon, expert bodies systematically assessed and synthesized research results to create formal treatment guidelines that can be applied by doctors across institutions. Usually, such guidelines come in the form of checklists that provide the most relevant information and suggest a sequence of interventions that was proven of its safety and effectiveness (Knaapen, 2014; Timmermans & Berg, 2003). Because of

[16] Often, the term is related to David Sackett et al.'s 1996 article "Evidence based medicine: what it is and what it isn't" (Sackett et al., 1996). However, Heiner Raspe notes that David Sackett did not invent the term but read it in a publication by Gordon Guyatt, who developed the principles of evidence-based medicine in his educational program (Raspe, 2018).

its consideration of economic aspects of medicine and its overall emphasis on the proper execution of procedures, Evidence-based Medicine was famously criticized as representing "cookbook medicine" (Sackett et al., 1996, p. 72; see als Knaapen, 2014, p. 823; Raspe, 2018).[17]

Besides the importance of guidelines and checklists, systematic literature reviews took center stage with the emergence of evidence-based medicine. In the early 1970s, Archie Cochrane criticized the whole profession and called for more systematic evaluation of medical interventions.[18] Although he was impressed by the results of controlled experiments, his vision was broader because not all medical interventions can be evaluated by randomized controlled trials. In addition, the number of available studies soon went beyond any level comprehensible by doctors or researchers. Rather, a systematic assessment of the effectiveness of health interventions must consider different forms of evidence and study types (I. Chalmers, 1993).

Although systematic reviewing and meta-analyzing predated the emergence of evidence-based medicine by decades, these genres benefitted substantially from the new paradigm in medical practice, medical research, and healthcare policy. Especially the epistemic promise to synthesize overall conclusions on specific questions was highly appealing to the proponents of evidence-based medicine, who had to find ways to determine the evidence on which medicine had to be based. In this regard, systematic reviews also became key components for cost-benefit analyses of medical interventions, or health technology assessments (Castañeda, 2019; K. E. Smith & Stewart, 2015). The role of systematic reviews for medical decision-making will be further elaborated in Chapter 3.

Review genres are no invention of the 20th-century biomedical sector. Rather, they already accompanied the emergence of modern science and the first traces of scientific periodicals during the 18th century (Csiszar, 2018). Review articles synthesize, discuss,

[17] During this work, I will occasionally make use of the concept of *cookbook science,* which is not commonly used or defined. However, other works that analyze the standardization of scientific practice use the cookbook metaphor as well (e.g., Fujimura, 1987, p. 278).

[18] Archibald Cochrane was a British epidemiologist who famously pleaded for the evaluation of health interventions in his book "Effectiveness and Efficiency: Random Reflections on Health Services" (Cochrane, 1972/1999). During his time as a doctor in a prisoner of war camp in the Second World War, he found that many interventions by the doctors did more harm than good and that many injuries healed perfectly without any intervention (Weßling, 2011).

summarize, or reflect on the current scientific literature. They attempt to paint the picture of the current state of science, reflect on bigger trends and schools of thought, or draw the lines of scientific debate (Blümel & Schniedermann, 2020; McMahan & McFarland, 2021). As such, reviews promise their audience to get an overview of the research field, an interpretation of its dominant trends, and an authoritative suggestion of future directions. Instead of reading vast amounts of primary literature, readers may only consult the reviews, which is a much less tedious and achievable task.

However, one can review different literatures—from experimental reports over conference reports up to whole monographs. Further, authors review the literature with varying methods and rigor and also different intentions (Myers, 1991). So, while there have been many ways of reviewing the work of others and various types of review studies in the history of science, medical researchers started to criticize and problematize the review literature of their field more vocally and distinctly during the 20th century. During the postwar era, many reviews were published by eminent authors from the field. Although their experience in the field was out of question, the studies they included in their "armchair literature reviews" (Glass, 1976, p. 4) were always a subjective selection. Knowingly or implicitly, they could neglect otherwise relevant or important studies, for example, if those contradicted the expert's own views on the topic (Virgo, 1971). As a result, two different reviews on the same topic or research question could end up considering completely different sets of primary literature and end up with different or even contradictory conclusions (R. J. Light & Pillemer, 1984). In addition, the growth of the body of homogenous trial literature created the need for a wider integration and aggregation of quantitative data (see Feldman, 1971; Mulrow, 1987).

In the late 1970s, Gene Glass used the term "meta-analysis" (1976, p. 3) to describe a novel technique for pooling study results, which was later considered as the first modern meta-analysis (Hunt, 1999). In the actual publication published a year later, statistical methods were used to pool quantitative data from multiple studies into an overall effect size (M. L. Smith & Glass, 1977). They proved the efficacy of psychotherapy by combining 833 different measures from 375 studies. The variables and their units were very diverse, ranging from physiological factors such as blood pressure or palmar sweating, over psychological scales such as one's own or the therapist's appraisal of well-being, up to sociological factors such as work achievements (Hunt, 1999). But scales and measurements of these variables were abstracted as magnitudes of improvement and statistically combined. However, the diversity in the variables combined in a meta-analysis was a critically discussed issue from the beginning and remains so until today. In general, meta-analysts have to assess the level of heterogeneity of their studies and

often try to keep it as low as possible. However, opinions about the heterogeneity of included studies differ by field and the respective epistemic expectation towards knowledge syntheses. Originally, Gene Glass defended himself, saying that it is adequate to "mix apples and oranges when we are trying to generalize about fruit" (Hunt, 1999, p. 61).

The method promised to establish scientific consensus on a given topic and rule out a variety of biases and problems of traditional review genres. Often, primary studies led to contradictory results (R. Light & Pillemer, 1986), and traditional review articles were deemed unable to cope with the vast amount of knowledge or resolve these contradictions (Feldman, 1971; Mulrow, 1987). It could also detect effects that were rejected by individual studies because their samples have been too small (Type II error). On the other hand, it could rule out those results that were retrieved due to chance because the dominant statistical regime already employed significance levels of 0.05, allowing for 1 in 20 results to be due to chance (Hunt, 1999; Newcombe, 1987; see also Ioannidis, 2005).

Although meta-analysis emerged and matured particularly in social sciences, it always played only a somewhat restricted role in that field. Being a statistical technique with a rather narrow epistemic scope, it is only of limited use when qualitative data or more complex results must be taken into account too (Hammersley, 2001; Hodgkinson, 2000; Lydahl, 2019; MacLure, 2005; Pirrie, 2001). In contrast, meta-analytic techniques became very successful in biomedicine, where they could be based on a landscape of highly standardized and formalized trial experiments and could be developed further (I. Chalmers et al., 2002).[19]

Despite the claims of robustness and validity, meta-analysis can still be biased or manipulated. In this regard, especially publication bias was of major concern, because the statistical methods alone do not determine which studies are included or excluded in the meta-analysis. Authors might already have a favorable result and be tempted to dismiss primary studies that report counter evidence, or they do not assess the often-

[19] The translational history of MA from the social to the medical sciences offers some intriguing anecdotes. For example, Morton Hunt writes that after performing the first modern meta-analysis in medicine, Dr. Thomas "Chalmers remained unaware that a similar methodology was being used in other fields. When the president of the Evaluation Research Society phoned him, he had won its 1982 annual research award for his meta-analyses, Chalmers, puzzled, asked, 'What are those?' and only learned that his statistical innovation was part of a notable development taking place in a number of sciences" (1999, p. 82).

inconsistent qualities of the included studies (Egger & Smith, 1998). Thus, it is not only the trial-producing industry that fueled publication bias. The first investigations of publication bias found that researchers contribute to this effect because they lose interest in negative findings or consider them unpublishable (Dickersin, 1987). However, meta-analytical techniques alone were not sufficient to overcome the selective inclusions of primary studies. In addition, even the first meta-analyses proved to be lengthy projects with tedious information retrieval and reference hunting (Hunt, 1999). But even the published record of primary studies itself may be biased, as, for example, negative findings remain unpublished or kept secret by pharmaceutical companies. Thus, two meta-analyses on the same topic can end up at different or even conflicting conclusions, much like two narrative reviews may do (de Vrieze, 2018).

To overcome the problems of meta-analysis, scholars developed systematic review methods. These provide a methodological toolbox for searching and retrieving literature, as well as appraising the relevance and quality of the retrieved studies (Torgerson, 2003). When doing a systematic review, researchers define a research question as well as the characteristics of the literature that is required to answer the question. These characteristics are turned into a search strategy, or a database search string that can be reported and used by others to reproduce the search results. After retrieving primary studies of interest, comparable characteristics are extracted from each study, and risks of study biases are estimated (Moreira, 2007).

The systematic retrieval and appraisal of studies is a task that is independent of the aggregation of results. For example, if studies are not eligible given the predefined quality criteria, they are not considered in the review. Notably, there are systematic reviews that end after the critical appraisal, concluding that no study met the inclusion criteria and that the research question cannot be answered without improving primary research (Yaffe et al., 2012). Soon, it became the most important methodological amendment to meta-analysis (Bohlin, 2012). Relatedly, adjacent genres such as scoping reviews end with providing an overview of the size and characteristics of the literature about a given topic in order to identify research gaps (Grant & Booth, 2009).

The systematic reviewing method itself does not determine methods of aggregation in principle. Formally, the core tasks of systematic reviewing revolve more around retrieving and selecting primary literature rather than the data analyses. This is also the reason why bigger projects may involve librarians, as we will see in Chapter 7. The independence of the review method from the aggregation method also leads to the many proposals for using systematic reviews to review qualitative or mixed-method research (Atkins et al., 2012; Jenicek, 1989). However, the output of the above-

mentioned clinical trial industry has led to a rather homogenous mass of randomized controlled trials that express their results in quantitative effect sizes. Therefore, most systematic reviews result in the application of a meta-analysis of quantitative data. The combination of systematic reviews and meta-analysis also led to the high epistemic status of the genre and sits at the core of methodological advancements, driven forward by prolific experts such as Iain Chalmers, Doug Altman, and Archie Cochrane (Egger et al., 1995/2007; Oakley et al., 2005; Straus et al., 2019).

A more well-known result of the accolade of systematic reviews that are based on randomized controlled trials is the foundation of the worldwide Cochrane Collaboration, named after Archie Cochrane, during the 1990s (Bero & Rennie, 1995; I. Chalmers, 1993). Cochrane consists of multiple publicly funded centers and Cochrane groups that perform systematic reviews of trials about basic healthcare interventions. Its main mission of creating rigorous systematic reviews that can be used for clinical guidelines led to its characterization as the "manifestation of a broader 'evidence-based medicine' (EBM) movement which has arisen since 1992" (A. Wyatt, 2002, p. 23; see also Solomon, 2015; see Askheim et al., 2017 for a more critical note).

The Cochrane Collaboration has established its own way of doing a systematic review. This method is based on a large methodological framework and is considered the most comprehensive and rigorous but also the most effortful way to do a systematic review today (Cochrane Handbook, see Higgins et al., 2019). As a result, Cochrane reviews are not just a particular review study type but also have become their own publication type and even document genre, which can result in large publications with structured bibliographies and extensive supplementary material and appendices (Moher, Tetzlaff, et al., 2007; Uman, 2011). It is important to note that many understandings of proper systematic reviewing are based mainly on formal rigor and extensive documentation of details. In their wake to approach these ideals, it is especially a Cochrane Review that can find the available trials as ineligible and that comes without results, as already mentioned (see also Yaffe et al., 2012). Likewise, Cochrane reviews have been found to be too extensive and effortful to use in some fields, so that they can be undervalued in clinical decision-making (Pagliaro et al., 2010).

During its course, the Cochrane Collaboration developed and implemented a variety of innovations that made Cochrane reviews diverge more and more from narrative reviews and even other forms of systematic reviews. For example, full-text Cochrane reviews are published in the Cochrane Database of Systematic Reviews (CDSR) and occur in scientific journals only in an abbreviated form and with a prominent reference to the full version in the Cochrane database (Starr et al., 2009). This enabled Cochrane to respond to the

requirements of a maximized level of methodological rigor. For example, the most prominent of these is the vision to constantly update systematic reviews whenever new trial results are published or methodologies change (Bashir et al., 2018; Moher, Tsertsvadze, et al., 2007; Shea et al., 2006). The most ambitious approach of this endeavor is the so-called *living systematic review,* which employs an a priori commitment to updates in a predefined frequency (J. H. Elliott et al., 2017). We will encounter some examples of the interplay of different methodologies and tools in chapter 7.

The stance of systematic reviews in medicine and medical research shows how scientific texts can unfold substantial social powers. Among medical researchers and knowledge translators and doctors alike, systematic reviews have become one of the most respectable sources of medical evidence, mostly due to their quest for consensus based on a formal methodology (see also Solomon, 2015). Similar to the history of randomized controlled trials, the emergence of systematic reviews gave rise to a community that formed and established new practices, new professional roles, and even organizations such as the Cochrane Collaboration. Although the goals and trajectories of complex networks of different practices, actors, and organizations may not be reduced to the publication of randomized controlled trials or systematic reviews alone, the texts can be seen as boundary objects that connect and mediate the different strands, as explained above.

1.4.3 Research genres as powerful devices

The idea that randomized controlled trials and systematic reviews can be considered publication genres plays an important role in understanding their social role and their performative capacities within research and medical practice. Belonging to a certain genre is the first step in being accepted as an eligible contribution, for instance, by drug regulators who expect trial results or guideline makers who require systematic reviews as an evidence base.

Similar to how scientific concepts create a complex relationship between things, words, and the background knowledge of writers and readers, genres create relationships between the intention of an author, the form of a scientific text, and the effects upon its audience. Genres have a representational character that tells their audience what contents or intentions it may expect from the text. In this way, the audience can efficiently assess the text before consuming its actual content (Bazerman, 1981, 1994). Therefore, genres can be characterized like the above-introduced “literary technologies.” Genre, too, can be wielded as representational and especially persuasive

device within scholarly discourse. For example, the information that a conclusion was drawn in a systematic review rather than in a primary article can already make audiences more likely to accept it.

However, the cases of randomized controlled trials and systematic reviews have illustrated how not only the creational origin but also the performative effects of genres are highly relational and context dependent. For example, a researcher who downloads the report of a trial might expect to add new scientific facts and beliefs to their set of knowledge. In contrast, a regulatory officer who finds the very same trial report on their desk might expect that they must take certain actions. Thus, the form of a genre remains the same, while its interpretative scheme, i.e., its social life, is context dependent (Diaz-Bone, 2002; see also Bazerman, 1994; Keller, 2006). Much like how scientific concepts work in general, genres need a unique and unambiguously used name or label to unfold their performative capacities. As we will see throughout this book, the PRISMA guideline plays an important role for the (re-)definition of systematic reviews. The somewhat more superficial practice of adding the name of the study type as a suffix to their titles, not least due to the guidelines, is also very prevalent in medical research (Tate & Douglas, 2011).

One source of a genre's performative capacities is its close connection to research designs and practices. If medical researchers perform specific research tasks, such as registering a protocol, allocating patients into treatment or control groups, or performing a significance test as part of a statistical analysis, particular concepts and descriptions of these tasks occur in a report thereof (e.g., Rennie, 1995; Marks, 1997; Moreira, 2007). Consequently, similar and recurring study designs, such as the randomized controlled trials and systematic reviews we have described, mean that similar combinations of words and descriptions are published again and again. In addition, the standardization of methodological techniques and study designs often involves the creation of technical terms and nomenclature so that there is always a similar level in the codification of knowledge (Swales, 1990; Whitley, 1985/2000; Zuckerman & Merton, 1972). So, it is not only the codified and standardized information itself, but the relationships between codifications that form patterns.

Subsequently, the recurrence of codified patterns will be interpreted as textual types and literary genres (Swales, 1990). In turn, the receptive community from which such genres emerge and in which they resonate also reproduces the genres. But the production of a homogenous body of literature also shapes the expectations of the consumers. Readers and audiences expect certain characteristics when they are confronted with a text of a kind, because they have acquired background knowledge

about the relationship between document types and document characteristics (Bazerman, 1994). For example, readers expect a randomized controlled trial to report its randomization procedure or the provision of its registration number. Such expectations influence how researchers assess publications and shape the contents of reporting guidelines, as we will see throughout this book.

If genres are powerful devices, then creating instantiations of certain genres is a way of acting in this world. Thus, although a genre identity is partly predetermined by the study type it presents, its manifestation and typification also lie in the hands of the authors. Today, publishing scientific texts still requires a substantial amount of creative writing, especially when it comes to describing, interpreting, and discussing seemingly objective results. So even when some parts of scientific papers are fully determined by research techniques or equipment, for example, x-ray images, MRI scans, or tables with statistical results (Wise, 2006), they are usually still described and interpreted in the published text.[20]

In such moments of description and interpretation, authors are free to choose the words and concepts to some extent. For example, they can spin the background and introduction of a study to describe its importance and relevance. They can choose which details of the methodology they reveal, and when they present and discuss the results, they can interpret them with a certain spin to accord them more weight (McGrath et al., 2017). Therefore, authors keep a substantial control over the information and intentions of texts (see also Bazerman, 1994). Unsurprisingly, some linguists argue that documents and documentation practices create and embellish independent lifeworlds rather than approximating the truth or real-world experience (R. Day, 2021).

But authors are not free to write as they like. Rather, their control over their texts is heavily moderated by the process of scientific publishing. The comments of colleagues, the scope of the journal, and, most importantly, the evaluation by peer reviewers are important influences that can change texts substantially. Therefore, the freedoms of writing are met by a responsibility to get published firsthand. Especially the pressure to get published plays an important role in modern academia and science evaluation, as we will see in the next section. However, authors can also face the requirements and

[20] In a more recent trend, so-called “nanopublications” are stripped from textual descriptions and interpretations altogether. This promises a quicker publication of contents and some level of independence from how the original authors might interpret the data (Auer, 2018; Velterop, 2010). Instead of bridging the gaps between research practice, reports, and the interpretation of their results, nanopublications suggest further separating both.

social expectations of different actors, such as pharmaceutical companies who need a trial report that is accepted by regulators or guideline development boards who want a proper systematic review. So even if financial interests may not influence the overall study results, they can incentivize authors to add a certain spin to their conclusions (Yank et al., 2007). These requirements can create substantial pressures on the creative gap between research practice and the report thereof, as I have elaborated previously.

We have seen how medical research is driven by a multitude of public or private expectations. Research in this sector is not done for its own sake but rather serves the interests of different stakeholders, explicitly or implicitly. This can mean that scientific authors must leverage the creative gap between what happened in the laboratory and the report thereof so that the expectations of their industry sponsors are met. But the pressure goes far beyond drug licensing and scientific marketing. The moral stance of medicine and the requirement to serve the common good of society require scientific research to be purposeful and relevant. All these influences incentivize researchers to turn towards highly reputable and impactful genres such as systematic reviews. Or they may feel inclined to achieve novel, outstanding, counterintuitive, and, above all, positive results by any means necessary. But research is a risky and uncertain enterprise so that results are neither reliably positive nor meaningful at all. Thus, in its worst case, the social pressure on medical research can get them engaged in over-interpreting, exaggerating, or even faking their research results.

The exact extent to which funders and industries shape the writing of science is not quantifiable. But it reached a level where it is perceived as a systemic problem that leads to growing concerns among medical experts. Many of the problematizations of the biomedical literature revolve especially around the influence of the pharmaceutical industry. Since medical experts became concerned about the funding effect in research during the 1980s (e.g., Davidson, 1986), metascientists collected more and more evidence about its quantities and implications that suggest a worsening of the situation in recent decades (Ivanov et al., 2017).

The publishing style of the industry can involve several behaviors that seem problematic from an epistemically honest and more scientific standpoint. In general, modern pharmaceutical companies manage complete communication pipelines, from the planning and conduct of the study to the science-based marketing of the final drug (Cosgrove et al., 2016; Sismondo, 2009).[21] Traditionally, metascientists have criticized

[21] The unscientific practices of industry-sponsored research extend beyond publishing activities. Even trial designs can be managed strategically. Treatments can be

the selective outcomes reporting and file-drawer problems that come with industry-funded studies. For example, companies withhold study reports completely if the drug was not effective or caused too many adverse reactions (I. Chalmers, 2006; Melander et al., 2003). But industry-managed studies can also make marginal improvements look novel, exaggerate their effectiveness, or downplay the risks, even if they were made public (Als-Nielsen et al., 2003; Cosgrove et al., 2016). The industry's influence on trial reports is so profound that some argue for a stronger separation of the two systems (Sismondo, 2008).

Over the history of medical research during the 20th century, the medical research domain and the medical industries have established tight and almost inseparable bonds. Almost every expert from any medical specialty is in close contact with industrial actors, so that separating research from industry would imply "turning their world … inside-out" (Sismondo, 2008, p. 1913). In addition, biomedicine and healthcare are costly endeavors. Thus, banning private interests from funding medical research altogether seems impossible.[22] But reporting standards for transparent and thorough reporting have also been developed for industry-based research (Wager et al., 2003). Further, there is a broader consensus in medicine that authors should report on who funded their project which they often fail to do or omit intentionally because they fear suspicions of bias and repercussions by their peers (Bekelman et al., 2003). For example, the disclosure of funding sources and potential conflicts of interest has become a necessary item in guidelines for randomized trials (Schulz et al., 2010), and systematic reviews (Liberati et al., 2009e), which we will see later in this book.

This section has viewed the history of medical research from the perspective of scientific publishing. It has shown how research practices and their outcomes, scientific publications, are socially contextualized, and how the social expectations placed on research procedures can influence their outcomes, not always for the better. However,

---

intentionally performed incorrectly so that the focal treatment appears more effective or safer (I. Chalmers, 2006).

[22] There are nuances and different perspectives when it comes to completely banning privately funded research. In an interview, John Ioannidis said that nobody who is interested in any form of reliable information on the health effects of smoking should even bother to read a study that was funded by the tobacco industry (Ioannidis, LaBlanc, 2024). This opinion is based on the decades-long influence of the tobacco industry on the evidence base about smoking. Rather than blatantly opposing results from public research, the industry-sponsored scientists leveraged the creative gap and common study limitations to cast doubt and uncertainty on the overall research field (Michaels, 2008).

we have also seen how this configuration characterized even the earliest versions of modern medicine. Thus, although the interrelation of public, private, and epistemic interests has indeed been troublesome throughout the 20th century, it does not paint the whole picture of the more recent transparency crisis that led to the birth of PRISMA and other reporting guidelines.

To acquire the last building block, the next section will draw our attention towards how some traditional scientific institutions have been transformed dramatically by the growth of modern science and the digitalization of scholarly communication.

1.5 The wealth of science and knowledge

1.5.1 The explosion of medical information

The second half of the 20th century witnessed a massive growth of the global biomedical research sector. This growth was comprised of the foundation of new research organizations, novel collaborations, and more researchers who got engaged with novel topics and methods. With the end of WWII, the US science adviser Vannevar Bush famously called the advancement of science and technology an *endless frontier,* and the United States and the Soviet Union were keen to compete on these battlegrounds (Pielke, 2010). The development of atomic bombs and the space race not only became famous arenas of national competition between the two superpowers (Bowler & Morus, 2005; Krauhs et al., 1989). Especially in the western world, they also initiated more intense collaborations between academic and non-academic spheres, such as the military or the private sector (Etzkowitz, 2003). But these developments also came with huge public investments into the academic scientific sector. In the United States, investments in the medical research sector increased from 161 million dollars in 1950 to 2.5 billion dollars only 18 years later (Junod, 2008; see also Pielke, 2007).

When the public science sector got a bigger slice of western societies' public budgets, expectations of financial accountability slowly trickled into science policies and organizational frameworks. The expectation that science would contribute to the common good and economic growth led to the concession of greater resources. This fueled the emergence of a science management culture in which research organizations were understood and governed as if they were businesses (Hallonsten, 2021; Münch, 2014). But in contrast to the private industrial sector, scientific growth and the advancement of human knowledge cannot be expressed in revenues, profits, or gross domestic products. Preclinical drug discovery wouldn’t be possible without the periodic table of elements, yet the idea of assessing its economic value sounds odd.

However, many clinical research projects or clinical trials to prove safety and efficacy are situated much closer to applicatory contexts and potential financial exploitation, as we have seen previously. But despite the many ways medical research can impact industry or society, translational research pipelines may take years or decades until they result in applicable treatments and measurable returns on investments. Further, even research-based patents can remain unused for an indefinite period (Rosenberg, 1974; Stephan, 2015; see also Morris et al., 2011; Sampat & Ziedonis, 2004). In addition, the idea of a successful knowledge translation is always a retrospective concept that is meaningful only when outcomes of research and innovation become visible, and their causes and effects are revealed. But much of the publicly funded research is basic research so that usefulness or applicatory horizons are not foreseeable, risky, and uncertain at best. Yet at the same time, governments not only require financial accountability of the scientific sector but also find themselves in a global competition in scientific excellence (Hallonsten, 2021). To win this competition, they attempt to create the social and political contexts in which science can thrive. This often boils down to the identification and incentivization of the most successful approaches towards research or the most productive and creative scientific workforce (Langfeldt et al., 2019; S. Moore et al., 2017; Vessuri et al., 2014).

But the governance and steering of national sectors requires goals. The strive for excellence, fruitfulness, or success in science has been described as the search for "the Holy Grail of science policy" (Van Leeuwen et al., 2003, p. 257). But in practice, the goals and objectives defined in science policies were far less transcendental. Any complex, contingent, or dynamic notion of scientific quality and scientific progress had to be translated into distinct criteria that can be facilitated, achieved, and governed by research organizations, project managers, and nationwide administrations (Münch, 2014).

But what could such criteria look like? Stripped from the ability to measure any translational impact of research directly, the public science sector had to find other, more immediate ways to evaluate research projects, organizations, or funding lines to set further paths and distribute additional resources (Partha & David, 1994; Kitcher, 2002). Subsequently, public science funders, publishers, and the wider biomedical research community conjointly established and fostered an evaluatory regime to decide which of the processes and products of public research can be considered most relevant (Langfeldt et al., 2019). Scientific texts and especially academic journal publications have taken center stage in this evaluatory regime, especially in the life sciences (Hornbostel et al., 2008). However, the necessity of evaluating and ranking individual research outputs was not solely a result of national competition and administrative customs.

Moreover, the sheer number of published outputs rendered any individual or qualitative assessment unfeasible.

The second half of the 20th century saw a massive growth in biomedical publications, a trend that grew into a sever information problem and which holds until today. Common tropes are that science in general and biomedical research in particular witnessed a "knowledge explosion" (Bazerman, 1983, p. 173; see also Bawden & Robinson, 2020; Hunt, 1999) or suffered from an "information overload" (M. Clarke, 2004, p. 100). Some scholars argued that published scientific information grew exponentially, which means that there was more seemingly relevant information than any single human reader could comprehend. Consequently, researchers and medical experts are unable to possess all the knowledge that is associated with their craft (de Solla Price, 1963; Kronick, 1976; Tenopir et al., 2011; Woelert, 2013). For example, even reading every publication on a comparatively small medical subfield such as echocardiography in 2010 (the science of ultrasound heart examination) would take more than eleven years (Fraser & Dunstan, 2010). Subsequently, experts provocatively asked, "Seventy-Five Trials and Eleven Systematic Reviews a Day: How Will We Ever Keep Up?" (Bastian et al., 2010).

This situation also worsened with the advent of digital publishing, which radically reduced the costs to produce and distribute publications, aligning with the general "rummage sale of information on the World Wide Web" (Bowker & Star, 1999, p. 7; see also Tenopir et al., 2011). In principle, electronic publishing promised to reduce publishing delays, publishing costs, and startup costs for new outlets, which paved the way for independent journals. In addition, digital library catalogues and preprint servers enabled the mass distribution of scholarly content and paved the way for a radical transformation of the scientific information market (Guédon, 2001; Woelert, 2013). Against this background, not only national science policies, funders, and research bureaucrats started to frame the assessment and evaluation of science as a technical problem. Rather, the researchers themselves were in dire need of tools and procedures to cope with the ever-growing amount of available information.

1.5.2 Authors and rewards in biomedical sciences

But why did the number of publications grow so rapidly, even exponentially? Essentially, the previously mentioned investments in the public science system resulted in a greater number of researchers, who wrote more publications. Since the dawn of modern science, publications have been testimonials for individual scientific achievements because authorship establishes a formal link between the outcome of a project and the name of one or more individuals. Thus, scientific reports do not just transport

information but also associate the topics, ideas, or findings with an inventor, originator, or investigator (Hesselmann et al., 2021). In this way, researchers gain attention and recognition in the scientific community and in wider society. The importance of authorship and the bond between author and idea have always been an integral part of formal scholarly communication (Csiszar, 2018; Shapin & Schaffer, 1985/2011). Moreover, this association has become increasingly significant today, despite more recent perspectives emphasizing the role of scientific teams in knowledge production.

Until today, being an author on a scientific publication means to have participated in a successful research project. Especially the publicly visible connection between the project that is reported and the name of the researcher provides a mark of effort, achievement, and intellectual property on a product that can be copied and reused at no cost once published. This feature is so essential that Robert K. Merton turned it into a part of the ethos of science: "The products of competition are communized, and esteem accrues to the producer" (Merton, 1973, p. 274). Or, as Paula Stephan wrote it in other words and decades later, “by giving it away, scientists make the research findings their own” (2015, p. 7). Both quotes illustrate the relationship between the publicness of scientific products and the authorship-based reward system.[23] In addition, attaching a name to a publication and its set of ideas also provides information about the intentions of the texts. As such, texts authored by researchers are usually considered as scientific texts that intend to inform or educate, rather than entertain their audiences, or display the witty and artful writing style of their authors (Csiszar, 2018; see also Bazerman, 1994). However, this intention is also shaped by the organizational affiliation and the publication venue.

Authorship also establishes a connection between a research organization and the scientific product (Di Leo, 2003; Stephan, 2015). Often, these affiliations not only signify

[23] This is an economics-based depiction of scientific knowledge as a product that has high production costs but, being public and easily reusable, provides only little competitive advantage. Further, ideas are one-time goods that, once made public, provide no additional value when repeated or reproduced. In a sense, the importance of novelty as publishing criterion, as well as the lacking incentive for scientists to do replication studies, second this characterization (Nosek et al., 2013). However, this depiction of knowledge as a product severely underestimates the costs of digesting, applying, and contextualizing scientific knowledge. Illustratively, some traditional studies in the economics of science found that companies need research staff not only to produce their own knowledge but also to *get access* to public knowledge in the sense that only trained scientists can comprehend and process research papers to the benefit of the company (W. M. Cohen et al., 2002).

scholarliness and scientificity. Rather, they are used by researchers and the general public alike as shortcuts for excellence and quality so that the reputation alone makes people believe the results (Origgi, 2017; Rosenbaum et al., 2008; see also Giddens, 1997). While those shortcuts and habits are fallible, they are not generally unfounded. Eminent institutions and labs are expected to offer more resources and training to their researchers. Due to their visibility, they are thought to have a greater responsibility to remain credible and have a reputation to lose. Publications coming out of those institutions are assumed to be more robust and of higher quality, which is backed up by investigations that show how elite institutions have lower rates of retracted articles or fewer issues with funding bias (Jørgensen et al., 2006; Lievore et al., 2021). Therefore, not only authors but also institutions can have a scientific reputation. For example, the Cochrane Collaboration introduced above built up a reputation for creating the most extensive and rigorous systematic reviews available. Even though readers might not be aware of the complete methodological processes behind a Cochrane Review, they can acquire an "entitled understanding" (S. Wyatt et al., 2017, p. 70) of its epistemic superiority. Therefore, Cochrane Reviews are highly regarded among medical experts and are often understood as final answers to certain research questions.

Since academic recognition is so crucial for the evaluation of science and distribution of resources, proper attribution of authorship and affiliation has become part of the jurisdictional activity of the science system. Authorship has become a form of reputational currency in the science system, and researchers sometimes end up in conflicts over it (Hesselmann et al., 2021). Any willful misrecognition of authorship has been deemed as unethical and can even lead to the retraction of the article (Bozeman & Youtie, 2016; Zhang & Grieneisen, 2013). A crucial role in controlling authorship and managing authorship conflicts is played by academic journals.

Throughout the history of science, authorship and reputation have always been entangled. Traditionally, many scientists communicated via books or exchanged letters. But with the rise of the learned societies such as the Royal Society during the 17th century, more and more scientific periodicals emerged, and journal publishing slowly proliferated into the backbone of formal communication in the natural sciences (de Solla Price, 1963). In contrast to books, journal papers were shorter and cheaper to produce, thus a more doable and compelling task for authors. By writing journal papers instead of books, they could efficiently communicate their results and ideas in an open and publicly visible forum, which made more and more scholars take this route. Thus, many traditional accounts consider the proliferation of journal publishing as a central dimension in the maturing of modern science (Bazerman, 1988; Csiszar, 2018; R. A. Day, 1989; de Solla Price, 1963; Golinski, 1998; Guédon, 2001; Kronick, 1976).

1.5.3 The gatekeepers of science

Academic journals are arenas where authorship and recognition become manifested. Especially those conceptions and conventions of authorship that were employed by academic journals prevailed, due to the importance of these publication venues. Subsequently, the definition and discussion of dimensions such as authorship and credit became part of a professional identity of science journal editors (Csiszar, 2018). Today, the International Committee of Medical Journal Editors (ICMJE) can be seen as the culmination point of this professionalization. With a focus on scientific publishing in the medical sciences, the committee develops definitions and sets standards. Publishers often adopt these standards and implement them across their entire journal portfolio—with some deviations permitted—which makes the medical publishing culture highly influential in science publishing in general (Hesselmann et al., 2021).

Journals represent intellectual strands and scholarly schools of thought, broader theories or methods, or sets of classification and nomenclature (Kronick, 1976). They do so by selecting what content will be published in what form. They provide the arenas in which new concepts and concept-knowledge links get established and become standardized within a scientific community. Taken together with their function in managing and controlling who can publish, journals largely define how scholars communicate and interact in formal arenas (Csiszar, 2018). For their central function in the science system, they have traditionally been described as the "gatekeepers of science" (Crane, 1967, p. 195).[24] Thus, important social and intellectual decisions are made in the editorial office of an academic journal.

Journal editors filter submissions not only based on topic and fit to the journals' scope but also with regard to textual forms and quality of writing. Further, they set the rules and organize peer review in which the submission is handed to external reviewers who appraise content and quality. Only after successful peer review are submitted manuscripts published with the names of their authors and organizations attached. Hence, authorship and scientific reward do not follow automatically from the production of a text but are subject to evaluation and assessment. Most importantly, peer review has become the focal point of quality control in almost all scientific disciplines and especially in biomedicine (Abramo & D'Angelo, 2011; Kronick, 1976;

[24] In contrast to the gatekeeping practices described here, Diana Crane analyzed how non-scientific factors influence editorial decision making and found that decisions may be biased towards submissions where the social characteristics of the authors match those of the journal editors.

Reinhart, 2012; see also Kaltenbrunner et al., 2021 for a discussion regarding STS). Subsequently, academic journals have become the representatives of the peer review system, so that their role in certifying the academic record is emphasized when new developments and technologies endanger their function as publishing venues (e.g., Priem, 2013).

But what does gatekeeping mean in this context? Academic journals decide which texts are published and which are not and only published texts become part of the publicly visible record and the stock of knowledge. Therefore, there are two ways in which journals shape scientific publishing on a global scale.

First, in a direct way, by accepting or rejecting manuscripts, they tell submitting authors what types, formats, and styles of writing they should choose. Journals often provide submission guidelines or manuscript guidelines to predefine such aspects and provide a transparent criterion for rejecting manuscripts. On a broader level, suggest what types of studies, research practices, or scientific content are eligible, for example, by focusing on specific study types such as the journal *Systematic Reviews* (Moher et al., 2012) or *The Journal of Open Source Software* (A. Smith, 2016). But on a more fine-grained level, journals have an important influence on the textual types and genres of the biomedical research literature. An important example is the introduction of standardized article formats, such as the "introduction, methods, results, and discussion" (IMRaD) format (R. A. Day, 1989; Sollaci & Pereira, 2004; Swales, 1990). But journals also set rules for proper statistical analyses and visualizations (Gardner et al., 1986; Stratton & Neil, 2005). Traditionally, journals also rejected manuscripts that were too long because of the higher printing and distribution costs when electronic publishing did not yet exist (Csiszar, 2018). Likewise, journals also play an important role in the dissemination and implementation of reporting guidelines in medical research, as we will see later.

Second, journals also indirectly shape scientific publishing practices. Successfully published papers and the underlying study designs represent types of research that are acceptable to journal editors and peer reviewers, i.e., the scientific community. Subsequently, such forms and types become templates for projects. Early career researchers and aspiring authors draw inspiration from published works and structure their projects accordingly. Even though the journal and its editors don't require it, templating other papers and following the dominant style can be misinterpreted as a requirement to get published. As a result, individual authors may co-produce the textual homogeneity of a field's literature the same way as its gatekeepers do.

But the homogenization of content and form did not mean the end of diversity and variety in doing and writing research. Rather, varieties and heterogeneities became standardized in new forms and genres of publications and also new types of journals. Content-wise, whole journals could focus on a particular theory or school of thought, so even at the beginning of modern science, “pitting theory against theory required pitting journal against journal” (Csiszar, 2018, p. 41). Instead of having a diverse set of contributions in each journal, academia grew a diverse set of journals.

Not surprisingly, the journal landscape further differentiated with the emergence of new theories, methods, or research disciplines. But even with regard to seemingly less intellectual aspects of science, such as the genres and forms of texts, specific types of journals emerged. Most notably, review journals emerged in the wake of the standardization of the primary research article. In this regard, the *Journal des Scavans*, one of the first modern journals, can be considered a review journal. However, a more sizeable number of these journals did not emerge before the 18th century (Csiszar, 2018). Although historical analyses found a “bewildering array of formats” among these outlets (Kronick, 1976, p. 20), prominent editors and publishers of these times had already started to categorize journals into primary and secondary, substantive and historical-bibliographic, or giving and taking journals, rather than categorizing individual articles (Kronick, 1976).[25]

Review journals and literature reviews in general also play an important role with regard to the growth of biomedical literature. Traditional narrative reviews provide authoritative interpretations of the current state of the field by eminent scholars (McMahan & McFarland, 2021; Restrepo Forero, 2003). Systematic reviews compress hundreds of pages of primary research into a single paper and promise easy and achievable access to a whole topic, thus providing a tool to "cope with the information explosion" (Hunt, 1999, p. 82). In that sense, systematic reviews and meta-analyses became appropriate tools, not only to overcome publication bias and subjective study selections as described above, but also to enhance the genre’s capabilities in aggregating knowledge and reduce the reading efforts of any single reader, as Gene Glass illustratively wrote about his meta-analysis:

[25] A journal might be categorized based on its individual contributions. But once the journal is categorized, its contributions may inherit the journal’s category in a non-technical sense. Readers might say the BMJ journal “Systematic Reviews” publishes only systematic review articles, which suggests the existence of such an article type. All the nuance and heterogeneity of the individual texts in that journal might get buried in the journal categorization (see Bowker & Star, 1999 for a general account on classification).

> *The armchair literature review in which one cites a couple dozen studies from the obvious journals can't do justice to the voluminous literature of educational research that we now confront.*
>
> (Glass, 1976, p. 4)

In general, reviews serve important intellectual functions for the scientific community or are crucial for the translation of medical knowledge, as explained above (see also Bensman, 2007). Review journals that co-define the format, content, and genre of reviews, as well as publish and curate this type of literature, are likewise important cornerstones in biomedicine's information machinery. Review journals are platforms for reviews, commentaries, and summaries of original research. They promise researchers and practitioners to keep an overview of the vast literature in medical research. For this reason, forms of reviewing and review journals have accompanied scientific publishing since its beginnings (Kronick, 1976; see also Garfield, 1987).

So far, we have seen how academic journals manage and regulate scientific knowledge. First, they provide communicative platforms and filter academic contributions by content and form, though not in a strict sense, as new trends and topics in both—content and form—may spark new journals. Second, they organize peer review, thereby manifesting and executing procedures of scientific quality assessment. Third, journals set and regulate the rules of formal scholarly communication, e.g., by defining authorship and assigning it to individual results or ideas. Taken together, academic journals are fundamental building blocks of the scholarly communication system and suppose a science-internal perspective on the value and relevance of medical knowledge. However, journals did not offer any counterweight against the explosive growth of the biomedical research sector. Rather, the journal publishing system and its capacities of absorbing and managing information were overwhelmed by the growing outputs, as Robert Day put it in the late 1980s.

> *Money meant research, and research meant papers. And our journals were virtually overwhelmed by manuscripts pouring out of our research laboratories.*
>
> *(R. A. Day, 1989, p. 18)*

Instead of gatekeeping and regulating the massive flow of scientific information, journals published more issues and articles, or new journals were founded (Mabe,

2003). More issues and journals meant more publishing opportunities for researchers. This made room for new theories and methodologies that can be discussed and paved the way for new trends and scientific fields. But most importantly, publishing opportunities implied career opportunities because researchers could participate in scholarly communication and be recognized and rewarded for their contributions (Biagioli & Lippman, 2020; de Solla Price, 1963, 1961/1975; Tenopir & King, 2014). For a prolonged time throughout the 20th century, the number of academic journals was correlated to the inflow of personnel and the public investments in science (Tenopir & King, 2014).[26]

The journal system went with the tide, and the academic libraries were next in line to deal with information overload. Librarians and information specialists were faced with judging the value and relevance of scientific publications and turning the classification and curation of scientific information back into an achievable task (Gould, 1998; see also Harris, 2005). But rather than accepting or rejecting individual publications, they focused on whole journals by including them in library catalogs. Given the growth of the journal landscape, selecting journals became a question of library budgets. Science publishing has long been established as a professional and revenue-driven industry with an interest to constantly increase prices, while the budgets of libraries and universities to license literature were limited (Csiszar, 2018; Garfield, 2006; Guédon, 2001; see also Larivière et al., 2015). Subsequently, librarians faced a *serials crisis* because libraries were unable to license and curate all existing scientific periodicals and publications (McGuigan & Russell, 2008). Moreover, they must select what they license and order from the publishers. Like research administrators, libraries had to find ways to determine and rank the value and relevance of scientific contributions.

Journal catalogues and journal lists became essential tools for selecting journals to include on library bookshelves. Lists of valuable or eligible periodicals accompanied the journal landscape since its early days. For example, the Royal Society's Catalogue of Scientific Papers from the 19th century employed strict eligibility criteria, providing some form of authority and valuation for its 1400 included journals (Csiszar, 2018). But lists of legitimate publication venues still play an important role in academia today. To

[26] Carol Tenopir and Donald W. King (2014) discuss how the growth in journals decreased to the end of the century, albeit electronic publishing made it easier than before to start and maintain a journal. However, digitalization also meant completely new publishing infrastructures and communication platforms that do not increase the number of journals. In addition, the number of authors per paper increased too, and can differ substantially as we will see in chapter 6.

sort libraries and guide spending, librarians employ approval lists such as *core collections*, *the Union List of Serials*, or *the World List of Scientific Periodicals* (Corby, 2003; Guédon, 2001; Kronick, 1976). On the opposite end of the spectrum, disapproval lists have also emerged recently. Most notably, Beall's list of predatory journals features journals that have been accused of no or insufficient forms of quality assurance, such as peer review. While the definitions and inclusion criteria of such lists are controversial, the scientific community tends to be suspicious of publications from journals included in them (Richtig et al., 2024; Strinzel et al., 2019).[27] Thus, approval or disapproval lists provide classifications that are used as shorthand to scientific qualities and characteristics (Bowker & Star, 1999). However, lists can still be long and unsuitable for filtering. Thus, libraries started to rank individual journals by the use of quantitative publication metrics.

1.5.4 Measuring biomedical science

Publication metrics are a technical and quantifiable interpretation of intellectual value and relevance. The original idea to develop journal metrics, such as the famous Journal Impact Factor (JIF), involved ranking all existing journals and licensing only those that achieved the highest ranks. This metric is based on the number of papers published in a journal and the number of citations these papers accumulated, restricted to a certain timespan such as two, three, or five years (Archambault & Larivière, 2009; Garfield, 1977b, 1998, 2006). With the Journal Impact Factor, citation-based assessments of scientific literature became common.

Citations promised to be a mathematically strict and seemingly objective method in determining the role of a publication or a whole journal for a scholarly community. References to former text as a phenomenon can be found in almost every scientific discipline, because they are based on the previous findings, discuss or refute them, or use them to underpin their own arguments. Sociologist of science Robert K. Merton originally described citation as a fundamental mechanism of the normative structure of science, assuming that scientists credit each other's work by citation. More citations then mean that the paper was more influential and important than other publications in the field (Aksnes et al., 2019). However, this normative theory of citation has been

[27] An illustrative culmination of the debate on journal lists can be found in a controversy over a bibliometric analysis published in Scientometrics, which naively used Beall's list as a proxy for predatory publishing—a disregarded practice in which quality assurance, mainly peer review, is sold but not performed. The issue became especially severe when Springer Nature, which had stakes in some of the journals on the list, intervened to have the article retracted (Abramo et al., 2022; Oransky, 2021).

contested. In fact, there is a multitude of social, cultural, and contextual factors. These factors make the idea of citations as derivatives of any particular social dimension in science implausible but rather suggest a rich citation culture (Wouters, 1999). In addition, authors may use citations strategically to establish links and alignments within their community (Woelert, 2013; Wouters, 1999, 2020) or persuade readers of their texts, for instance, "bringing friends in" (Latour, 1987, p. 31). A "social constructivist theory on citations" (Tahamtan & Bornmann, 2022, p. 2) remains open to a multitude of motivations and reasons that can influence citation behavior simultaneously and which may even conflict with each other. Despite original calls for more sociological or STS-inspired research, the scientometric field still lacks a universal theory for citation. However, these accounts have uncovered different causes and motivations for citations and do not back up the idea of a single, uniform theory of citation (Wouters, 1999).

However, citations somehow still suggest uniformity and homogeneity to many scholars and public observers of science. The very ubiquity of citation as a phenomenon and the seeming formal isomorphism it has achieved throughout the history of science—citations look very similar across different textual genres and fields of science—may also fuel impressions of a homogenous practice. However, disciplinary scientists and research managers, who often have a scientific background themselves, may just impose their own citation behavior on their colleagues, scientific peers, and other scientific fields. It has been argued that such biases become averaged out if citation and publication metrics are aggregated on organizational or national levels (Aksnes et al., 2019). Against this background, citation metrics have become tied to the promise for an exogeneous form of science governance that circumvents questions of epistemic culture and disciplinary idiosyncrasies and which can be performed by outsiders, such as information experts, librarians, and research evaluators (Whitley, 2011).

Subsequently, publication and citation metrics have been used as tools and judgement devices for evaluating the productivity and impact of countries, organizations, and even individual researchers. Here, the interpretation of citation impact as a quality marker was another reduction that sparked substantial criticism. Similar to how citations have been interpreted as aggregations of different citation behaviors, the highly complex and diverse notion of research quality has been reduced to citation impact, which is, in turn, accessible by technique rather than being bound to expert-based judgements (Aksnes et al., 2019; Aksnes & Rip, 2009; Weingart, 2005; Wouters, 2020; see also Hojat et al., 2003). As a consequence, metrics-based evaluation of science is often contrasted and discussed with peer review, which is the traditional form of peer judgement and quality assurance in modern science (Reinhart, 2012; Wouters, 2020). Even though systematic investigations revealed correlations between peer judgement and impact metrics, the

accolade of citation impact as a representative of scientific quality remains controversial and contested (see also Garfield, 1998; Woelert, 2013).[28] Moreover, the equation of quality with impact has become a central misunderstanding in modern research that has the potential to distort how we envision and conceptualize scientific progress, as we will see below.

The promise of scientometric indicators to provide a seemingly objective and comparative representation of intellectual relevance was also driven by the emergence of bibliometrics and quantitative science studies as research fields or expert communities. The history of scientometrics and metrics-based research evaluation is closely tied to the emergence of information science as a profession and academic discipline, because the former can be described as an expansion of the latter (Biagioli & Lippman, 2020). The growth of public and private bodies during the postwar era called for new techniques in indexing, archiving, and retrieving technical documents. The emerging profession of information science sold secondary bibliographic services such as automatic indexing and translation services to universities, the chemical industry, and the US military. The Cold War period, i.e., the "golden age of information science in the US" (Burke, 2007, p. 13), provided the stability in markets, sponsors, and tasks that allowed for information science to mature as an academic field. However, the profession has ever since been concerned with processing scientific papers and scientific information, which grew substantially after World War II, as explained above. Therefore, even though there have been calculations and assessments of scientific literature that bear the intellectual moment of modern scientometrics prior to the 20th century,[29] it was the maturing of the information science and the technological advancement of that time that allowed for more systematic and large-scale approaches (Wouters, 1999). Not surprisingly, Eugene Garfield's indexing project not only benefited from the information explosion and the need for professional information retrieval but also from his automatic and computer-aided approach (Csiszar, 2020; Wouters, 1999).

The indexing of scientific papers and abstracting of their information allowed taking a macroeconomic perspective at scientific production. During the 1950s and 60s, science historian Derek de Solla Price started to use publication data to quantify the

[28] A more discussion about the technical aspects and the epistemic limitations of bibliometric analysis can be found in chapter 2.

[29] We have seen some examples from the 18th century (Csiszar, 2018; Hornbostel et al., 2008; e.g., "Zur Statistik Der Naturwissenschaftlichen Literatur," 1870). In his history of scientometrics, Yves Gingras (2016) explains how some libraries compiled small citation analyses of the chemical literature during the 1920s and 1930s to base their licensing decisions on data.

development of modern science (Storer, 1974; S. Wyatt et al., 2017). But soon the attention shifted towards citation data, which allowed for a more advanced type of information retrieval, e.g., beyond keyword-based retrieval. Subsequently, Garfield published and marketed the Science Citation Index (SCI) since the 1960s. However, the academic community was more interested in using the citation data to study scholarly communication itself (Wouters, 1999). Subsequently, Garfield started to market the SCI to sociologists and historians of science (Csiszar, 2020).

The use of citation data by sociologists of science amplified the entanglement of scientometrics and metrics-based research evaluation. Scholars such as Vasily Nalimov or Robert K. Merton studied publication networks and analyzed the structure of scientific communities and their reward systems (Kulczycki, 2024; Wouters, 1999). But especially Merton's efforts were fueled by his mission to develop a set of science indicators with which the National Science Foundation could report and justify its investments to the US Senate (Csiszar, 2020). Subsequently, Garfield became more open about the potential of citation data to improve the efficiency of public spending during the 1970s. But these ideas sparked controversy among academics because they suggested an entirely new public perspective on the roles and goals of public science. Likewise, Merton was aware of these issues and uttered criticisms at Garfield's Institute for Scientific Information (ISI), where he was a board member, albeit advocating Garfield's work in public (Csiszar, 2020; Storer, 1974; S. Wyatt et al., 2017).

Over the course of the 20th century, bibliometrics became an academic field and individual profession. Despite the entanglement of measurement and valuation, the field encompassed diverse strands throughout its evolution. On the one hand, there were scholars predominantly interested in the social study of scientific communities and the role of scholarly communication (Kulczycki, 2024). On the other hand, a rising caste of bibliometric experts and research evaluators turned towards the quantification and metrification of scholarly communication for the purpose of assessment and evaluation (Hornbostel et al., 2008).[30] This dual role provided academic legitimation but also secured the necessary funding to further develop methods, create data sources, and train future bibliometricians. In that sense, the field of bibliometrics gave rise to new

[30] We might speak of a third group: information scientists who are concerned with the technical specification of research information, databases and data sources, or the validation of bibliometric methods and indicator formulas. Although this strand might be as disinterested in social structure of science as it is in science evaluation, it is nevertheless motivated by the important role and value of bibliometric data in modern academia (Leydesdorff et al., 2016).

research centers and offered novel career pathways within academia (Petersohn & Heinze, 2018).

However, academic bibliometricians also engaged in error analysis and validation studies of bibliometric data, methods, and indicators (Aksnes et al., 2019). Subsequently, the field matured in terms of available data sources and quality thereof, its methodological toolbox, and sociological knowledge about scholarly communication and the behavior of researchers. In this way, some bibliometricians developed and nurtured a more critical attitude towards naïve understandings of science indicators and metrics-based research assessments, or overly confident perspectives on their value and meaning.[31] This attitude is not only based on the many errors and problems that render bibliometrics as doable for science studies but problematic for evaluation. Rather, the very idea that the bibliometric profession can be understood in separation from the politics and practices of valuation, i.e., the potential for an exogenous form of science governance, is highly controversial and has been called "the failure of a paradigm" (Wouters, 2018, p. 534; see also Archambault & Larivière, 2009).

The pathways and successes of information science and bibliometrics paved the way for a wider shift towards metrics-based science evaluations. This shift introduced new goals and conceptions of relevance to the academic sector, which can be efficiently assessed by (disciplinary) outsiders, such as librarians, business analysts, and research managers (Burke, 2007; Langfeldt et al., 2019; Münch, 2014; Weingart, 2005; Willmott, 2011). Publication and citation metrics were soon used for the evaluation of research activities, e.g., from institutions over projects to individuals (Geuna & Martin, 2003; Weingart, 2005). In general, funding and hiring schemes of science policies strive for excellence, as explained above. The introduction of research metrics into national funding schemes and targeted federal interventions is intended to incentivize and encourage scientists to become more productive, innovative, and excellent, thereby contributing to national advantage and societal good (Langfeldt et al., 2019; S. Moore et al., 2017). The quest for performance enhancement implies a maximization of the return on investment, which means defining (achievable) outcome goals, incentivizing researchers and organizations working towards these goals, and providing additional rewards to those who have

[31] There is some nuance to the definition of a naïve perspective of bibliometrics. Though, one is bound to the idea that scientific progress or a similar important dimension of science can be assessed by metrics in at all. But this idea does not allow for alternatives and contingencies if portrayed at the level of individual researchers. Thus, while the validity and use of person-level metrics is often rejected, bibliometric macro analyses at the national or organizational level are considered as less critical (Moed et al., 2004; Van Leeuwen et al., 2003).

performed best (de Rijcke et al., 2016; S. Moore et al., 2017; Münch, 2014; Wouters, 2017). Against this background, quantitative science indicators provided a technical and seemingly objective measure of the relevance of scientific work.

Evaluatory uses of bibliometrics often get contrasted against more or less system-wide peer assessments. Common dimensions in these comparisons are the costs and efforts required for comprehensive and system-wide peer assessments. For example, Giovanni Abramo and Ciriaco D' Angelo (2011) argue that even if bibliometric data is biased and inaccurate, it is no less reliable than peer review but substantially cheaper on a macro level. The most prominent example is the British Research Excellence Framework (REF), which combines evaluatory bibliometrics and peer assessments in a massive and nationwide evaluation framework. But the inclusion of formal indicators on the national level varies substantially by country so that we can find systems that put different emphasis on either peer review, publication output, or citation impact (Abramo & D'Angelo, 2011; Hammarfelt & De Rijcke, 2015; Mathies et al., 2020). However, it must be noted that most national science policy systems that consider science indicators focus on funding success indicators rather than citation-based indicators (de Rijcke et al., 2016; Krempkow & Schulz, 2012).

Together librarians, bibliometricians, and research evaluators have developed and established measures to cope with the information explosion. An important factor that accelerated both the growth of scientific information and the technical countermeasures has been the advent of digital publishing. With the internet era, publishing ceased to imply printing and shipping out stacks of paper, which drastically reduced the publishing costs and provided wider and easier access to producers and consumers of scientific texts (Csiszar, 2018). Thus, while I have focused on the motivations that drove librarians, bibliometricians, and research evaluators here, others have argued that the interplay of information overload, novel digital technologies, and research evaluation transformed the scientific information economy. For example, Peter Woelert argued:

> *The same period that saw the spreading of technologies that made possible the effective electronic storing and dissemination of published research, and which in the process made available a large body of electronic data, also saw a significant expansion in the actual use of formalized, standardizing systems to evaluate and steer research activities.*
>
> *(Woelert, 2013, p. 358)*

All in all, the second half of the 20th century can be characterized by the emergence and maturing of a socio-technical governance layer on science, or a technical gaze on the academic system. This gaze means to see scientific practices and products in an abstract and derivative way that ignores many peculiarities and contingencies to allow for a horizontal and comparative perspective. This development can be interpreted as the emergence of a new definition or interpretation of the value and relevance of scientific work. For example, we have seen how librarians used journal metrics to make decisions about library catalogs, or how research evaluators use metrics to assess productivity, efficiency, and excellence at a general level.

The concept of scientific quality or relevance that stems from publication metrics must be interpreted in hindsight to broader implications that go beyond counting papers and citations. Or to put it more provocatively, why should we bother about today's librarians, bibliometricians, or research evaluations being fascinated with publication metrics? There are two answers. First, the emergence and proliferation of publication metrics have direct and profound implications on the science system and the way knowledge is produced today. We will explore these direct implications in chapter 1.6. There, we will turn to the individual researcher and explore the constitutive effects of bibliometrics-based research evaluation that led to a gaming behavior of researchers, journals, and organizations alike. In this phenomenon individual researchers can utilize their degrees of freedom, i.e., the gap between doing and writing science, to increase their indicator performance. However, the metrics-based reconceptualization of scientific relevance also introduced an indirect but much more profound shift in how value and relevance of science are perceived. Against the background of a metrics-based regime, this shift implied a new social contract for science that is highly controversial because it suggests separating between intended and unintended consequences instead of promoting a plurality and diversity of research goals (Csiszar, 2020; Griesemer, 2020). While I will explain this shift next, chapter 4 will elaborate on a similar value shift that came with the promotion and measurability of transparency as a new goal for science.

1.6 The fertile soil for futile science

1.6.1 Working for the metrics

Previously, I have explained how the growth of the public research sector and the resulting flood of scientific publications gave rise to publication metrics and turned them into a form of understanding the value and relevance of scientific knowledge. Now, I will address some of the more direct implications of this metrics-based concept of relevance, e.g., the impact of research evaluation and the resulting metrics culture. After this, I

hope that I have provided a comprehensive overview of the different interests involved in modern biomedical research and how these interests came with their own quality notions, promoted different social patterns and technologies within scientific publishing, and, when being in conflict, provided the fertile soil for the perception of a crisis in medical publishing.

Individual researchers became attentive to the indicators and publication metrics that were offered by information scientists and promoted by research evaluators. Since good indicator performance could provide additional (or necessary) funds for further research, the whole scientific community became attentive to these indicators (Mirowski, 2018). This implies that scientists have taken up the idea that scientific quality and excellence manifests in measurable and assessable forms such as citation impact. Especially in modern biomedicine, the meaning of the *best-performing actor* is heavily shaped by market-based reasoning around the inclusion in high-impact journals, leading to "folk theories" about Journal Impact Factors (Rushforth & de Rijcke, 2015, p. 135; see also Van Leeuwen et al., 2003). The (highly self-referential) idea behind such folk theories is that authors who published their articles in high-impact journals faced greater challenges because such journals were stricter in terms of quality assurance and peer review. It also implied that they had excelled in competition, since most researchers wanted to publish in a relevant, highly visible journal. Consequently, the idea of high-quality research work became equivalent to getting published in high-impact journals. For this reason, many researchers perceive citation impact and scientific quality as highly related or even equivalent (Aksnes & Rip, 2009). [32] Consequently, journal metrics became a handy selection criterion that helped researchers to navigate the flood of scientific papers.

The rapid growth of academic literature has affected researchers just as much as it has affected librarians or evaluators. Due to their limited reading time, researchers and doctors faced the challenge of sifting through the vast amount of information. In fact, they were the first in line to suffer from the information overload that I have discussed previously. To cope with this information overload, researchers started to use electronic publication databases and platforms that enable them to filter and order search results (Guédon, 2001; Woelert, 2013). These results are ranked by relevance criteria that often involve publication metrics such as citations or alternative metrics, i.e. mentions in blogs

[32] Although this notion is rejected by bibliometricians, they have conspired to some extent, as there are bibliometric studies that statistically prove the correlation between citation-based indicators and other forms of quality assessments (see Abramo & D'Angelo, 2011).

or news items (Sachse, 2019, 2025; see also Breuer et al., 2022). In such settings, citation values and journal rankings influence the reading behavior of individual researchers or doctors. Instead of reading every paper in the field of echocardiography (Fraser & Dunstan, 2010), a doctor or researcher could decide to read just the papers from the highest-ranked journals, since they believe that these journals provide the most valuable and relevant papers (Biagioli & Lippman, 2020; Teplitskiy et al., 2022). However, metrics-based conceptions of quality and relevance have even more profound implications for the producers of papers than for the readers.

Publication metrics have become production goals in modern research faculties. Research faculties with formal performance-based funding systems often consider publication output and citation impact, or even Journal Impact Factors directly (McKiernan et al., 2019; Woelert & McKenzie, 2018). For example, medical faculties in Germany have included Journal Impact Factors in their formulas for performance-based funding, which attributes additional rewards to employees who publish in prolific journals (Brähler & Strauß, 2009). In their introduction to the volume Gaming the Metrics, Mario Biagioli and Alexandra Lippman put it illustratively:

> *As an author, I can "trade" articles with certain impact factors into a job, and the institution that employs me can then "trade" those publications (together with hundreds or thousand more by other faculty of the same university) into a better national or global ranking, which may be subsequently traded into more students, donors, contracts, and so on.*
>
> *(Biagioli & Lippman, 2020, p. 7)*

Even more profound examples are university PhD regulations. These regulations commonly employ journal lists that specify which journals must be targeted so that a publication is considered as eligible for a dissertation (Willmott, 2011). Some of these programs even involve a scoring system, in which a final score must be achieved before a PhD can be granted. Journals with higher rankings award more points, and publishing more papers can offset lower scores. This assumes that the quality and quantity of knowledge are interchangeable dimensions with regard to the number of invested working hours. Thereby, early career researchers and PhD students are trained to think with indicators. Relatedly, quantitative indicator performance has become an argument for promoting the value and importance of new professional roles and career paths. For example, in the case of scientific literature reviewers (Garfield, 1977a; Sambunjak &

Puljak, 2010), data stewards (Mooney & Newton, 2012), or research software engineers (Druskat et al., 2021).

Researchers must be attentive to publication metrics under the current information economy in science. But what is so special about this development? Today, the proliferation and embeddedness of publication metrics in evaluation frameworks and wider research culture are highly criticized. A central tenet of this criticism is that a more general and incidental attentiveness to metric performance turned into a more pivotal form of attention and a peculiar strategic goal for researchers. As a result, they not only invest time and resources to learn about the metrics and try to game the system (Biagioli & Lippman, 2020). Moreover, the aim to perform well can impact epistemic practices because researchers prefer those projects that promise the biggest returns or interpret their results in a likewise favorable way.

A famous criticism of this development speaks of unintended consequences or goal displacement (Csiszar, 2020). Critics argue that the strategic adaptation of scientific actors to the metrics-based performance system is a form of misbehavior or management deficit. This criticism is often related to Goodhart's law, which implies that a measure becoming a target ceases to be a good measure (e.g., Fire & Guestrin, 2019). However, this line of criticism has been contested. It has been argued that narratives of unintended consequences and Goodhart's law in academia narrowly focus on the measure but not on what is measured. Both narratives inhibit realist interpretations of citation metrics. They inherently conceptualize intended behaviors and consequences with respect to scientific goals and then identify and blame moments of deviation. However, these accounts fail to uncover how metrics-based research evaluation reshaped the social contract between science and society into a relationship between goals and their achievement. In this regard, Alex Csiszar recalls an argument made by Yaron Ezrahi at one of the very first conferences on science indicators, held in 1974:

> *Broad public views about the legitimate ends of scientific activity cannot in general be separated from public representations of scientific activity and its productivity. For that reason, public indicators of science affect not only the practice of research, but also the public legitimacy of science.*
>
> *(Csiszar, 2020, p. 38)*

In recent decades, researchers have uncovered many ways to game the system of performance-based evaluation. More direct adaptations to the metrics culture can be

found in the production and publication of scientific documents. For example, by the *salami-slicing* of publications, i.e., the fragmentation of research results into multiple publishable units (Fire & Guestrin, 2019). This disperses project results to multiple places and documents, which increases the authors' published output while it also increases the readers' efforts to comprehend and contextualize the overall results of the project. Similarly, researchers may publish another paper that provides only a little more insight than their previous one, or even no additional value in the case of multi-publication or self-plagiarizing (Woelert, 2013). In general, however, researchers are pressured to turn as much of their efforts and work tasks into codified forms of documentation as possible. Undocumented or unpublished work becomes invisible. Otherwise, a scholar risks low evaluative scores and may observe her position being cut. This forces researchers to either publish or perish (Van Dalen & Henkens, 2012, p. 1282).

Publish or perish can go further than salami-slicing. Collaborating with high-impact researchers can boost one's own indicator performance so that the oeuvre of a researcher becomes more important than her expertise or research. Likewise, team-internal hierarchies or the discussion of ideas and results can be guided by evaluation metrics (Rushforth & de Rijcke, 2015). Researchers who are sensible to metrics may also choose trendy topics and projects that promise the highest impact. In turn, they shun more risky research projects with uncertain and probably negative outcomes, which, as we have seen above, are often not publishable (see also Gläser & Laudel, 2007). But when they publish, they accept cuts requested by editors and peer reviewers and thus subordinate the opportunity to get a paper out under informational needs (Bohorquez et al., 2024).

Scientific publications also experience content changes. I have shown above how a writer's (envisioned) audience shapes the content and writing style of her text. Now, the audience of scientific texts grew by research evaluators and by a different type of reader—academic peers who cite the work, thereby making it relevant and valuable. Readers are not mere consumers of content anymore but donors of relevance and academic recognition because citing and getting cited has become a form of currency exchange (Wouters, 2020). In that sense, scientific texts and their authors have entered a race for attention that influences what and how texts are written, published, and promoted (Migheli & Ramello, 2018).

Discovering novel and strong treatment effects or proving counterintuitive relationships does not only draw attention from healthcare policymakers and pharmaceutical industrialists, as shown above. Moreover, such findings of course draw a lot of attention from researchers who might cite the work (Intemann, 2020; McGrath et al., 2017). But

while major breakthroughs are risky and not plannable, authors can frame their interpretation of results or conclusions in a more noteworthy, novel, or positive way. This might not turn their publications into award-winning pieces of knowledge but spins the amount of attention they get in their favor. Other researchers then cite the study and increase its scholarly (citation) impact, either by going along or criticizing the results, irrespective of if they were groundbreaking or exaggerated.[33]

The impact extends to academic journals. The JIF sets the journals at fierce competition, and editors must take measures to ensure that their journals draw attention and increase their JIF (T. R. C. Davis, 2004; García et al., 2018; Weingart, 2005). To do so, they can take advantage of some structural or systemic biases in the calculation of the JIF and turn them into viable strategies to improve the JIF of their own journals (Falagas & Alexiou, 2008). For example, they might invite eminent researchers to write review articles (Blümel & Schniedermann, 2020) or require submitting authors to cite other articles from that journal or an affiliated journal, engaging in so-called citation cartels (P. Davis, 2012; Willmott, 2011).

Taken together, the growing attention to the outputs and impacts had a profound impact on intellectual work and academic culture. In that sense, scholars have argued that publication indicators and formal assessments change the environment in which they are implemented (Espeland & Sauder, 2007; Woelert, 2013). Although there is only little systematic evidence about such factors and influences, it is argued that the metric tide made researchers basically start "thinking with indicators" (Müller & de Rijcke, 2017, p. 158; see also Hessels et al., 2019). Such ways of thinking with indicators and the corresponding behaviors might distract researchers from purely intellectual routes, or at least those they would have taken without the pressure to publish or perish.

The pressures and changes in behavior and research culture are commonly associated with research evaluation, the management of academic bodies, or the definitions and funding formulas of national science policies. We have seen how these actors fueled the development and implementation of evaluation indicators. In this regard, it has been argued that these forms of science governance set the wrong incentives, that there is

[33] An illustrative, yet ironic example can be found in scientometrics itself. Park et al. (2023) argued, based on a citation network analysis, that science becomes less disruptive over time, i.e., how science is in decline. The famous Nature journal paper grew wide reception but is infamous in the scientometrics community because, among other issues, of its operationalization and interpretation of the calculated indicators (Leibel et al., 2024; Macher et al., 2024).

an “evaluation gap” (Wouters, 2017, p. 108),[34] or a “performance paradox” (Dahler-Larsen, 2014, p. 972). In other words, there is a mismatch between what is measured and what is actually relevant to the epistemic enterprise. Such arguments suggest that there are different types or interpretations of the relevance and value of scientific work. By now, this should sound familiar to the readers of this chapter, as we have considered a multitude of roles and goals of medical knowledge production in Western societies. With regards to the evaluation gap, however, STS scholar Karin Knorr-Cetina differentiated between recognition and credit:

> *Thus, recognition presumably reinforced the truth-seeking, achievement-oriented behavior considered most essential to the scientific system. In contrast, credit is defined as a symbolic capital acquired by scientific agents through the imposition of technical definitions and legitimate representations of scientific objects on the field.*
>
> (Knorr-Cetina, 1981, p. 70)

1.6.2 The complex crisis of scholarly communication

To this point, we have painted a wider picture of the impact of publication metrics on modern academia. I have explained how the flood of scientific information challenged researchers, librarians, and information specialists alike so that they utilized science indicators as technical solutions. Further, we have seen how bibliometrics became a scientific field that turned science indicators into epistemic objects that are explored, validated, and discussed with scientific methods. At the same time, indicators were used to evaluate science at different levels. In the wake of this development, policymakers and funders created another interpretation of scientific relevance and renegotiated the social contract between science and society.

Although many of these developments are observable in wider academia, they are most pressing in the life sciences, the biggest sector, which also encompasses biomedicine. Even more, medical research has always been faced with translational needs and questions of societal relevance, which contributed to its receptiveness for performance evaluation and science indicators (Rushforth & de Rijcke, 2015). Against this background, the proliferation of science metrics cannot be rooted in any single cause.

[34] This is a different gap from the one between conducting and reporting a study which was introduced on page 17.

Moreover, we should be hesitant to blame instances of gaming the metrics on any hegemonic and neoliberal ideology solely pushed by modern science policies and research evaluation frameworks (Bacevic, 2018; Hammarfelt & Hallonsten, 2022). Instead, we must dismantle the audit culture and attention economy with precise analysis and identify why certain phenomena emerge and take hold. We must consider contexts, i.e., the fertile soil that co-produces judgement devices such as publication metrics and novel definitions of scientific quality and relevance. In this regard, it has been suggested to take Goodhart's law for granted and then focus on the measures that actually become targets (Csiszar, 2020; Griesemer, 2020).

All in all, evaluators, journals, librarians, and the researchers themselves have engaged in a downwards spiral that focuses so much on the metrics rather than the contents and circumstances of research (Falagas & Alexiou, 2008; Weingart, 2005). We have seen how the metrics-based concept of scientific relevance and value lies at odds with traditional conceptions of scientific quality or epistemic value. But it also lies at odds with more societal concepts of quality and relevance of medical knowledge:

> *I would ask those who designed the impact factor and those in the administration departments of universities who consider The Journal of Hand Surgery of so little importance whether they would be concerned if they discovered that a surgeon who is about to operate on their hand does not subscribe to any journal of hand surgery and has not read any such journal in the past 10 years, preferring to read Nature instead.*
>
> *Timothy R. Davis in his editorial "An Age of Enlightenment or Information Overload", Journal of Hand Surgery (2004, p. 522).*

The quote illustrates how the content of published knowledge is subordinated under its evaluation in the form of the Journal Impact Factor. Journals from a small field, such as the Journal of Hand Surgery, do not achieve impact factors comparable to Nature, which publishes papers from all medical domains (see also García et al., 2018). Yet if we need hand surgery, we will surely prefer the worst hand surgeon over the best general practitioner in town.[35]

[35] I highly recommend the full editorial by Timothy Davis because it is not only written in an entertaining way. Moreover, the parting editor tries to manipulate the JIF by citing all publications from the previous two years: "I hope that my successor, David Elliot, will enjoy this post as much as I have." (T. R. C. Davis, 2004, p. 523) This attempt was

Much of the criticism and problematization of the medical literature can be attributed to the mismatches between epistemic ideals, the demands of knowledge consumers, and contemporary research evaluation. These mismatches, far from being unique events or modern trends, could be better understood as a constant exchange between the various poles and influences that have always shaped science, particularly biomedicine. In this regard, Samuel Shapiro depicted medical publishing as an ethical dilemma in 1985.

> *Over the course of a 15year experience with research on the effects of drugs, we learned that the issues are important to many constituencies. Broadly, these include practicing physicians, the research community, regulatory agencies such as the U.S. Food and Drug Administration, drug manufacturers, and the public. To publish on drug effects is to risk the anger or pleasure of each. Each constituency has its own unique interests, and consciously or unconsciously exerts its own unique pressures.*
>
> *(Shapiro, 1985, p. 365)*

However, while we have seen many historical examples, many of the underlying issues are fundamental aspects of the configurations in modern science and therefore still prevail. Moreover, the criticisms and problematizations have been uttered more voicefully and vocally since the 2000s, as we have seen in the introduction. Furthermore, they appear to have culminated in an unprecedented level of alarmism, fundamentally questioning the value of biomedical research. This alarmism has permeated the codified literature and garnered discussion in prominent forums. For example, in an influential series in the famous journal *The Lancet* (I. Chalmers et al., 2014), or by otherwise influential authors (see I. Chalmers et al., 2002; Glasziou et al., 2014; Lund et al., 2016; Roberts & Ker, 2015). Often, experts criticize “research waste” (e.g., Glasziou et al., 2014, p. 267), thereby chiming in on traditional theories of how the public sector in general and the healthcare sector in particular waste taxpayers’ money (see also Tanenbaum, 1994).[36]

detected by Clarivate and the journal was not provided with a Journal Impact Factor in the following year.

[36] We will come back to the striking similarities between the contemporary metascientific discourse and the discourse on therapeutic efficacy in the emergence of

Criticisms also highlight the discrepancy between traditional concepts of relevance and current incentives. For example, the high epistemic status of research syntheses has been called into question by bemoaning "the mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses" (Ioannidis, 2016, p. 485). Although the systematic review method has been developed to cautiously assess primary research and consider its potential biases, reviews can't be done if the overall existing literature is not sufficient (Kirkham et al., 2020).

Ironically, the three pressures we have examined so far sometimes also contradict each other. For example, I already mentioned how some systematic reviews do not include any studies or that medical practitioners find Cochrane Reviews uninformative. In such cases, the work was driven by scientific standards and methodological rigor, which came at the cost of what clinical practitioners may consider as relevant. What researchers consider highly sophisticated and methodologically strict work might be rendered useless or irrelevant by doctors and healthcare practitioners because they find it too strict or constrained. This tension or tradeoff has been described as a fundamental gap between medical research and the application of medical knowledge (Cartwright, 2007). With respect to the inclusion and exclusion criteria of randomized controlled trials, Steven Epstein summarizes the distinction as follows:

> *Strict inclusion/exclusion criteria are sometimes used to create a more standardized and homogeneous research population for a study, on the argument that the more researchers reduce the number of variables that might affect a study, the easier it will be for them to distinguish 'signal' from 'noise.' On the other hand, experts on clinical trials who are concerned with the 'external validity' of a trial—the significance of its findings for large numbers of people in everyday life settings, and not just for the smaller number of people who happened to participate in the clinical trial—often tend to favor more relaxed entry criteria. These experts argue that heterogeneous research populations are better models of actual conditions and that studies with homogeneous populations may result in scientifically elegant experiments that lack much real-world significance. There is no solution, in principle, to the problem of how to resolve this trade-off between experimental rigor and generalizability of findings.*

modern public healthcare in the concluding section, "8.2 From meta-analysis to meta-science".

(Epstein, 2007, p. 49)

Another contradiction can be identified with regard to the demands of medical therapy and drug development. Basic research has been found lacking in validity and replicability by applied researchers, so that there is an unbridgeable rift or "valley of death" (Seyhan, 2019, p. 1) between both strands. This criticism was also prominently uttered by researchers from the pharmaceutical industry. On the one hand, the industry requires positive results from clinical trials to prove efficacy and license new drugs and get their returns on investment, which fuels pressures on researchers towards more favorable or even breakthrough results. On the other hand, the industry depends on the validity and replicability of basic or preclinical research results for developing effective drugs and treatments in the first place (Prinz et al., 2011).[37] To them, exaggerated and irreproducible claims can mean additional costs, because they must validate these findings and even completely drop initially promising projects that already costed money.

The more recent problems in the biomedical literature did not just refuel single-sided allegations against the pharmaceutical industry or modern research evaluation. Instead, a nuanced and informed understanding of the interdependence of diverse influences and demands in modern biomedical research called for a canny and constructive perspective on the problems and potential solutions. Instead of trying to expel non-scientific influences from science, experts sought integrative solutions that do not require system changes (see also K. C. Elliott & Steel, 2017). Practically, this can mean reinforcing scientific quality goals by recodifying them as new standards or incentives for researchers. In this regard, John Ioannidis concluded in an interview:

> *Incentives are a core feature in driving what we want to get out of science. Scientists are very bright people, very smart, and they will do their best to try to fit whatever orders are given to them. So if they're told, 'You need to get extreme, extravagant, extraordinary results,' they will come up with them because that's what will allow them to continue doing what they enjoy and what they love. I'm not saying that they're criminals in doing it. They just want to do science and continue to do science. So, we need to find ways that our*

[37] This study illustrates the replication attempts by researchers from BAYER and was featured in the Wall Street Journal (Naik, 2011) where it led to a culmination of a more medicine-oriented sphere of the replication crisis.

> *reward and incentive system is aligned with the expectation not being spectacular but flawed results, but accurate results.*
>
> *(Ioannidis, LaBlanc, 2024)*

Such recodifications have become a central tenet of the metascientific community as well as the wider open science movement. This movement promotes epistemic values such as openness, transparency, and accessibility but does so on a broader and cross-disciplinary level (Fecher & Friesike, 2014; see also Hosseini et al., 2022). In particular, developments such as open access to scientific literature or the sharing and reusing of scientific resources such as data and software code are facilitated and promoted under open science. Thus, after developing infrastructures to make open science possible and easy, open science enthusiasts urge that it be made normative, rewarding, and ultimately required (Nosek, 2019) Their aim is to reemphasize scientific and intellectual conceptions of relevance and scientific validity while pushing back those definitions that originate elsewhere and incentivize strategic behavior, gaming the metrics, or even misconduct (Nosek et al., 2013). Their main tools are new metrics and devices that represent intellectual values such as openness, transparency, and rigor, which are assessed by novel metrics and standards (e.g., Nosek et al., 2015). Therefore, we might find yet another pattern or tension in the contrasting notions of scientific relevance, as well as another narrow definition of universal epistemic and societal goals of science (Csiszar, 2020; Griesemer, 2020). One that is based on traditional metrics such as numbers of publications and citations, and another one that is based on formal assessments or metrics of open, transparent, and rigorous science. We will see how reporting standards may contribute to the latter paradigm throughout the rest of this book.

All in all, the historical and social background I have provided so far has shown how there is a creative gap between performed research and reports thereof and how this gap is affected by the external and internal contexts of biomedical research. Non-scientific actors define the relevance of biomedical research with respect to intentions. Pharmaceutical companies require testimonials of drug efficacy, doctors need knowledge about patients and disease, and policymakers require a sound evidence base to assess and legitimize whole healthcare programs. But also, the research system in a narrower sense increasingly established measures to judge the value and relevance of individual contributions.

Measures became necessary because the biomedical research sector faced substantial growth in personnel, organizations, and published research papers. But researchers

were not just victims of their own success who now need some management and proper incentivization. Instead, research communities and individual researchers co-constructed the new evaluative regime in many cases. In this regard, I have unveiled the many and sometimes conflicting trajectories of current modes of biomedical knowledge production. We have seen how a diverse set of influences shape research practices and the biomedical literature in a way that certain scientific stories, literature genres, and publication venues can become powerful and proliferate. But we also got a glimpse of how some of these influences can be at odds and result in effects that are often described as unintended consequences (Dahler-Larsen, 2014).

Reporting guidelines have developed and evolved against the backdrop of a diverse and complex set of influences and demands in modern biomedicine. The previous sections have shown how there is a diverse and broad palette of intellectual, social, and political factors at play in modern biomedical research. Likewise, the emergence, dissemination, and proliferation of reporting guidelines are presumably shaped by a diverse set of factors as well. A proper analysis of reporting guidelines as a broader social phenomenon should thus go beyond the search for single causes and determinants. Instead, a comprehensive and holistic understanding requires focusing on how the guidelines, as complex social phenomena, get embedded in the contexts and practices of biomedical research.

Analogous to acknowledging the social complexity of the phenomena under study, a proper analysis must remain open to the various perspectives and opinions about the current biomedical research sector that will likely be encountered. This also involves any form of criticism. Although there are fierce criticisms of how pharmaceutical industries and private interests influence science for the worse, we have also seen how detrimental and problematic aspects are often co-created by different actors in (unintentional) collaboration. Thus, any sociological assessment of reporting guidelines must be epistemically broad and robust in its methods but also truthful and modest in its interpretative mode. For this reason, the next chapter will introduce the project from a methodological and theoretical perspective.