Introduction

Weigl used to do everything perfectly, whatever it was. He spent money on only two passions: fishing and archery. However, the fishing rods, trout flies, and bows and arrows that he acquired on his travels around the world never quite met his standards. For him, the only way to be sure of having good equipment was by designing it oneself. In his research, he was just as much of a perfectionist as in his passions, and he mobilized his considerable ingenuity to do whatever was necessary in order to achieve his goals. At the beginning of the twentieth century, researchers were trying to force lice to eat germs in order to create a vaccine against typhus – a deadly epidemic that killed more of Napoleon's soldiers as they retreated from Moscow in 1812 than the Russians did. Weigl, with characteristic inventiveness, literally turned their reasoning upside down, saying "we'll stick them up its [the louse's] ass" (Allen, 2015, p. 22).

Rudolf Weigl – a Polish biologist from Lviv – invented the first effective vaccine against typhus, which saved countless lives during the interwar period. As part of his research, he employed people to serve as lice feeders; they had cages with typhus-carrying lice strapped onto their thighs. These feeders were most often Polish intellectuals, Jews, and underground fighters. In the group of over 2,000 feeders, there were two outstanding men: a mathematician, Stefan Banach, the founder of modern functional analysis, and a microbiologist, Ludwik Fleck, who developed the concept of "thought collective" crucial for Thomas Kuhn's later notion of the "paradigm shift." Weigl was nominated many times for the Nobel Prize, but he was never awarded it. Once, he refused to be a candidate because according to him, his discovery was not among the highest-ranking ones. Another time, during World War II, he opposed becoming a candidate from Hitler's Germany (Wincewicz et al., 2007). For Weigl, it was the discovery that was science's most important outcome, not awards or publications.

Weigl's approach to publishing academic papers was radical. He believed that the actual research consisted in doing science and making discoveries, whereas writing

was torture and a waste of time. While a student, he had been forced by his supervisors to publish papers just to keep receiving his scholarship. However, he drove his own students to despair because his way of doing science reduced the number of publications they could put on their résumés (Allen, 2015). What is now called "salami publishing" in science – that is, dividing a large study that could be published as a single research paper into smaller papers – Weigl called "duck shit": just as ducks leave a lot of traces while walking about in the yard, scientists hastily publish articles with partial results that are the product of undeveloped thought. It was Weigl's belief that the true value of research manifested itself in its impact on society and not in publications. Being neither the means of exerting an influence on society, publications were also not one of science's key pillars. And yet today, in most academic environments, they are considered the engines of science's development.

Publication-Oriented Science

Today's science is publication oriented; it is communicated, organized, financed, governed, and evaluated through publications. Seen from the current centrality of publication to science, Weigl's approach appears idealistic and unconvincing. I would put it even more forcefully: Many researchers and policy makers perceive science as confined to publications, forgetting the real people who do the research, make discoveries and inventions. These are the people who work in institutions that together constitute "academia." However, academia should not be understood as a kind of collection of geniuses but rather as an international collective endeavor that involves thousands of researchers, technicians, students, and administrative staff members working in a wide variety of places.

Nonetheless, the image of science as the work of geniuses is quite popular, and the lives and views of well-known scientists are used to promote differing visions as to how science should be organized, financed, and governed. In 2013, Peter Higgs – a Nobel Prize laureate in physics – declared that no university would have employed him in today's academia because he would not be considered "productive" enough (Aitkenhead, 2013). This statement has been used as evidence that a "publish or perish" research culture has conquered science, and to back up the argument that academia must cease to be publication-oriented. Under a culture of publish or perish, academics are pressured into producing large numbers of publications, not only to succeed but merely to survive in their work environments. Yet academia has been publication oriented for many years. In her tribute to William J. McGuire, social psychologist, Banaji (1998) refers to McGuire's bet with a colleague that he would not publish a single paper until after receiving tenure at the University of Illinois. McGuire was tenured in 1960 and, one year later, published ten papers that he had already written but had not submitted for publication.

Banaji (1998) mentions that "this act of daring made him an instant hero of many of us when we were graduate students."

Although such arguments are compelling, they rarely offer solutions for how researchers can survive in academia. Instead, they end up reducing the discussion to the level of the absurd. Taking such arguments even further, one could say that Albert Einstein would not have been awarded tenure in the present days given that his major works were not published in English or – and here we reach the ultimate phase of pure absurdity – that Socrates would not have been granted a project because he did not publish any papers. In this way, the academic community uses ahistorical reflection to build various myths which it then deploys in its defense of the profession's autonomy.

Since the Manhattan Project started to produce the first nuclear weapons during World War II, science has changed irrevocably. In the mid-twentieth century, it entered fully into the era of big science which was characterized by a rapid growth in the number of institutions, researchers, discoveries, and publications (Price, 1963). At the same time, discussions on the role of science in society started to be shaped by definitions of scientists' responsibilities toward society. As Shapin (2012) shows, there never was an Ivory Tower and yet the call for scientists to leave it became the leitmotiv of twentieth-century reflection on the impact of research.

Research initiatives like the Large Hadron Collider or the Gran Telescopio Canarias need financial support from many countries. Because of this, science has begun to be both financed and carried out internationally. Science's ongoing development requires tools through which governments can distribute public funds and evaluate the results of provided inputs. However, science, the research process, and scholarly communication are too complex to have all their features reduced to a single model. Because bureaucrats seek clear and undemanding solutions with which to justify their decisions, the models elaborated for this purpose need to be simple and based on readily comprehensible elements. They should, moreover, be easy to explain to the general public. It is thus very tempting to use publications as science's touchstone: one can easily count them and say who has published more.

Scholarly communication is one of the key pillars of science. However, it does not only manifest itself in the publication and dissemination of research results through academic papers. Scholarly communication begins with reading, discussing, and arguing. Presenting and reviewing preliminary research results are also important phases of scholarly communication which does not end at the time of publishing. Indeed, as a circular process, scholarly communication cannot be reduced to any one of its phases. Yet in actuality, it is subjected to precisely such a reduction, which in fact takes place twice. First, when researchers and policy makers start to think that disseminating research results means only publishing academic papers. Second, when they identify journal articles as the

key elements of research and science themselves, to which they can be reduced. Finally, researchers and policy makers end with publications that are abstracted from research which represent the science. This synecdoche represents a pivotal feature of today's academia. For instance, Dahler-Larsen (2015) argues that one of evaluation's constitutive effects is to provide an erroneous image of what is actually going on, and to support the overarching assumption that research should be understood as production (e.g. of new papers, patents etc.). More importantly, such abstracted publications can be counted and the results can be used to justify various decisions. Nonetheless, someone still needs to decide how publications will be measured and what measures will be applied. In this moment of deciding what (e.g., all academic publications or only peer-reviewed ones?), by whom? (e.g., policy makers, researchers), and how (e.g., should all co-authors take full credit or should it be distributed among authors?) publications are evaluated, the power of measures is established. Measures and measuring are the technologies of power (i.e., instruments embodied as sets of protocols, indicators, and policy aims) that are used by global actors, states, and research institutions to evaluate science and through this, to fund, control, and govern the whole science sector.

For Weigl, counting publications in order to measure the true value of research constituted an offense against of the very essence of science itself. Today, the "metricization" of science, that is the introduction of metrics into research and academia, is global. A great deal has been written about this phenomenon: starting from studies on the quantification of social practices (Espeland & Stevens, 2008; Muller, 2018), histories of measuring science (Godin, 2005, 2009), audit cultures in higher education (Power, 1999; Shore & Wright, 2003; Strathern, 1997), research on the impact of using indicators in science (De Rijcke et al., 2016; Müller & De Rijcke, 2017), monetary reward systems in academia (Quan et al., 2017), through critiques of university rankings (Pusser & Marginson, 2013; Sauder & Espeland, 2009; Yudkevich et al., 2016), to the consequences of local uses of research evaluation systems in various countries (Aagaard, 2015; Aagaard & Schneider, 2017; Butler, 2003b; Kulczycki, Rozkosz, & Drabek, 2019). The community of scholars that focuses on using metrics for monitoring, reporting, managing, and – most often – evaluating research and researchers, produces a discourse that shapes science policy and has an impact on research and academia. Thus the way in which metricization based in publication metrics is discussed influences the system of science itself.

Academia constitutes a complex system which has its own history and heritage. Although science is global, inside this global system, various local, institutional, national or regional structures can be found, and each of these has its own specific background and features. Discussions on using measures, metrics, and evaluating research show what tensions might occur when local or national conditions meet global values and challenges. Moreover, these discussions reveal how the demands

of the global market in higher education can change researchers' work and ways of communicating their research. Nonetheless, the responses – or in other words, implementations of global demands – vary by virtue of their dependence on specific cultural and historical contexts.

Aims of This Book

In this book, I aim to offer an alternative position to the discourse on using publication metrics and measuring science that is being produced by the community of research evaluation scholars. This book focuses on research evaluation systems to make scholars and policy makers aware of two key blind spots in this discourse.

The first blind spot relates to the absence of the Soviet Union and post-socialist countries in the histories of measuring science and evaluating research, despite the fact that these countries have played a key part in this history from its very inception. In these countries, a distinct discourse on using metrics in the system of science – based on the scientific organization of scientific labor and central planning – was articulated.

The second blind spot relates to thinking about global differences in studies of the transformations of scholarly communication. I show that the contexts in which countries face the challenges of publish or perish culture and questionable (or so-called predatory) journals and conferences, should be taken into account in discussions about them. In order to identify and explore both blind spots, we must therefore include state(s) as an explanatory factor in the study of the effects of research evaluation.

Moreover, I argue that in today's academia, research is influenced not only by the global context of the knowledge-based economy or the idea of accountability in public funds. The way in which research is done can also be mediated through national and local science policies. For instance, in some countries, among them Canada, France and the United States, researchers experience research evaluation mostly when universities measure their publication productivity or when they are faced with a tenure procedure. In other countries, for instance in Australia, Argentina, Norway, Poland and the United Kingdom, research evaluation is also experienced through national research evaluation systems which influence both university strategies and tenure (or tenure-like) procedures.

These blind spots inspired me to elaborate an alternative perspective on the history of research evaluation and its effects going forward. This perspective is based on the concept of the "evaluation game," which refers to the ways in which researchers, academic managers, and policy makers react on being evaluated and when they act as evaluation designers. While, as stated above, the research evaluation game is a global phenomenon, it was not my intention to write a book about the global phenomenon of

research evaluation and measurement that would attempt to account for every distinct situation in every part of the globe. The world of science is far too complex to capture in a single book. Rather, as a member of the community of research evaluation scholars, my aim is to intercede within the discourse on research evaluation in a way that promotes a deeper understanding of the social world of science, thereby contributing to creating a better environment for research and researchers.

In this book, I conceptualize doing research and managing academia along three intersecting planes, in each of which the simultaneous significance for doing science is also emphasized: (1) global (supranational), (2) national, and (3) institutional/local. In this way, this is a *glonacal* perspective (see Marginson & Rhoades, 2002) in which conceptualized phenomena are characterized by global, national, and local planes.

The first plane involves global transformations of academic labor (e.g., the pressure to produce more and more publications or to publish mostly in English). The main actors here are supranational institutions like the World Bank, the Organization for Economic Co-operation and Development (OECD), global companies producing bibliographic and bibliometric data sources like Elsevier and Clarivate Analytics, and institutions producing university rankings such as Shanghai Ranking Consultancy (Shanghai Ranking) or Times Higher Education (World University Rankings). These organizations transcend the state form and through their own global influence play an important role in the universalization of publication metrics in science as well as in shaping national systems by providing global exemplars and solutions like field classifications or bibliometric indicators (cf. Godin, 2005). In this book, I argue that one can observe some of the manifestations of characteristics specific to countries of the Eastern Bloc or to Western countries at this supranational level.

On the second plane, research evaluation systems influence the day-to-day work of members of the academic community in a given country. At the same time, through their administrators and policy officers, states have to react to and interact with transformations and values manifesting at the global level that might cause varied tensions. These might, for instance, include tensions around the promotion of publishing research results in English (a goal of scholarly communication rooted in global-level values) or national languages (a goal rooted in state values that cultivate local culture and heritage or aim to reduce inequalities in academia). In this book, states (especially at the national levels) are perceived as characterized by a high degree of sovereignty and agency. This view might be criticized as ahistorical or even essentialist (cf. Jessop, 2002) considering that the state has lost its sovereignty and has become more of an instrument in the hands of various – often global – interest groups. In other words, one might say that globalization has weakened the agency of the state and made it rather a medium of and for the global context. Moreover, New Public Management – the group of ideas that is transforming

relations across government, public institutions, and society – might be perceived as a departure from the traditionally understood state and its central (not in the Soviet sense but in the sense of the welfare state) management of the public sector.

On the one hand, I second this view of the state's weak position. In Chapter 2, I argue that in fact the state should not be the basic unit of analysis and could instead be replaced by a broader unit, that is the world-system (Wallerstein, 2004). On the other hand, however, the state functions as an actor on the global stage that is both influenced by the global context and influences its own local contexts. This is why I argue that doing research and managing academia in the era of research evaluation systems should be understood as they are in this book: taking place through three intersecting planes, that is the supranational, national, institutional and locals (levels of practice). At the third level of academic practices, academia's institutions and people work, react, and adapt to changing conditions of academic labor. In this book, I am especially interested in reactions and resistance (at the local level) caused by state policy instruments, that is by research evaluation systems (at the national plane) influenced by global transformations (at the global plane).

Building on the above elaboration of the three intersecting planes, this book proposes a new theoretical framework for understanding the effects of research evaluation systems and using metrics in academia. I have written it in order to explore how the process of evaluating science – mostly through the prism of publications – shapes the production and communication of scientific knowledge in and through universities and research institutes, in a way that renders them akin to political institutions of the state. In asserting this, I am referring to the fact that they require state resources, the state provides them with certain benefits, and institutions in turn gain some authority from the state (Neave, 2012).

Research evaluation systems are science policy instruments used to measure academia's performance. I use the term "academia" to indicate members of a community who share common beliefs, norms, and values regarding science and research. This community includes primarily (1) professors and researchers from higher education institutions and research institutes as well as (2) these institutions and (3) their staff management. Individuals became a part of academia when they learn how to act (and what to believe) in a way that is acceptable to the academic community. Institutions become part of academia when they are defined as political institutions of the state in the higher education or science sectors.

Turning now to the effects produced by the operation of research evaluation systems, these can be either intended and unintended. Intended effects can be understood as goals accomplished and successful public interventions. However, when it comes to investigating the unintended effects, this cannot be reduced to tracking and reporting the unforeseen or unpredicted side effects of preplanned interventions. This is due to the fact that unintended effects originate not only in

social interventions themselves but also – among others – in the context in which such policies are implemented (e.g., unstable conditions of academic labor or scholarly communication reduced to publishing papers). Moreover, the intentions of policy designers, policy makers, or stakeholders are rarely explicitly communicated. Thus it can be difficult to assess whether certain effects were intended or not. Within evaluation studies, the distinction between intended and unintended has been criticized (Dahler-Larsen, 2014) from the position that all effects exert an impact on the evaluated reality. Nonetheless, I argue that this distinction might still be useful for understanding the science policy perspective through which policy makers assess the results of their efforts.

There is a long tradition, within both administrative and organization theory, of exploring the dysfunctional consequences of performance measurement (Ridgway, 1956). During the first half of the twentieth century, the byproducts and impact of performance measurements were analyzed in numerous areas that ranged from American and Soviet industries to public policies. Even then, studies showed that the use of a single measure was not adequate and should be replaced by the use of composites, that is multiple and weighted criteria or – as was later suggested – by multiple indicators. This was because no single indicator could ever reveal more than a small part of the multidimensional picture that is composed through research (Martin, 1996).

This knowledge and experience were utilized within New Public Management (Hood, 1991), which transformed performance measurement substantially into outcomes-based performance management (Lowe & Wilson, 2017). For researchers on evaluation systems, New Public Management is an often-cited reference point, which designates the central implementation of research evaluation systems. Its mention serves to emphasize a process of transforming relations across the government, public institutions, and society. In the second half of the twentieth century, these relations shifted, moving from an era of professional autonomy coupled with bureaucratic systems toward the promotion of efficiency in the production of public services (Lowe & Wilson, 2017).

I use the theoretical framework that I elaborate through three key steps to explore transformations of scholarly communication caused by the process of measuring and evaluating science.

First, I present the concept of the *evaluative power of the state* as a ground for developing the framework in which the effects of research evaluation systems can be investigated. Evaluative power is the capacity of the state to influence and shape the key area and to change practices of individuals and institutions. While evaluative power can also be identified as the power of global actors (e.g., companies developing citation indexes like Web of Science Core Collection [WoS] or Scopus), in this book, I focus mostly on evaluative power as a characteristic of states.

Second, I present the concept of the *evaluation game* through which the reactions provoked by evaluative power manifest themselves. In this way, the evaluation game is one of academia's responses to the changing context of academic labor. For instance, some forms of evaluation game manifest through the establishment of new and questionable journals or publishing within them only to fulfill expectations produced by the research evaluation regime. Thus I demonstrate that the evaluation game and its consequences are (un)intended effects of the design and use of research evaluation systems.

Third, I rethink the history of the measurement and evaluation of science and argue that understanding the consequences of research evaluation systems better requires the incorporation of an omitted part of this history. In other words, I show that performance measurement in the science sector is not only a hallmark of Western science but was in fact first implemented at a national level in Russia, and then later in the Soviet Union and Eastern Bloc countries. Bringing this heritage into the spotlight is a necessary step for understanding why in countries with similar research evaluation systems – like Australia and Poland, for example – the reaction and resistance against the systems manifests itself in diverse forms of evaluation game, and further, why researchers perceive the same elements of those systems (like the use of bibliometric indicators or peer review) in substantially different ways. For instance, as Mishler and Rose (1997) show, distrust is the predictable legacy of Communist rule and in the postcommunist societies of Eastern and Central Europe, trust in experts is substantially lower than in other societies. Thus, when in the 1990s, Poland implemented sweeping reforms in the science sector, skepticism about peer review was one of the key obstacles to rolling out a performance-based research funding system (cf. Jablecka, 1997). Further, Sokolov (2020, 2021) argues that in Russia, the use of quantitative indicators was an expression of distrust by the state of scientists' capacities to act as evaluators. Moreover, the study of research evaluation history in Russia and Eastern Bloc countries is made more relevant by the fact that the current wave of Chinese modernization is marked by Soviet heritage.

Finally, building on the above three discussions, I go on to examine how research evaluation systems shape scholarly communication in contemporary academia, and how various practices evident in the evaluation game can be used as tools for understanding these changes.

Power and Its Dark Side

This book is the result of research carried out at the intersection of three perspectives. When I started to investigate research evaluation systems, I was a philosopher interested in the communication mechanisms and policy agendas that shape

academia in the context in which I worked. At a certain point, I became a social scientist who moved from philosophy to social sciences and investigated scholarly communication, science policy, and research evaluation. In this way, I combined two perspectives. First, the perspective of a researcher who works in academia and whose work is evaluated and performance measured. Second, the perspective of a social scientist who critically investigates research evaluation systems and scholarly communication as they are implemented through various bibliometric and scientometric indicators. In consequence, I was being measured at the same time as I was also measuring my peers, for instance, by showing how publication patterns in social sciences and humanities were changed by research evaluation systems (Kulczycki et al., 2018, 2020). Thus I investigated research evaluation systems and criticized some science policy instruments by arguing that the process of constructing the measures served as means of sustaining the evaluative power of the state. These two perspectives allowed me to understand my own situation in academia better.

At some point, however, my critique of the Polish research evaluation system was recognized and acknowledged by policy makers. Having conducted two research projects on the effects of research evaluation, I was asked to serve as a policy advisor and – for some period – as a policy maker. In other words, I was put in a situation in which I was able to use my critical studies and recommendations to suggest how the Polish research evaluation system could be improved. I decided to take on this task and help – in the eyes of some of my colleagues – "power's dark side" to reproduce an oppressive system which I had been examining and criticizing for years. This third policy perspective showed me how difficult it is to design, implement, and use various science policy tools and why each social intervention always triggers both resistance and other types of reaction.

The experience that I gained from being located at the intersection of three standpoints showed me first, that while a dialogue between the academic community and policy makers might be fruitful, it is a demanding task and second, that in their current form, policy instruments are strongly shaped by researchers' demands. Before joining this dark side of power, I had thought that all the poorly designed policy instruments were the product of policy makers' own work and that they never listened to researchers and did not care about their opinions. Later, I learned that policy makers often design solutions exactly as suggested by the academic community for which "my own field's perspective" is the only acceptable perspective. The problem is that some researchers are not aware of the structural effects produced by changing only one element of the system, and it is therefore difficult for them to formulate useful recommendations for policy makers. On the other side, policy makers do not always understand that the solutions suggested by natural scientists might not work in the humanities and they should therefore