Serendipities
Journal for the Sociology and History of the Social Sciences

ARTICLE

Making Rankings Credible: How Ranking Creators Navigate Trials Of Strength

Elisabeth Strietzel and Can David Tobias

elisabeth.strietzel@uni-bielefeld.de, can_david.tobias@uni-bielefeld.de

Abstract

In this article, we take a closer look at the production of social knowledge beyond conventional academic sites in two NGOs that produce rankings of nation states. Whereas extant literature on rankings has mainly focused on the critique of methodologies or the inherent power of numbers, we suggest viewing them as webs of practices aimed at generating credible knowledge about the world. Using Latour's concept of trials of strength and applying it to two in-depth case studies, we investigate these practices empirically to unpack how these organisations attempt to create epistemic credibility.

Keywords

Rankings, trials of strength, knowledge production, credibility practices

INTRODUCTION

Rankings and similar evaluation practices have proliferated in many different areas of social life (see Beaumont and Towns 2021; Ringel et al., 2021a). Global governance is no exception: global performance indicators, including rankings of nations, have become increasingly common since the 1990s (Kelley and Simmons 2019). For example, the Global Benchmarking Database by the University of Warwick listed a total of 334 global indicators in 2021.[1] Many areas of global policy making are marked by quantitative indicators, and their relevance as tools of governance that provide orientation in an increasingly complex global world has been recognised by social scientists (Cooley 2015).

The proliferation of global performance indicators in recent decades has raised the question of how to make sense of them from a social scientific perspective. A lot of writing has focused on the task of debunking, problematising, and calling into question the credibility of global policy rankings, stressing either the questionable methodological foundations, the choice of what is measured, the veiled power structures behind these indicators, or other features deemed problematic (e.g., Beaumont and Towns 2021; Bhuta et al., 2018; Merry 2016). What has been less present is the

1 https://warwick.ac.uk/fac/soc/pais/research/csgr/benchmarking/database (03.05.2024)

question of what the creators of these rankings do on a practical level to make them credible for policy and media reporting.

In this paper, we therefore take a different approach. Drawing on two case studies, we investigate how ranking creators work to create epistemic credibility – that is, how they support their claims of representing the world and try to make them count as credible knowledge. Instead of treating them as problematic or non-scientific *by definition*, we propose an empirical perspective treating rankings as practices of social knowledge production. Drawing on the definition of social knowledge by Camic et al., (2011), which includes non-academic agents like think tanks or policy institutes, we investigate questions on the specific practices in which the people behind these devices engage to make their rankings count as credible knowledge.

To address this question, we borrow a concept from Bruno Latour's sociology of science: *trials of strength* (1987, 1990). This concept implies a practice-based understanding of knowledge production. Objectivity is not attributed or denied by a sociological observer, but rather treated as a contingent outcome of negotiations and struggles among participants in particular settings. Analogous to a scientific experiment, the rankings we study can be seen as claims to knowledge: the people behind them invest considerable resources into forming durable networks of practices. This involves developing arguments, defining criteria, collecting data, observing and investigating their respective political environments, anchoring their judgments in defensible principles, or telling a coherent story about their findings and why they matter. By transferring this concept from scientific practices to the making of rankings, we align with Camic and colleagues in using the conceptual tools developed in Science and Technology Studies to make sense of broader forms of "social knowledge practices" (Camic et al., 2011:2). Rather than focusing on whether rankings produce objective facts, we will demonstrate that this quality is the outcome of continuous practical achievement that is "relative to trials of strength in specific settings" (Latour 1987:78). The ranking creators have developed different practices aligned with Latour's conceptualisation of trials of strength. At stake in these trials is the credibility of the rankings: do they represent credible – that is, in this instance: objective – knowledge or not?

To address this question, we proceed in several steps. First, we review how existing literature on rankings and indicators has approached the question of their epistemic credibility. We argue that fixating on particular aspects of rankings, such as their (imperfect) methodologies or their use of numbers, has led many researchers to overlook the complex practical work carried out by the creators to ensure their credibility and acceptance as knowledge in political arenas. Second, we briefly introduce the cases we studied – two rankings published by relatively small civil society organisations addressing climate change and financial transparency respectively – as well as the methodological approaches we used to study them. Third, we analyse how claims to epistemic credibility are articulated in the two cases, building on the concept of trials of strength. We look at how the ranking creators work to pass trials of strength in different regards: (1) by explicitly *situating* them in their respective policy environments, (2) by telling a specific story about the world that underpins the *operationalisation* of their indicators, (3) by consciously *managing the non-standardised* elements of their production and (4) by *recruiting other actors* to solidify credibility. We close by reflecting on the implications and further connections of this approach and how it can be used to better understand the production of rankings and other forms of social knowledge.

THE EPISTEMIC STATUS OF RANKINGS AND INDICATORS IN THE LITERATURE

Social scientists have investigated different aspects of rankings and indicators. Authors have pointed out the "reactivity" and side effects of law school rankings (Espeland and Sauder 2007), the embedded power relations and how they are shaped by particularistic worldviews (Rottenburg and Merry 2015), or how rankings have fostered competitive relationships among actors (Brankovic et al., 2018). The focus of this literature encompasses several key areas, including the history, the effects, the underlying assumptions behind the production, and the epistemic nature of rankings.

For this paper, we focus on contributions that address the question of epistemic credibility. We identify three strands of literature: (1) we consider arguments from what we call a "methodology-focused perspective" on rankings, which is made up mostly of scholars from international relations and political science (e.g., Cooley 2015; Kelley and Simmons 2019; Beaumont and Towns 2021); (2), we discuss a line of thinking we call the "power of numbers perspective", which comes mostly from anthropology and sociology (e.g., Rottenburg and Merry 2015; Merry 2016; Bhuta et al., 2018); and (3) we turn to contributions we label "rankings as social practice", which aligns with our perspective but mostly focuses on university rankings (e.g., Mehrpouya and Samiolo 2016; Brankovic et al., 2022; Hamann and Schmidt-Wellenburg 2020).

METHODOLOGY-FOCUSED PERSPECTIVE

The methodology-focused perspective mostly focuses on the epistemic status and the scientific quality of indicators with the goal of improving this quality (Cooley 2015); offering caution against their potentially harmful effects (Kelley and Simmons 2019); or sometimes simply describing them as "bad science" (Broome et al.; 2018:516). A common idea is that rankings are insufficient when measured against scientific criteria (Cooley and Snyder 2015; Beaumont and Towns 2021).

While studies in the methodology-focused perspective have rightfully pointed out the shortcomings of many indicators, we find that it tends to be too narrow to grasp them as a social phenomenon with its own internal complexity and empirical variation. While rankings do indeed claim epistemic credibility, focusing on critiquing and improving the methodology of rankings overlooks the social dynamics and processes. Our approach differs from this perspective in that we are interested in the specific practices and strategies in which ranking-producing organisations engage in order to make it possible for their products to count as credible knowledge.

POWER OF NUMBERS PERSPECTIVE

The second perspective we identify is the power of numbers perspective. It focuses on the use of numbers and quantification as the primary factor in making rankings count as credible knowledge. According to these scholars, numbers possess a certain inherent quality of persuasion and objectivisation (Broome and Quirk 2015). There is a widespread notion that indicators and rankings are articulations of certain (problematic) macro trends, such as neoliberalism (Davies 2015), audit culture (Shore and Wright 2015a, 2015b), or new public management (Desrosières 2015), though the latter two are often seen as sub-trends of the former. Connected with this point is a recurring emphasis on the power structures ingrained in numeric evaluation – while numbers tend to appear as neutral representations of the world, they are all but neutral, as ideology shapes what gets counted (Merry 2016).

While these studies have important merits in sensitising us to the embedded political agendas of rankings, we argue the social processes that lead to making rankings credible are not reducible to the

effects of numbers alone. Instead, practices of quantification are part of a web of practices involved in creating robust claims (Latour 1987). One example is the practical work invested by some ranking organisations to explain their reasoning behind the numbers in the methodology papers, or press releases. As we show in the empirical section of this paper, numbers often do not speak for themselves, but have to be embedded in narratives about social problems and their solutions to become relevant in political arenas.

Likewise, treating rankings as mere manifestations of audit culture or neoliberalism tends to overlook empirical variations in the networks of practices (see Wilbers and Brankovic 2023). To make this clearer, consider the differences between the PISA rankings published by the Organisation for Economic Cooperation and Development (OECD) on one hand, and the Financial Secrecy Index discussed at more length in this paper, which was initially published by a small group of activists and designed as a critique of economic inequality and the prevailing discourse on tax evasion (Hansen and Vestergaard 2018). Whereas one may plausibly argue that the former represents a certain form of neoliberal thinking,[2] it would hardly be convincing to classify the latter as such. It is part of a political project that centres on inequality as a social problem and advocates for a robust tax state as the solution. Treating both indicators only on the basis of their utilisation of numbers, while disregarding the entire web of institutions and practices around them, risks losing the potential of empirically investigating the specific ways in which practices are employed to produce and maintain rankings in specific cases.

SOCIAL PRACTICES PERSPECTIVE

Finally, the rankings as social practices perspective is the closest to our own: rankings are neither seen exclusively as forms of knowledge that need to be improved by social scientists, nor as statistical technologies of power, but rather as bundles of practices that constitute an interesting object of sociological inquiry. In particular, this literature emphasises the need for studies on the level of organisational practices (Brankovic et al., 2022). Some studies focus on practices of legitimisation such as "discursive work" and the ability of ranking-producing organisations to respond convincingly to criticism (Hamann and Ringel 2023) or their ability to link global with national policy arenas (Hamann and Schmidt-Wellenburg 2020). With a few exceptions though, the bulk of these contributions focuses exclusively on rankings and indicators in higher education. While we broadly align with the conceptual approach of these studies in focusing on organisation-level practices, we apply this perspective to global policy rankings, which have, so far, rarely been studied in this way (though see Mehrpouya and Samiolo 2016 for a noteworthy exception).

THE CASE STUDIES: CLIMATE CHANGE PERFORMANCE INDEX AND FINANCIAL SECRECY INDEX

The findings presented here are a result of two case studies that we have conducted on the Climate Change Performance Index and the Financial Secrecy Index. Both cases have in common their production by relatively small NGOs and can be characterised as "boundary organisations" (Medvetz 2012), meaning they wield a particular form of influence that depends on a mixture of ties to different societal domains, such as politics, media, or academia. The rankings can be seen as "public interventions" (Eyal and Buchholz 2010) that leverage expert knowledge - among other things - to influence public discourse and ultimately shape policy decisions in their respective areas.

2 Though this would still not be a sufficient explanation for its success, in our view.

The Climate Change Performance Index (CCPI) is a ranking that has been published yearly since 2005 by Germanwatch e.V. and is currently published in cooperation with the NewClimate Institute and Climate Action Network International. The aim of the publication is to be "an independent monitoring tool for tracking countries' climate protection performance"[3]. To this end, the 63 countries and the EU responsible for more than 90 percent of global greenhouse gas emissions (GHG-emissions) are assessed on the basis of a total of 14 indicators.

The Financial Secrecy Index (FSI) has been published by the Tax Justice Network biannually since 2009. According to its website, the FSI ranks "jurisdictions most complicit in helping individuals to hide their finances from the rule of law".[4] To do this, it combines a qualitative component, namely an assessment of countries' legal frameworks regarding financial transparency, with a statistical component called the "Global Scale Weight", which is a figure representing the number of financial services offered by a particular country to non-residents per year.

A commonality between the cases is the institutional position and the size of the organisations behind these rankings. Both Germanwatch e.V. and Tax Justice Network are non-profit organisations that rely on external, mostly project-based funding, for which they have to apply on an ongoing basis. To secure funding, they have to demonstrate the impact of their work and frame it as being oriented towards a specific goal (Krause 2014). Impact, in this sense, refers to the engagement by relevant stakeholders, such as other civil society actors, government officials, and the media. Both rankings are produced by teams comprised of less than 15 people.[5] The rankings have been continuously published for more than ten years and therefore allow insights into how ranking creators work to make sustained publication possible (see Ringel and Werron 2021).

Unlike official intergovernmental organisations, the funding of these non-profits is not guaranteed, and they do not command the resources, staff, or reputation of organisations like the International Monetary Fund (IMF) or the World Bank. Scholars have observed that even the IMF, a large intergovernmental organisation, relies on its reputation as an expert institution to influence member countries (Barnett and Finnemore 2004). This is even more relevant for small non-governmental organisations, which rely on their credibility as knowledge producers, while simultaneously lacking the resources, prominence, and direct channels to governments typically possessed by intergovernmental organisations.

METHODS

Our empirical approach is characterised by the triangulation of different methods (Flick 2011) and data sources used to gain a comprehensive understanding of the practices of ranking production (Schatzki 2012). In line with the dictum "follow the actors" (Latour 2007), we draw on different data types to make sense of how ranking creators assemble their knowledge claims across different sites – including, for example, internal discussions about press releases as well as public-facing communication on the ranking's website. To generate insights into the daily routines and activities of the participants, each of the authors conducted ethnographic fieldwork over several months at one of the organisations described above, producing observation protocols of everyday work (Breidenstein et al., 2013). This approach was complemented by semi-structured, expert interviews

3 https://ccpi.org/ccpi-philosophy-team/ (03.05.2024)

4 https://fsi.taxjustice.net/ (03.05.24)

5 The number varies each year depending on available resources and the specific time of the production process. Both ranking teams use temporary workers like interns to support the core team.

to gain access to participants' perspectives on their work (Kaiser 2014), and practice-oriented document analysis (Asdal and Reinertsen 2022) to integrate the many written documents produced by our research objects.

Aligning with our methodological approach, the argument we develop in this article is based on a mixture of different types of data. The participant observations, interviews and documents have informed our thinking about ranking production. For the analysis presented in this paper, we mainly draw on publicly available documents, complementing these with other data types where needed. In short, the presented analysis would not have been possible without the contextual knowledge acquired through field work.

The ethnographic data was collected in two phases in each of the cases allowing for an initial phase of open observation and a second phase of selective and focused observation. For the FSI, virtual fieldwork lasted eight months in total. During these phases, Can David Tobias was actively involved in the ranking production work on a part-time basis,[6] conducting participant observation and generating fieldnotes and observation protocols. In total, 20 interviews were completed, 14 with members of Tax Justice Network and six with external informants. In the case of the CCPI, the participant observation was conducted by Elisabeth Strietzel over two phases, each lasting three to four months. The first phase entailed full-time participation as a member of the ranking team, while the second phase involved multiple observation periods across the ranking cycle, with no active involvement in the ranking production. For the CCPI, 27 interviews with 18 different informants were conducted. In both cases the organisations and participants were aware of the conducted research.

Our interpretation is based on an in-depth examination of the material, followed by multiple cycles of interpretation and discussion between the two authors. To facilitate systematic insight across the data, we employed the qualitative analysis software MAXQDA to code the data, including field notes, transcribed interviews, and collected documents. Crucially, the comparison and discussion of similarities and differences across the two cases enabled us to identify common themes and interesting variations. The empirical categories that we present in the analytic part of this paper are the result of an ongoing comparative discussion between our cases (Kelle and Kluge 2010). The goal of our research design is to conduct two explorative case studies (Savin-Baden and Major 2012:154f.) in which to explore the research gap regarding the sociological understanding of the production of (nation state) rankings. While this in-depth view allows us to gain insights into the inner working of ranking production, this study design does not allow insight into the effect of the two rankings, an outside perspective on the ranking, or insights into commonality of the found practices across different rankings.

THE CLIMATE CHANGE PERFORMANCE INDEX AND THE FINANCIAL SECRECY INDEX IN TRIALS OF STRENGTH

In the following section, we turn to the empirical analysis of how the epistemic credibility is constructed in the cases we studied. We use Latour's concept of trials of strength (1987; 1990), applying it to global policy rankings. This lens helps us to view epistemic credibility (or objectivity) not as a fixed fact but as the result of ongoing efforts and negotiations by actors. According to Latour,

6 The production of the FSI takes place entirely on a virtual basis, meaning all members work from their home office and interact through digital devices like Zoom, Slack or Sharepoint. Part time means that the ethnographer was involved in the organization for 3 days per week on average.

each (scientific) claim to credible knowledge undergoes three different trials of strength that are anticipated by actors, in which the robustness of the claim is put to the test: one on the level of narratives, one on the level of evidence, and finally one in front of an audience where all three are put to test. The gist of this triangular conception is explained in this way by Latour: "An experiment is, however, none of these trials alone. It is the *movement* of the three *taken together when it succeeds or separated when it fails*" (Latour 1990:61, emphasis in original). Actors therefore need to anticipate all three trials to assemble a robust claim.

In our understanding, trials of strength do not only refer to *public* trials in which a claim is tested and other actors attempt to dismantle it, but also to the practices of anticipating dissent devised by the author of a claim. As experienced members of their political arenas, the organisations we studied are well aware of the public scrutiny they could face (and have faced in the past). Therefore, their strategies of preparing the rankings to withstand this scrutiny are as much part of their trials of credibility as the public challenges they face.[7] The former is the focus of this paper.

With this understanding of how epistemic credibility can be constructed by actors, we look at how the ranking creators we studied prepare the ranking to resist trials of strength in four different regards (see table 1): gaining credibility by *connecting* to higher facts in their respective policy environments; *balancing* mechanical objectivity and trained judgment; devising a specific *reasoning* in their design of indicators that rests on a narrative about how the world works; and *recruiting* allies to solidify the credibility of the rankings.

Table 1: overview of empirical practices

Trials of strength (Latour 1990)	**Cross-case practices**	**CCPI**	**FSI**
Evidence	Connecting claims to an established fact	Keeping nation states accountable to the Paris Agreement	Introducing fairer, more universal criteria than OECD's tax haven blacklists
	Managing non-standardized judgment	Externalizing judgment	Proactive transparency of non-standardized judgment
Narrative	Reasoning behind operationalization	Climate mitigation as a responsibility of nation states	Scale of financial business defines responsibility of nation states
Audience	Recruiting allies for publication	Mobilizing partners from civil society	Mobilizing partners from civil society

To set the stage for the analysis, we first give one example of criticism faced by the two rankings in the past, showing the contested nature of their epistemic credibility. Following this, we show how trials of strength play out in our two cases.

7 Latour also analyzes practices of anticipation with regards to his concept of trials of strength, and not public trials exclusively (see 1987; 1990)

WHY TRIALS OF STRENGTH MATTER IN RANKING PRODUCTION

To understand this drive for objectivity, it is essential to comprehend the scrutiny to which the rankings are subjected. It is only through the ability to effectively anticipate criticism that ranking teams are able to attain a status of expertise (see Ringel et al., 2021b). By passing trials of strength the rankings can be regarded as credible. The following example demonstrates how one of the two cases is subjected to criticism and how this question of credibility is relevant to ranking creators.

In 2019, the Australian Prime Minster challenged the credibility of the CCPI. This came after the country received the worst climate policy rating of ranked countries a few weeks earlier.[8] When asked about the CCPI report in a press conference on, inter alia, the climate policy of the country, the following exchange took place:

> JOURNALIST: Prime Minister, just on climate change, you said there's a global effort, but Australia ranks last in the world on climate policy in a new global index. Isn't that an indictment of your Government's response?
>
> PRIME MINISTER: No, I completely reject that report. We don't accept that.
>
> JOURNALIST: You don't accept the report?
>
> PRIME MINISTER: No.
>
> JOURNALIST: Why not?
>
> PRIME MINISTER: Because I don't think it's credible.[9]

This criticism can be seen as part of trials of strength about the credibility of the ranking. The journalist uses the CCPI to question the Australian Prime Minister's claims of adequately addressing climate change. The ranking results continued to be used by other actors to call for political action from the Australian government which indicated that the critique was not successful. Just a few months later, Greenpeace included the ranking results in a press release about protests at Australian embassies and consulates to claim that the Australian government is not taking sufficient actions on climate change.[10]

This example is not an isolated event in the history of both cases, nor in that of rankings more generally. As Ringel and colleagues (2021b) have shown, this dynamic also applies to rankings more broadly, as many ranking-producing organisations devise strategies to react to and resist public contestations. (Policy) rankings can therefore be seen as a form of social knowledge characterised by its predisposition to attract criticism as well as by conscious efforts to anticipate this and defend itself. The example illustrates why the question of credibility is important for ranking creators. If the specific ranking is not seen as credible, it cannot be used as a tool for advocacy work. This demonstrates the need for epistemic authority of rankings to intervene in political discourse.

8 https://www.theguardian.com/australia-news/2019/dec/13/author-of-report-ranking-australia-worst-on-climate-policy-hits-back-at-pms-claim-its-not-credible (03.05.2024)

9 https://web.archive.org/web/20210115174305/https://www.pm.gov.au/media/press-conference-melbourne-vic (03.05.2024)

10 https://www.greenpeace.org/international/press-release/28338/greenpeace-holds-global-protest-telling-australian-government-act-on-climate-phase-out-coal/ (03.05.2024)

CONNECTING TO A MORE ESTABLISHED FACT

In *Science in Action* (1987: 74-79), Latour describes how a claim is challenged by a "dissenter" on individual parts of an experiment. As the dissenter scrutinises one element of the argumentative chain after the other, the claim is defended by *linking* the claim to a principle, idea, or fact, which has already been *established*, so that the dissenter is forced each time to challenge the accepted principle. Latour argues that a claim to epistemic credibility gets validated (or objectified) when its link to more widely accepted claims is maintained:

> Every time the visitor [i.e., the dissenter] followed a lead he reached a point where he had either to quit or start a new controversy about a still older and more generally accepted fact… This means that when the doubter tries out the connections, all these other facts, sciences and black boxes come to the Professor's rescue (Latour 1987:77–78).

Looking at the Financial Secrecy Index, we can find an example of how this plays out in the case of a ranking. Proponents of the Financial Secrecy Index often argue that it constitutes a 'non-political' ranking, relying on 'objectively verifiable criteria'.[11] Yet, these statements do not stand for themselves, but can be understood by an informed observer as performative speech acts that "do" something in social reality that goes beyond their propositional content (Austin 1971): they implicitly distinguish the FSI against a specific other, the tax haven blacklists.[12] These blacklists were first introduced by the OECD in the late 1990s and represented the first attempt by a major international organisation to directly name specific countries for their alleged complicity in transnational tax evasion. By singling out smaller jurisdictions beyond the financial centres of the world – and especially by not including well-known tax havens like Switzerland, Luxemburg or the US state of Delaware – these blacklists caused major political turmoil and a strong backlash from a variety of actors, including tax experts from civil society (Mayne and Kimmis 2000; Christensen 2007). A key point in the critique of the blacklists was their neglect of more powerful countries: even though OECD member states like Luxemburg or Switzerland were known tax havens, they did not appear on the blacklist. By contrast, less powerful countries like Liberia, Panama or Samoa were included. Critics took this as proof that the lists did not apply the same criteria to everyone, and instead were politically tainted (Dean and Warris 2020). This explains the emphasis on being a 'non-political' ranking on the part of the Financial Secrecy Index.

The following quote by Alex Cobham, chief executive of Tax Justice Network, in which he criticises the EU's ongoing blacklisting effort in 2017, exemplifies the way the Financial Secrecy Index is explicitly framed as the better – that is, more impartial – alternative to the blacklists:

> There's a long and largely ignominious tradition of tax haven blacklists, mainly at the OECD and IMF. They've tended to be subjective efforts, naming economically smaller jurisdictions with less political power, and steering well clear of major financial centres – regardless of their behaviour. The Tax Justice Network established the Financial Secrecy Index in 2009, largely as

[11] These wordings were used frequently by participants in conversations and interviews with the ethnographer as well as in public communication, for example this blog: https://taxjustice.net/2017/11/27/blacklisting-the-eu-paradise-lost/ (03.05.2024))

[12] The concept of a blacklist establishes a discrete, two-tier hierarchy between 'good' and 'bad': in the case of the tax haven blacklists, a number of countries are listed as non-compliant, while those not appearing on the list are implicitly marked as compliant. Those not on the list are thereby 'whitelisted', i.e., not considered to be doing anything wrong (hence the backlash against the non-listing of countries like Switzerland or Luxemburg). Thus, in contrast to a ranking, what matters is not *where* a given country is on the list, but whether it *is* on the list at all.

> an attempt to put transparent and objectively verifiable criteria in place, to allow a fair comparison of jurisdictions – a level playing field, if you like.[13]

In more abstract terms, the critique of the blacklists revolved around the question of fairness and universalism. While claiming to represent a global-universal view on the issue of tax evasion, in the eyes of many observers they were clearly tainted in the interest of powerful OECD member countries. Hence the critique of a "political" list that is not based on fair and universal criteria, but a mere articulation of pre-existing power constellations: if you are powerful enough, you can do what you want and still not land on the list. Instead, what was needed according to the critics, was a "level playing field" that treated all countries in an equal way and applied the same criteria, regardless of their political influence.

It is precisely this argument that represents one of the leading legitimisation strategies of the Financial Secrecy Index. By placing both smaller countries as well as the biggest economies of the world on a "secrecy spectrum" (Cobham et al., 2015:283) while considering the scale of their financial business, the creators of the Financial Secrecy Index translated the critique of the blacklists into an argument for the credibility of their ranking. In this way, the creators of the Financial Secrecy Index connect it to a more established fact: that of fairness and equal treatment. The language of a "non-political list" and "objectively verifiable criteria" represents this connection.

In contrast to the FSI, the CCPI is embedded in a very different way in its political environment. The methodological construction of the CCPIs indicators is aligned with the 2015 Paris Agreement, which has been the basis of the latest methodology used since 2017. The Paris Agreement marks a departure from the Kyoto Protocols approach of climate action focusing on regulation to "a framework for making voluntary pledges that can be compared and reviewed internationally, in the hope that global ambition can be increased through a process of 'naming and shaming'" (Falkner 2016:1107). Linking the methodology of the ranking to this UN institution and to the consensus reached by nation-states during negotiations, the CCPI translates this consensus into the definition of targets to assess the performance of countries in climate mitigation. This illustrates how the CCPI was reshaped to mobilise international and domestic pressure to push for the pledges that lay at the heart of the Paris Agreement.

One aspect in which the connection between the CCPI and the Paris Agreement can be traced throughout the ranking's publication is the phrasing "well below 2°C". This is a phrasing based on the Paris Agreement designating the goal set during the climate negotiations to limit global average warming to this target (Leemans and Vellinga 2017). Since the Paris Agreement, this phrasing can be found in the texts published alongside the ranking. Referring to this phrasing as a mark of successful climate action, the CCPI team incorporates the ranking into the broader framework of global climate efforts.

Compared to the FSI, which creates a connection to an established concept in its policy field through critique, the CCPI creates a connection by aligning the ranking criteria with the established Paris Agreement. This means that for this aspect, in order to call the criteria of the CCPI into question, critics would have to either show how they do not accurately represent the goals of the Paris Agreement – i.e., sever the links of the CCPI's claim with the more established fact of the agreement – or they would have to question the Paris Agreement itself, and subsequently find themselves in an

[13] https://taxjustice.net/2017/11/27/blacklisting-the-eu-paradise-lost/ (03.05.2024)

entirely different discursive arena. In Latour's terms, the Paris Agreement acts as an ally to the CCPI strengthening the rankings in a trial of strength.

By comparison, the FSI does not tie itself as directly to established institutions in its policy field, but rather draws its main rationale from criticising the tax haven blacklists. At the same time, however, we can see how the index also connects itself with a "more established fact": by incorporating the *level playing field*-argument into its methodology, the FSI invokes the universalistic discourse of global justice and fairness. Again, critics would have to either show how the FSI does not accurately represent this principle, or they would have to challenge the principle of a level playing field itself – that is, in this case, of equal rules applying to all countries.

REASONING BEHIND OPERATIONALISATION

Another aspect of the credibility of rankings is the reasoning behind the selected indicators. It refers to why the indicators are chosen, what is measured, and how this relates to the evaluated issue. As we have discussed above, scholars in the powers of numbers-perspective have rightfully pointed out that there are always contingent decisions involved in choosing what is measured and how it is weighted. In the following section, we look at how this reasoning is used in strengthening the claim to credibility. This reasoning can then also be subjected to tests of strength, as critics can call into question the connections it makes. If the connections do not hold, the credibility of the ranking is dismantled.

One way of gaining credibility through the reasoning behind the operationalisation, and selection of indicators, can be seen in how the rankings aim to record improvements over time. This aspect of the social practice of rankings has been pointed out by Ringel and Werron (2021). As serial comparisons – repeated publishing with updated data – rankings create comparison over time. The competition is therefore not only between the ranked entities, but also between current and past performance in the ranking.

In the case of the CCPI, the rankings methodology is aimed at capturing the process behind lowering GHG-emissions. The CCPI consists of four different categories of indicators: (1) GHG-emissions, (2) energy use, (3) renewable energy, and (4) climate policy. Each of these categories is measured by a set of indicators designed to assess current status and trends over time. The category of GHG-emissions is weighted the most with 40 percent, reflecting the notion that this is the category that ultimately needs to improve to mitigate climate change. But rather than only measuring GHG-emissions, the ranking team includes three other categories (weighted 20 percent each). This choice of indicators and their weight in the evaluation contain the way in which the ranking team envisions climate mitigation. It makes clear that the climate crisis should be addressed by nation states through climate policy aimed at lowering energy use, expanding renewable energy, and ultimately lowering GHG-emissions.[14] This is not implicitly included in the construction of the ranking but expressed in documents accompanying the rankings publication, as found in this description:

> Whether these policies are effectively implemented, can be read – with a time lag of a few years – in the country's improving scores in the categories "Renewable Energy" and "Energy Use" and

[14] Technological improvements such as carbon capture, or individual consumption limitation like carbon footprints are therefore not seen as the main way to climate change mitigation. The CCPI ranking creators clearly emphasis the responsibility of nation states to act rather than to individualize the solution as is done in the concept of the carbon footprint.

> lastly in positive developments in the category "GHG Emissions"... This weighting scheme allows the CCPI to adequately capture recent changes in climate policy and newly achieved improvements on the way to reduce GHG emissions. As GHG emissions reductions are what needs to be achieved for preventing dangerous climate change, this category weighs highest in the index (40percent) (CCPI Background and Methodology, 2024:4).

In presenting these steps towards climate mitigation, the team proposes a model of measuring and promoting the process of climate action by national actors. They argue that this way of measuring accounts for the time and processes it takes to lower GHG emissions on a national level. This narrative of change needs to pass tests of strength to support the status of the ranking as an objective measurement of climate action.

Dissenters could sever this connection between the four categories and/or between the indicators used to represent these categories in questioning this narrative. If successful, the ranking would be marked as unreliable, but if this narrative survives trials of strength the ranking is one step closer to being seen as credible. For example, we could imagine a dissenter who contests the assumption that national climate policy is needed to promote the lowering of GHG emissions. Rather than focusing on nation states, one could point to needed technological advancements of carbon capture or the use of nuclear energy.[15] If this was successful at severing the link between climate policy and the other categories, the ranking itself would fail the trial of strength. To be successful in severing the connection the dissenter would therefore either need to present a stronger network or be able to dismember the connections the narrative proposes.

The FSI, on the other hand, promotes an image of tax evasion that is not only incongruent with other representations found in its policy area, but in part directly challenges them (Seabrooke and Wigan 2015). Similar to the critique of the tax haven blacklists described above, it also sharply contrasts with the Corruption Perception Index (CPI). The main point made by the proponents of the FSI regarding the difference between the two indices is that the CPI represents corruption as an individualised problem ultimately caused by national elites – overwhelmingly in countries of the global south.[16] The FSI is intentionally designed to incorporate this critique, as it aims to frame the problem of tax evasion not as one of individual countries, but rather as a shared global problem, to which countries contribute in varying degrees. According to this reasoning, the rank of a country does not reflect the transparency of financial legislation alone, but also its responsibility towards other countries. This is measured in the number of financial services provided to non-residents in a given period. On the level of indicator design, this ambition to include a measure of the responsibility of countries is reflected by the so-called global scale weights, yielding a very different picture of the problem in which countries like the US, Germany or Japan rank among the worst spots – contrasting with the CPI, which ranked Syria, Venezuela and Somalia in the three lowest spots in its 2023 edition.[17]

[15] While this section focuses on an *imagined* dissenter, there are many instances where the creators of the CCPI are confronted with real dissenters proposing alternative paths for climate change mitigation in general or directly challenging the choice of indicators. The creators of the ranking pay close attention to the different lines of argument and evaluate whether they need to change the choice of indicators. It is not necessary to understand one specific dissent in order to understand the motivation of the ranking creators to clarify their reasoning. Rather, using Latour's concept, we can understand that the creators of the ranking anticipate various (imagined) dissenters in the conceptualization and communication of the ranking, and therefore develop practices to make their claim defensible.

[16] https://www.taxjustice.net/cms/upload/pdf/0701_Mirror_Mirror_corruption.pdf (03.05.2024)

[17] https://www.transparency.org/en/cpi/2023 (03.05.2024)

The reasoning behind this choice has been called into question several times. In particular, the connection between a large number of financial services exports and a higher responsibility towards other countries to maintain transparent laws has been attacked. One example of this is a report by Cayman Finance, an organisation representing the financial sector of the Cayman Islands, published in 2020 after the jurisdiction had been ranked in the worst spot of the FSI.[18] Among other criticisms, the report argued that the number of financial services exports is in fact a measure of success, in that it shows how well a country has attracted foreign capital. Therefore, it argued, the FSI unfairly penalises countries with successful financial sectors (Mansour 2020). In an official statement published on its website, TJN responded to this challenge by argumentatively reinforcing the link between global scale weight and responsibility towards the global problem:

> Secrecy jurisdictions often twist this analysis [that includes the global scale weights, author's note] on its head, claiming they are being treated unfairly for 'successfully' attracting money from non-residents… These claims miss the central purpose of the Financial Secrecy Index's ranking of countries by their supply of financial secrecy as opposed to the secrecy of their financial laws: All countries have a responsibility to safeguard against financial secrecy. The more financial activity a country seeks to pull in from other countries' residents, the greater the responsibility that country has to make sure its financial sector is not abused by people to evade or avoid tax in other countries or to launder money they've obtained elsewhere through illegal means. (Mansour 2020)

The fact that TJN staff invest the effort to publicly react to these challenges and take up the point of the global scale weights underlines that the narrative is not self-evident and needs to be explained and defended against scrutiny – or in other words, to stand in trials of strength. The intervention by Cayman Finance can be seen as an attempt to sever the narrative connection between "contribution to the global problem tax evasion" and "size of financial sector", while TJN's reaction represents a reinforcement of that connection, as it is an integral part of the FSIs underlying narrative.

Taking both cases together, the credibility of rankings does not only rest on the relationships the ranking teams established with higher concepts within the policy field alone, but also on the way it succeeds in arguing for a reasoning behind the operationalisation. In the case of the CCPI, this can be found in the way the ranking imagines a way for nation states to transition to lower GHG emissions. The FSI, on the other hand, suggests a specific understanding of the problem of tax evasion as genuinely global, to which countries contribute in different measures and correspondingly carry more or less responsibility to make their financial systems transparent. Both of these reasonings are neither self-evident not completely consensual in their respective policy areas. They need to be defended and underpinned by convincing narratives that can be articulated each time the operationalisation is scrutinised.

MANAGING NON-STANDARDISED JUDGMENT

Another way in which the CCPI and FSI strive to create epistemic credibility is that they both make the problem of non-standardisable elements of the ranking process transparent. Many authors argue that measuring nation state performance according to standardised criteria is by definition problematic, as the world does not lend itself smoothly to standardisation (Timmermans and Epstein 2010; Merry 2016). In line with this, the ranking creators we studied devise practices to deal with

[18].https://caymanfinance.ky/2021/09/22/flawed-cayman-finance-challenges-credibility-of-tax-justice-networks-biannual-financial-secrecy-index/ (03.05.2024)

the problem that their standardised criteria and methodologies – which are, as we have shown above, a crucial prerequisite for their credibility claim – do not easily apply to all individual nation states in exactly the same way. Sometimes information is harder to find, ambiguous, or sometimes the particular situation might be different.

Borrowing from Daston and Galison (2007), we can reformulate this problem as the tension between mechanical objectivity on the one hand and trained judgment on the other. In the case of the FSI, this can be found in the internal evaluation practices of the ranking team. The CCPI, in contrast, externalises judgments to third parties allowing for so-called country experts to highlight issues specific to the respective nation state. In other words, they have both developed distinct strategies to manage the unavoidable need for non-standardised judgments.

In the case of the FSI, the tension between mechanical objectivity and trained judgment is addressed by conscious effort to make explicit where the limits of the former are reached and where the latter is needed. The need of referring to trained judgment in some instances is acknowledged in the following passage of the methodology document:

> In cases of conflicting information, we resorted to reasoned judgement – while recognising the necessary subjectivity of the approach. Where this was the case, therefore, we aim to provide full transparency about criteria and interpretation. As a result, in addition to references to all underlying sources, the database reports also include a large amount of supporting information and notes relating to data analysis (FSI Methodology 2022: 18).

For each datapoint that ultimately makes up the ranking, the FSI staff publishes a note aimed at explaining the reasoning behind that particular judgment. The majority of these notes refer to the standardised criteria that were applied to reach the specific score.

However, there are instances that do not allow a mechanised application of criteria and instead require extensive interpretive work to match a given legal situation with the standardised FSI criteria. One such example is the assessment of the implementation of a so-called beneficial ownership register for US companies in the FSI edition of 2022. The US had just passed a law introducing such a register – yet the FSI team ultimately decided it was insufficient as it had too many exemptions.[19] The judgment, which led to a poor score in the ranking, is accompanied by a multi-paragraph text on the index website providing a point-by-point explanation of the reasoning. As this example illustrates, the situations requiring such interpretive work are explicitly marked as such by the authors of the FSI. In other words, the limits of mechanised objectivity are acknowledged and there is a conscious effort of *managing* these situations in a proactive way. Thus, the situation in which non-standardised judgment becomes necessary is rationalised on the terms of the author – forestalling possible critique of subjectivity and aiming to strengthen the claim.

In the case of the CCPI, non-standardised judgments can be seen in the texts that are published alongside the ranking. These texts use what we call externalised judgment to explain the ranking results. This externalised judgment is collected in an annual survey sent out by the ranking creators to local partners. The local policy experts, most of whom work in climate change NGOs, answer questions on various aspects of their country's national and international climate change policies, such as fossil fuel phase-out, implementation of climate change laws, or land use policies. This survey is the basis for the policy indicators and informs the text explaining the ranking results of the

[19] https://fsi.taxjustice.net/country-detail/#country=US&period=22 (03.05.24)

evaluated countries. In the so-called country text, written for each of the ranked nation states, the input of the country experts is intertwined with that year's ranking results, and past performances.

Each text is written by the ranking team, reviewed by country experts, proofread by an editor, and finally published with the ranking on the launch day. The texts connect the quantitative indicators with the specific political context and demands of local NGOs. The evaluation used for this is based on the individual judgment of the participating country experts. This is emphasised in the country texts as the individual assessments of policy, and calls for concrete policies are attributed to country experts. Country experts are therefore assigned the capability to make a trained judgment about the country's climate policy while the ranking creators do not provide this kind of judgment. As one example, aspects of Denmark's general climate policy are highlighted as evaluations of the country experts: "All the experts agree that implementing a tax on agricultural production would be a crucial step towards lowering the country's high emissions in this sector".[20] Demands by the country experts for further action are then framed as a prerequisite for continued good performance in the ranking.

In giving external voices a platform to demand specific policy improvements, the ranking team make a clear distinction between their evaluation based on mechanical objectivity and the trained judgment of country experts. These trained judgments are therefore externalised and a differentiation is made between the expertise required for these judgments and the expertise of the ranking team. As such, the claim of credibility is strengthened as the ranking team extends the connection to other actors to support trained judgments and therefore makes clear that they are not experts in climate policy for all ranked nation states.

SOLIDIFYING CREDIBILITY BY RECRUITING ALLIES

Another element contributing to the credibility of the rankings is their endorsement by external actors. Around each edition's release, the ranking creators actively seek allies to buy into the claim and to multiply it within their respective context, especially national publics. Not only does it expand the public reach, but we argue it is vital for them to be reproduced by actors beyond the original creators to strengthen credibility. As Latour writes: "An idea or a practice cannot move from A to B solely by the force that A gives it, B must seize it and *move* it" (Latour 1988:15f.). Each positive reference made to the ranking by a local civil society actor, a journalist writing for a national newspaper, or a politician engaging in a discussion; solidifies the credibility of the ranking, and as such, actors make their endorsement visible to wider audiences. In this sense, we interpret the efforts of the ranking creators to find and mobilise allies to adopt the ranking in their own communication as a preparation for trials of strength with regards to audiences (Latour 1990:62).

In our cases, the presentation does not take place on a single stage, but on multiple stages, ranging from global political fora like the annual Climate Conference to local civil society actors using the ranking in their advocacy work. For each of these stages, the ranking creators try to find allies – mostly local civil society organisations, but also journalists – to act as intermediaries and amplify the ranking for each particular setting. On the part of the ranking creators, the main task in this regard is to recruit allies willing to adopt the ranking for their own projects and thereby put their weight behind its credibility (Callon 1984).

As Ringel (2021) has shown, the releases of rankings in today's world are often "painstakingly orchestrated performances" (ibid.: 56). Ranking creators invest considerable resources into planning

[20] See https://web.archive.org/web/20240409190443/https://ccpi.org/country/dnk/ (03.05.2024)

launch events, crafting visualisations, or finding storylines to align the rankings with current public debates. This is also true for the cases we studied. In the following paragraphs, we offer a brief overview of the practices devised by these organisations aimed at increasing engagement by external actors.

The CCPI team uses the yearly climate negotiation as a platform to hold a press conference at the COP venue. The ranking creator encourages local partner organisations to write and publish their own position about the ranking and its results. To make this easier for partner organisations, they offer a pre-release package to those who participate in the policy evaluation to give them time to prepare their own reaction to ranking results. The pre-release package includes information on the results of the ranking as well as graphics illustrating the results. It is not compulsory for partners to use these materials, but they are encouraged to use them as they see fit. This decentralised media strategy offers locally working partner organisations the benefit of preparing their own (social) media content for the publication of the ranking.

The team behind the FSI follows a similar strategy of publication, the main difference being the lack of a similar event to the climate conference for global tax governance. Prior to the official launch date of the Index, the in-house team of the Tax Justice Network uses various strategies of getting external actors on board. When preparing a global press release for the ranking, the team discusses which issues might be relevant to the transnational community of tax policy actors. At the same time, they reach out to regional and national partner organisations and journalists to share the release date and ranking results with them ahead of time. Additionally, the press team of TJN, similar to the CCPI, prepares so-called "partner kits" which summarise the ranking results for a given country. They also co-host region-specific launch events with some of these partners, presenting the central findings of the ranking to audiences mobilised by the local partners. The importance of getting these intermediaries to engage with the ranking is visible throughout the publication process.

Taken together, these publication practices underline the importance of engaging with external actors for ranking creators. In both cases, the creators invest significant resources to encourage intermediaries like national civil society partners or journalists to disseminate the ranking in their own name. This not only expands the range of the ranking and helps to make it relevant to more stakeholders but also reinforces its credibility. The concept of trials of strength thus helps us to understand the practice of recruiting other NGOs to the ranking, not only as a means of enhancing its visibility, but also as a strategy for expanding the network of defenders against criticism. Each positive engagement solidifies the credibility of the ranking, therefore getting external organisations to sign on to it is vital to strengthen its credibility.

CONCLUSION

In this paper, we have set out to study the question how ranking creators work to create epistemic credibility – that is, how they support their claims of representing the world and make them count as credible knowledge. Instead of assessing the credibility of the rankings ourselves, we have focused on the practices and strategies engaged by the actors to build credibility and convince their audiences. To this end, we have presented findings from two ethnographic case studies of rankings produced by globally-oriented civil society organisations: the Climate Change Performance Index (CCPI) and the Financial Secrecy Index (FSI).

We have shown how the people behind the CCPI and the FSI strive to create epistemic credibility in four ways. (1) By embedding claims in their respective political environments, ranking teams create

a connection to established facts which makes it harder to argue with them. (2) The process of managing non-standardised judgments illustrates how ranking producers account for aspects of the evaluation work that cannot be standardised while still protecting their claims against critique. (3) The reasoning of operationalisation is underpinned by specific narratives about the respective policy area, as they provide the connection between concepts, data, and imagined effect of the ranking. (4) The recruitment of other actors to support the publication of the rankings reinforces the claims of the ranking teams.

As we have stated, this is an explorative study that does not claim to be exhaustive. There are many more ways in which ranking creators construct credibility, as well as other actors and sites that have a part in trials of strength, and could be investigated in future research. Our main goal in this article has been to demonstrate the fruitfulness of applying practice-based and empirically open conceptions of knowledge, like Latour's notion of trials of strength, to rankings. By treating rankings as projects of knowledge production that can be studied using the conceptual tools developed in the sociology of science (Camic et al., 2011), we hope to contribute to a practice-based, empirically driven study of rankings. As Latour points out, the attributes of subjective and objective "are relative to trials of strength in specific settings" (Latour 1987:78). In line with this, we avoid subsuming the rankings under any pre-conceived label such as "quantification" or "neoliberalism", but instead focus on the concrete practices and strategies the ranking creators devise to pass these trials, including – but not limited to – the use of numbers. Moreover, we have suggested one should take the social position of the authors of such claims into account: in the cases we studied, their marginal institutional position as well as the comparatively small number of resources they command, require them to come up with creative strategies to pass trials of strength and create political traction.

Regarding the question of generalisation, we see this as a tentative step towards a practice-based understanding of rankings and their claims to credibility on the basis of two case studies. More systematic and comparative studies of rankings in different fields – like, for instance, higher education or the arts – are needed to gain a more general understanding of the various credibility strategies pursued by ranking creators. This article provides first insights into practices that ranking creators utilise to strengthen the credibility of rankings in trials of strength. Rankings are often characterised by scholars as "technologies of governance" (Merry 2011) or tools of "governing by numbers" (Shore and Wright 2015b). While we agree that researchers as well as journalists and other users of indicators should be alert to the political agendas implicated in them, we argue this does not apply exclusively to rankings, but to any form of social knowledge production which aims to influence public policy and media reporting. At the same time, we suggest to not always make this insight the starting point of empirical analysis of rankings. Instead, they can be studied through the lens of a practice-focused sociology of knowledge that carves out the strategies of strengthening and legitimising claims of credible knowledge. As we have shown, the studied ranking creators invest resources to establish their credibility and expertise, which is an essential building block in their attempt to influence governance.

By unpacking the credibility practices in our empirical cases of ranking creators, we contribute to a deeper understanding of these practices, and this understanding can be extended to other cases and areas of social knowledge production. How do actors who wish to make other actors or audiences buy into their claims equip those claims with credibility, so that they do not simply appear as articulations of opinion or ideology, but as knowledge about the world? In particular, the approach we have taken on rankings could also be applied to all kinds of policy reports and other outputs where various types of actors, like NGOs, interest groups, think tanks, or government agencies try to

intervene in political processes by providing expertise that goes beyond mere opinion. We would therefore argue that our approach of applying the concept of trials of strength to knowledge claims beyond academia is a promising avenue for future research on credibility practices in different forms of social knowledge production.

REFERENCES

Asdal, Kristin, and Hilde Reinertsen (2022) *Doing document analysis: A practice-oriented method*. Los Angeles, London, New Delhi, Singapore, Washington DC, Melbourne: SAGE.

Austin, John L. (1971) *How to do things with words: The William James lectures delivered at Harvard University in 1955*. Cambridge, Mass.: Harvard Univ. Press.

Barnett, Michael N., and Martha Finnemore (2004) The power of liberal international organizations, in M. N. Barnett and R. Duvall (Eds.) *Cambridge studies in international relations, Power in global governance*. Cambridge: Cambridge University Press, 161–184.

Beaumont, Paul, and Ann E. Towns (2021) The Rankings Game: A Relational Approach to Country Performance Indicators, *International Studies Review* 23(4): 1467–94.

Bhuta, Nehal, Debora V. Malito, and Gaby Umbach (2018) Introduction: Of Numbers and Narratives. Indicators in Global Governance and the Rise of a Reflexive Indicator Culture, in D. V. Malito, G. Umbach, and N. Bhuta (Eds.) *Palgrave handbooks, The Palgrave handbook of indicators in global governance*. Cham: Palgrave Macmillan, 1–30.

Brankovic, Jelena, Leopold Ringel, and Tobias Werron (2018) How Rankings Produce Competition: The Case of Global University Rankings, *Zeitschrift für Soziologie* 47(4): 270–88. https://doi.org/10.1515/zfsoz-2018-0118 .

Brankovic, Jelena, Leopold Ringel, and Tobias Werron (2022) Spreading the gospel: Legitimating university rankings as boundary work, *Research Evaluation* 31(4): 463–74. https://doi.org/10.1093/reseval/rvac035 .

Breidenstein, Georg, Stefan Hirschauer, Herbert Kalthoff, and Boris Nieswand (2013) *Ethnografie: Die Praxis der Feldforschung*. Konstanz, München: UVK-Verl.-Ges; UVK/Lucius.

Broome, André, Alexandra Homolar, and Matthias Kranke (2018) Bad science: International organizations and the indirect power of global benchmarking, *European Journal of International Relations* 24(3): 514–39. https://doi.org/10.1177/1354066117719320 .

Broome, André, and Joel Quirk (2015) Governing the world at a distance: the practice of global benchmarking, *Review of International Studies* 41(5): 819–41.

Callon, Michel (1984) Some Elements of a Sociology of Translation: Domestication of the Scallops and the Fishermen of St Brieuc Bay, *The Sociological Review* 32(1_suppl): 196–233. https://doi.org/10.1111/j.1467-954X.1984.tb00113.x .

Camic, Charles, Neil Gross, and Michèle Lamont (2011) The Study of Social Knowledge Making, in *Social knowledge in the making* in C. Camic, N. Gross, and M. Lamont (Eds.). Chicago, Ill.: University of Chicago Press, 1–40.

Christensen, John (2007) Mirror, Mirror on the Wall, who's the most corrupt of all?, (https://www.researchgate.net/publication/237827314_MIRROR_MIRROR_ON_THE_WALL_WHO'S_THE_MOST_CORRUPT_OF_ALL) (accessed: 10 Oct. 2024).

CCPI Climate Change Performance Index 2024: Background and Methodology (https://ccpi.org/download/climate-change-performance-index-2024-background-and-methodology/) (accessed: 3 May 2024).

Cobham, Alex, Petr Janský, and Markus Meinzer (2015) The Financial Secrecy Index: Shedding New Light on the Geography of Secrecy, *Economic Geography* 91(3): 281–303. https://doi.org/10.1111/ecge.12094 .

Cooley, Alexander (2015) The emerging politics of international rankings and ratings. A framework for analysis, in *Ranking the world. Grading states as a tool of global governance* in A. Cooley and J. L. Snyder (Eds.). Cambridge: Cambridge University Press, 1–38.

Cooley, Alexander, and Jack L. Snyder, editors (2015) *Ranking the world: Grading states as a tool of global governance*. Cambridge: Cambridge University Press.

Daston, Lorraine, and Peter Galison (2007) *Objectivity*. New York, New York State, Cambridge, Massachusetts: Zone Books.

Davies, William (2015) Spirits of neoliberalism: 'competitiveness' and 'wellbeing' indicators as rival orders of worth, in R. Rottenburg, S. E. Merry, S. J. Park, and J. Mugler (Eds.) *Cambridge studies in law and society, The world of indicators. The making of governmental knowledge through quantification*. Cambridge: Cambridge University Press, 238–306.

Dean, Steven A., and Attiya Waris (2020) Ten Truths About Tax Havens. Inclusion and the" Liberia" Problem, *Emory Law Journal* 70: 1659.

Desrosières, Alain (2015) Retoraction: how indicators feed back onto quantified, in R. Rottenburg, S. E. Merry, S. J. Park, and J. Mugler (Eds.) *Cambridge studies in law and society, The world of indicators. The making of governmental knowledge through quantification*. Cambridge: Cambridge University Press, 329–353.

Espeland, Wendy N., and Michael Sauder (2007) Rankings and Reactivity: How Public Measures Recreate Social Worlds, *American Journal of Sociology* 113(1): 1–40.

Eyal, Gil, and Larissa Buchholz (2010) From the Sociology of Intellectuals to the Sociology of Interventions, *Annual Review of Sociology* 36(1): 117–37.

Falkner, Robert (2016) The Paris Agreement and the new logic of international climate politics, *International Affairs* 92(5): 1107–25. https://doi.org/10.1111/1468-2346.12708 .

Flick, Uwe (2011) *Triangulation: Eine Einführung*. Wiesbaden: VS Verl. für Sozialwiss.

FSI Financial Secrecy Index 2022: Methodology. (https://web.archive.org/web/20231119065642/https://fsi.taxjustice.net/fsi2022/methodology.pdf) (accessed: 11 Nov. 2025)

Hamann, Julian, and Leopold Ringel (2023) The discursive resilience of university rankings, *Higher Education* 86(4): 845–63. https://doi.org/10.1007/s10734-022-00990-x .

Hamann, Julian, and Christian Schmidt-Wellenburg (2020) The double function of rankings. Consecration and dispositif in transnational academic fields, in C. Schmidt-Wellenburg and S. Bernhard (Eds.) *Routledge research in transnationalism, vol.* 38, *Charting transnational fields. Methodology for a political sociology of knowledge*. Abingdon, New York: Routledge, 160–177.

Hansen, Hans K., and Anne Vestergaard (2018) On the Contest of Lists and Their Governing Capacities. How 'Tax Havens' Became 'Secrecy Jurisdictions', in *Education by the numbers and the making of society. The expertise of international assessments* in S. Lindblad, D. Pettersson, and T. S. Popkewitz (Eds.). New York, London: Routledge, 35–52.

Kaiser, Robert (2014) *Qualitative Experteninterviews: Konzeptionelle Grundlagen und praktische Durchführung*. Wiesbaden: Springer VS.

Kelle, Udo, and Susann Kluge (2010) *Vom Einzelfall zum Typus: Fallvergleich und Fallkontrastierung in der qualitativen Sozialforschung*. Wiesbaden: VS Verlag.

Kelley, Judith G., and Beth A. Simmons (2019) Introduction: The Power of Global Performance Indicators, *International Organization* 73(03): 491–510. https://doi.org/10.1017/S0020818319000146 .

Krause, Monika (2014) *The good project: Humanitarian relief NGOs and the fragmentation of reason*. Chicago, London: The University of Chicago Press.

Latour, Bruno (1987) *Science in action: How to follow scientists and engineers through society*. Cambridge, Massachusetts: Harvard University Press.

Latour, Bruno (1988) *The pasteurization of France*. Cambridge, London: Harvard Univ. Press.

Latour, Bruno (1990) The Force and Reason of Experiment, in H. E. LeGrand (Ed.) *Australasian studies in history and philosophy of science, vol.* 8, *Experimental inquiries. Historical, philosophical and social studies of experimentation in science*. Dordrecht: Kluwer, 49–80.

Latour, Bruno (2007) *Eine neue Soziologie für eine neue Gesellschaft: Einführung in die Akteur-Netzwerk-Theorie*. Frankfurt am Main: Suhrkamp.

Leemans, Rik, and Pier Vellinga (2017) The scientific motivation of the internationally agreed 'well below 2 °C' climate protection target: a historical perspective, *Current Opinion in Environmental Sustainability* 26-27: 134–42. https://doi.org/10.1016/j.cosust.2017.07.010 .

Mansour, Mark Bou (2020) Slander and failed advertising campaigns how tax haven officials responded to the financial secrecy index 2020 (https://taxjustice.net/2020/03/05/slander-and-failed-advertising-campaigns-how-tax-haven-officials-responded-to-the-financial-secrecy-index-2020) (accessed: 3. May 2024).

Mayne, Ruth and Jenny Kimmis (2000) Tax Havens: Releasing the hidden billions for poverty eradication, (https://policy-practice.oxfam.org/resources/tax-havens-releasing-the-hidden-billions-for-poverty-eradication-114611/) (accessed: 10 Oct. 2024).

Medvetz, Thomas (2012) Murky Power: "Think Tanks" as Boundary Organizations, in D. Courpasson, D. Golsorkhi, and J. J. Sallaz (Eds.) *Research in the sociology of organizations,*

vol. 34.2012, *Rethinking power in organizations, institutions, and markets*. Bingley: Emerald, 113–133.

Mehrpouya, Afshin, and Rita Samiolo (2016) Performance measurement in global governance: Ranking and the politics of variability, *Accounting, Organizations and Society* 55: 12–31. https://doi.org/10.1016/j.aos.2016.09.001 .

Merry, Sally E. (2011) Measuring the World, *Current Anthropology* 52(S3): S83-S95. https://doi.org/10.1086/657241 .

Merry, Sally E. (2016) *The seductions of quantification: Measuring human rights, gender violence, and sex trafficking*. Chicago, London: The University of Chicago Press.

Ringel, Leopold (2021) Stepping into the Spotlight. How Rankings Become Public Performances, in L. Ringel, W. N. Espeland, M. Sauder, and T. Werron (Eds.) *Research in the sociology of organizations,* volume 74, *Worlds of rankings*. Bingley, UK: Emerald Publishing, 53–76.

Ringel, Leopold, Wendy Espeland, Michael Sauder, and Tobias Werron (2021a) Worlds of Rankings, in L. Ringel, W. N. Espeland, M. Sauder, and T. Werron (Eds.) *Research in the sociology of organizations,* volume 74, *Worlds of rankings*. Bingley, UK: Emerald Publishing, 1–23.

Ringel, Leopold, Julian Hamann, and Jelena Brankovic (2021b) Unfreiwillige Komplizenschaft: Wie wissenschaftliche Kritik zur Beharrungskraft von Hochschulrankings beiträgt, in S. M. Büttner and T. Laux (Eds.) *Leviathan Sonderband, vol.* 38, *Umstrittene Expertise. Zur Wissensproblematik der Politik:* 386–407.

Ringel, Leopold, and Tobias Werron (2021) Serielle Vergleiche: Zum Unterschied, den Wiederholung macht, *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie* 73(S1): 301–31 https://doi.org/10.1007/s11577-021-00750-2 .

Rottenburg, Richard, and Sally E. Merry (2015) A world of indicators: the making of governmental konlwedge through quantification, in R. Rottenburg, S. E. Merry, S. J. Park, and J. Mugler (Eds.) *Cambridge studies in law and society, The world of indicators. The making of governmental knowledge through quantification*. Cambridge: Cambridge University Press, 1–33.

Savin-Baden, Maggi, and Claire H. Major (2012) *Qualitative research: The essential guide to theory and practice*. Milton Park, Abingdon, Oxon, New York: Routledge.

Schatzki, Theodore R. (2012) A Primer on Practices, in J. Higgs, R. Barnett, S. Billett, M. Hutchings, and F. Trede (Eds.) *Educational Research E-Books Online, Collection 2005-2017, ISBN, vol.* 6, *Practice-Based Education. A Perspectives and Strategies*. Leiden, Boston: Brill | Sense, 13–26.

Seabrooke, Leonard, and Duncan Wigan (2015) How activists use benchmarks: Reformist and revolutionary benchmarks for global economic justice, *Review of International Studies* 41(5): 887–904. https://doi.org/10.1017/S0260210515000376 .

Shore, Cris, and Susan Wright (2015a) Audit Culture Revisited, *Current Anthropology* 56(3): 421–44. https://doi.org/10.1086/681534 .

Shore, Cris, and Susan Wright (2015b) Governing by numbers: audit culture, rankings and the new world order, *Social Anthropology* 23(1): 22–28. https://doi.org/10.1111/1469-8676.12098 .

Timmermans, Stefan, and Steven Epstein (2010) A World of Standards but not a Standard World: Toward a Sociology of Standards and Standardization, *Annual Review of Sociology* 36: 69–89. https://doi.org/10.1146/annurev.soc.012809.102629 .

Wilbers, Stefan, and Jelena Brankovic (2023) The emergence of university rankings: a historical-sociological account, *Higher Education* 86: 733–50. https://doi.org/10.1007/s10734-021-00776-7 .

ACKNOWLEDGEMENTS

We would like to thank Jelena Brankovic, Malte Neuwinger, Tobias Werron, Leopold Ringel, Jasmin Weissberg, Vivian Vollbrecht, the anonymous reviewer, the special issue editors Vincent Gengnagel, Christian Schmidt-Wellenburg and Frédéric Lebaron, the participants of the author workshop, and the participants of Tobias Werron's research class for their generous comments, feedback and contributions. In addition, we are grateful to the organisations for generously allowing us to take part in their work and study it over several months.

AUTHOR BIOGRAPHY

Elisabeth Strietzel is a PhD candidate at the sociological faculty of Bielefeld University. Her research focuses on the connection between practices of evaluation and comparisons; climate change; and global governance.

Can David Tobias is a PhD candidate at the Faculty of Sociology, Bielefeld University. His research focuses on the sociology of expertise, practices of knowledge production in policy-related contexts, and qualitative methods.