



Universiteit
Leiden
The Netherlands

Filling in the gaps: The interpretation of curricula vitae in peer review

Kaltenbrunner, W.; Rijcke, S. de

Citation

Kaltenbrunner, W., & Rijcke, S. de. (2019). Filling in the gaps: The interpretation of curricula vitae in peer review. *Social Studies Of Science*, 49(6). doi:10.1177/0306312719864164

Version: Publisher's Version

License: [Creative Commons CC BY-NC 4.0 license](#)

Downloaded from: <https://hdl.handle.net/1887/80181>

Note: To cite this publication please use the final published version (if applicable).

Filling in the gaps: The interpretation of *curricula vitae* in peer review

Social Studies of Science

1–21

© The Author(s) 2019



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/0306312719864164

journals.sagepub.com/home/sss**Wolfgang Kaltenbrunner** 

Center for Science and Technology Studies, Leiden University, Netherlands

Sarah de Rijcke

Center for Science and Technology Studies, Leiden University, Netherlands

Abstract

In this article, we study the use of *curricula vitae* (CV) for competitive funding decisions in science. The typically sober administrative style of academic résumés evokes the impression of straightforwardly conveyed, objective evidence on which to base comparisons of past achievements and future potentials. We instead conceptualize the evaluation of biographical evidence as a generative interplay between an historically grown, administrative infrastructure (the CV), and a situated evaluative practice in which the representational function of that infrastructure is itself interpreted and established. The use of CVs in peer review can be seen as a doubly comparative practice, where referees compare not only applicants (among each other or to an imagined ideal of excellence), but also their own experience-based understanding of practice and the conceptual assumptions that underpin CV categories. Empirically, we add to existing literature on peer review by drawing attention to self-correcting mechanisms in the reproduction of the scientific workforce. Conceptually, we distinguish three modalities of how the doubly comparative use of CVs can shape the assessment of applicants: calibration, branching out, and repair. The outcome of this reflexive work should not be seen as predetermined by situational pressures. In fact, bibliographic categories such as authorship of publications or performance metrics may themselves come to be problematized and reshaped in the process.

Keywords

comparison, CVs, evaluation, peer review

Correspondence to:

Wolfgang Kaltenbrunner, Center for Science and Technology Studies, Leiden University, P.O. Box 905, Leiden, 2300 AX, The Netherlands.

Email: w.kaltenbrunner@cwts.leidenuniv.nl

Introduction

The central mechanisms underpinning the reproduction of the scientific workforce in most countries – academic rituals such as tenure committee meetings and peer review for funding programs – involve processes of competitive selection (Lamont, 2009; Musselin, 2009; Whitley, 2000). Researchers deemed accomplished in a given area of study are tasked with assessing and selecting typically younger peers to become future leaders of institutions, projects, and people. These evaluative processes rely on instruments meant to inform comparisons between often large numbers of applicants. Chief among these instruments is the academic CV (*curriculum vitae*), which documents individual career trajectories in terms of categories that are abstract and partly quantifiable, such as authorship of publications, citation-based metrics, and previously acquired grants.

In this article, we study how referees mobilize personal experience of scientific practice in the interpretation of academic CVs. Peer review is often said to be a uniquely suitable form of quality control because of the ability of scientists to contextualize the merits and potential of their colleagues on the fly (Nowotny, 2014; Polanyi, 1962). But how exactly does that work? How do referees draw on their understandings of scientific activity to qualify biographical evidence of applicants, and how does this affect what forms of comparison are enacted? Such an analytical focus, we argue, constitutes an important complement to the previous studies that have looked at how CVs, or parts thereof, such as bibliometric information, are used in peer review. Some studies have argued that CVs' information is used to reduce complexity of evaluative situations, for example by allowing applicants to be ranked on the basis of productivity or citation-based metrics (Cañibano et al., 2009; Hammarfelt and Rushforth, 2017; Musselin, 2009; Sonnert, 1995). While this captures an important potential use of CVs and indicators, it is equally important to avoid slipping into a functionalist perspective – we should not assume that referees will automatically engage in reductive forms of comparison when asked to take evaluative decisions under resource and time constraints. To be sure, previous studies point out that referees often reflect on what information from a given CV should be used or not used in a particular case, and that the expertise of reviewing may gradually be redefined as referees develop technical knowledge of indicators (Hammarfelt and Rushforth, 2017). This presupposes, however, that such a redefinition will leave untouched the substance of the categories and their indicators. Put differently, the emphasis on complexity reduction implies a more or less unilateral determination of evaluative practices through a stable framework for representing academic career trajectories, or, perhaps, a choice not to use this framework at all.

In this article, we conceptualize the evaluation of biographical evidence as a generative interplay between an historically grown, administrative infrastructure (the CV), and a situated evaluative practice in which the representational function of that infrastructure is itself interpreted and established (Krüger and Reinhart, 2017). We propose to think of the use of CVs in peer review as a doubly comparative practice, where referees not only compare applicants with each other or to an imagined ideal of excellence, but also to their own experience-based understandings of practices and the conceptual assumptions that underpin CV categories. This actually means that the relation between CV categories and scientific practice is renegotiated and updated in every review situation. The

outcome of this interplay should not be seen as predetermined by situational pressures, as biobibliographic categories may themselves come to be problematized and reshaped in the process.

Biographical evidence of scientists as a basis for peer review

Although peer review has been the subject of longstanding philosophical and normative discussions (Bornmann, 2008, 2011), attempts to empirically open up and theorize the black box of evaluative decision-making in funding contexts are relatively recent (e.g. Guetzkow et al., 2004; Lamont, 2009; Langfeldt, 2001; Reinhart, 2010; Van den Besselaar, 2018). Lamont (2009) bases her landmark contribution on an empirical study of peer review panels in five major American funding programs. She provides a detailed analysis of the conversational dynamics among members of review panels. These interactions gradually give rise to particular ways of conceptualizing quality and potential in proposed research projects. Review processes are pictured as a situated practice constrained by the need to process large numbers of applicants in a very limited amount of time. The focus on conversational interaction, however, also results in a certain analytical neglect of the role that particular supplementary materials such as CVs play in the review process, and of how referees actually go about interpreting biographical information.

Focusing on a slightly different type of evaluative practice, namely academic hiring processes, Musselin (2009) dedicates more explicit conceptual attention to how referees make use of academic CVs. She argues that members of tenure committees initially tend to look for disqualifying criteria in a CV to reduce the number of applicants. In a second step, referees compare researchers on the basis of positive indications. Musselin's analysis (2009: 127ff) here draws on the concept of 'judgment devices' as proposed by Karpik (1996, 2010), that is, mechanisms to facilitate purchase decisions in markets of incommensurable and only partly price-dependent goods (movies, art, medical services, luxury goods, etc.). Examples of judgment devices include reviews by professional critics, rankings or established brand names, all of which can act as mechanisms for customers to delegate the assessment of the quality of a good to other actors who are considered more competent. Musselin argues that the decision-making processes of tenure committees are generally analogous to the techniques through which buyers choose between goods in the above-described markets. Referees regularly draw on CVs, combining various judgment devices to select individual applicants out of a pool. This includes aspects such as the number of articles a candidate has published, the reputation of the publication venue as determined by citation indices, and the perceived prestige of the institution that awarded a doctoral degree (Hammarfelt and Rushforth, 2017; see also Sonnert, 1995). In other words, the central categories that constitute the academic CV here are primarily pictured as devices through which singular phenomena – the unique biographies of researchers – are transformed into at least partially comparable entities.

While this type of analysis constitutes a very useful first step, there are also aspects it tends to downplay. First, by emphasizing the complexity-reducing potential of categories and indicators, the judgment device perspective limits what constructivist scholars would

call the ‘interpretive flexibility’ of the CV as an administrative technology (Pinch, 2010). Categories are pictured as essentially stable and unchanging entities that either shape evaluative practice or that referees sometimes choose to ignore (Hammarfelt and Rushforth, 2017). Second, it pictures the use of particular indicators in the review situation as primarily guided and constrained by the immediate need to select candidates. Factors such as disciplinary culture and personal experience only enter the analysis to the extent that referees in different fields will privilege different types of indicators (e.g. historians privilege monograph publications over journal articles).

However, unlike customers in markets who tend to delegate their judgment to more competent actors, a distinctive feature about peer review is, by definition, that referees are themselves considered experts in how to compare academic career trajectories. One of the most important rationales for peer review is the highly specialized character of scientific work, which suggests that only individuals sufficiently grounded in a field should be entrusted with defining evaluative criteria and interpreting the merits and potential of their colleagues (Merton, 1973 [1942]; Nowotny, 2014; Polanyi, 1962). Naturally, this principle of self-reproduction can also be seen in more ambiguous terms. The tendency of researchers to assess contributions of younger peers in terms of quality criteria derived from their own scientific experience may contribute to a structural conservatism, not least regarding how scientific work is represented for administration and review (Cole, 2000; Fuller, 2000; Kuhn, 1962; Serrano Velarde, 2018). The spread of formal research evaluation practices has given the longstanding discussion about the ambivalent role of experience-based judgment in peer review an interesting new spin. In the understanding of many academics, the expertise of human peers should provide a safeguard against mechanical reliance on quantitative comparison in evaluative settings. But there is also a growing number of empirical studies (Müller and de Rijcke, 2017; Rushforth and de Rijcke, 2015) that demonstrate precisely the thorough embedding of metrics in routine epistemic decision-making across fields. Arguably, the problem is no longer whether particular indicators are used in peer review in the first place, but rather what exact representational function referees accord to them.

While disagreeing in a number of important respects, then, these diverse perspectives also cohere in emphasizing the significance of more basic conceptual questions: How do researchers mobilize their lived experiences when engaging in review work? And how do such interpretive practices (re)shape the categories according to which career trajectories are compared in peer review? In the following section, we propose a conceptual framework that allows us to turn these questions into a tractable empirical problem.

The interpretation of CVs as a doubly comparative practice

In this article, we treat the CV as an infrastructural administrative technology that encourages referees to look at and compare scientific biographies in terms of a number of standardized biobibliographic categories. However, to appreciate the agency of referees in interpreting such information, it is necessary to acknowledge its constructed character. Publication lists, citation scores, or track records of acquired funding may appear as natural forms of representing and comparing academic career trajectories, but only

because they hide the preceding efforts necessary to warrant their temporary stability (Stengers, 2011; Verran, 2011).

The work that is necessary to construct comparative instruments has been analyzed in exemplary fashion in a recent article by Schinkel (2016). Schinkel's case study traces how scientists achieve comparability of historical and contemporary climate data through artfully interweaving preceding comparisons, thereby stabilizing similarity/difference relations that are considered useful for the purpose at hand. Particularly important here is cordoning off a space of relevant analytical variables through ensuring that certain key elements can be considered immutable across time and space. Once this 'comparity work' is accomplished, it is black-boxed in devices that provide a mobile framework through which users can look at material and make sense of it in terms of the previously stabilized 'ontological object space' (Schinkel, 2016: 377). However, mobilizing a comparative instrument is not a mechanical matter. As with any type of scientific equipment, wielding it effectively requires users to be intimately familiar with their technology, and with the properties and possible contingencies of the phenomena to which it is applied.

Gad and Jensen (2016) theorize that the reliability of particular comparative instruments is often locally established by the users, depending on the situation and the larger assemblage of practices in which they may be embedded. Another way of putting this is to say that applying a comparative instrument even in relatively routinized circumstances actually requires users to perform a *situated* comparison – namely one between the assumptions built into the technology and the conditions experienced in the situation in which it is used. This situated comparison can serve to adjust or calibrate the instrument for the purpose at hand (Deville et al., 2016; Schinkel, 2016: 15–16). It can also, however, create situations where comparative instruments are found to be fundamentally inadequate, for example, because they do not take into account unexpected properties of the encountered phenomena. When this happens, users are forced to reflect on the shortcomings of the technologically instantiated comparative framework, and they may literally or figuratively open up the instrument, in the sense of disassembling either its material or conceptual building blocks (Mayernik et al., 2013; see also Morita, 2014). Such a reflection does not necessarily mean that the comparative activity is permanently stalled. Instead, the perceived initial inadequacy of an instrument in a given situation can prompt the users to rethink their comparative practice (Krause, 2016; see also Dewey, 1939). Through their situated reflexive effort, users may end up adding to or altering the stabilized acts of comparison that make up the instrument, and in the process give rise to new objects of comparison that make more sense in the given situation.

We suggest that the use of CVs in peer review can be usefully analyzed in these conceptual terms. For one, taking the type of analysis provided by Schinkel as a model, the categories that constitute the academic CV can be considered to consist of particular kinds of black-boxed comparity work. CVs are usually divided according to a number of elements: educational and employment history, publication lists and citation-based metrics, previous successful grant applications, administrative and teaching experience, etc. The categories can actually be seen as a standardized script that tells researchers how to represent and interpret information about academic career trajectories. In Schinkel's (2016) terminology, this framework illustrates the practice of cordoning off a space of relevant comparative variables through declaring some elements as immutable across

time and space. A first fundamental assumption implicit in the CV is that the individual scientist is a basic organizational unit that exists in the same form across different research practices and fields, and that provides a meaningful unit of evaluative comparison for peer review. A further generalizing assumption is that the individual scientist can be usefully characterized by a number of more specific properties, such as successful grant applications, authorship of publications, and particular citation metrics. Having co-evolved with the gradual institutionalization of scientific work and intellectual property conventions (Biagioli, 2000; Biagioli and Galison, 2003; Csiszar, 2017), these bibliographic categories presuppose a range of comparative assumptions in their own right. This includes the idea that scientists carry responsibility and ownership for the claims they circulate in academic publications (Biagioli and Galison, 2003), as well as the idea that citations indicate the relevance of these contributions to the scientific community (Wouters, 1999). CV categories are sustained by a host of distributed activities, part of which are invisible to scientists (Paradeise and Filliatreau, 2016). Infrastructural services such as ORCID and commercial actors such as Clarivate Analytics and Elsevier constantly generate and maintain bibliographic information, thus allowing for relatively easy retrieval of preformatted citation and publication data. This distributed work in turn is crucial to sustain the impression that categories such as authorship of publications are a natural (rather than constructed and painstakingly maintained) basis for comparison (Lampland and Star, 2009; Stengers, 2011).

However, referees appear to rarely mobilize CV information in a strictly mechanical fashion, for example in the sense of simply adding up publication figures and comparing candidates on that basis. Instead, much as scientists adjust and calibrate their scientific equipment to the conditions of field sites to ensure effective use, referees tend to qualify the abstract accounts found in CVs on the basis of an experience-based knowledge of scientific practice. Such a situated comparison is in part prompted by the discrepancy between the standardized character of the CV and a constantly changing reality of scientific work across very diverse fields (Becher and Trowler, 2001; Galison and Stump, 1996; Knorr-Cetina, 1999; Whitley, 2000). CV categories, in other words, are too unspecific to make much sense on their own terms; they need to be contextualized on the fly. For example, when interpreting productivity and publication output of an applicant, referees will draw on their understanding of the collaborative and authorship conventions dominant in an epistemic culture. Similarly, when interpreting the success of a candidate in attracting funds, they will contextualize the respective parts of a CV in their knowledge of the material and human resources necessary for research in a given field, as well as their knowledge of relevant funding sources. But by arguing that referees *compare* their personal experiences with the assumptions that underpin CV categories, we also wish to convey that this is more than just a cognitive necessity for interpreting abstract accounts. The review situation should also be understood as a 'test' in Boltanski and Thévenot's (2006) sense, that is, a principally underdetermined situation where individuals must decide between competing normative possibilities. The use of the CV requires referees to reconsider repeatedly the legitimacy of certain forms of representing biographical trajectories, and perhaps to challenge evaluative categories where they are found to rub against personal experience.

In the following analysis, we discuss three different ways in which the review situation can unfold. This empirical selection is meant to present distinct and analytically interesting situations, and is not empirically exhaustive. First, we analyze an arguably widespread practice where referees draw on their first-hand understanding of scientific practice to ‘calibrate’ their qualitative or quantitative expectations towards a CV. This often goes along with a specific dynamic in which personal experiences are actively realigned with the conceptual assumptions of the CV in the course of the review. Second, we focus on the interpretation of CVs in Big Science. This allows us to explore how referees qualify biographical evidence in research fields whose highly collaborative organization starkly contrasts with the conceptual focus of the CV on the individual scientist. In the last part of the empirical section, we analyze the situation in which the comparison between CV categories and the referees’ personal experience results in an outright discrepancy, in the sense that certain categories are perceived as a distortion of good evaluative practice.

Sources and methods

The empirical material for this article was collected as part of a larger project in which we analyze the review process in a prestigious fellowship program at a German university. The program covers a broad range of disciplines across the natural sciences and engineering, and it regularly attracts many applicants from all over the world. For the present article, we draw on a subset of the material we have collected. This includes the complete application documents (CV, short research proposal including budget plan, letters of recommendation) and review reports from 14 applications. In addition, we draw on anonymized semi-structured interviews with 11 referees and 5 applicants. Lasting between 45 and 90 minutes, the interviews were recorded and transcribed in full.

The fellowship program operates with an evaluative process that is divided into three phases. In the first phase, the program officers screen applications to sort out obviously underqualified candidates. The remaining applications are sent out to a group of four to six external referees each. These referees assess applications individually, without meeting, and provide assessments through written reports. In addition to originality and feasibility of the research proposal, the evaluation form requires referees to judge the scientific merits and potential of the candidates on the basis of their CVs. Referees are asked to explicitly compare the applicants to the most accomplished researchers in their field: ‘Does the candidate have the potential to belong to the worldwide top of his/her field?’ ‘To what extent have the achievements of the candidate gone beyond the state of the art?’ Referees provide elaborate answers in open-ended text boxes. Applicants may suggest qualified referees from their field, provided that there are no conflicts of interest through collaborative or personal relations. Further referees are chosen from among previously successful applicants and on the basis of suggestions by in-house researchers. In the third and final phase, the review reports are forwarded to a separate committee of reputed academics (typically an even mix of local and international researchers) who are tasked with establishing a definitive ranking on the basis of the detailed assessments of the external referees.

Our analysis here focuses on the evaluative activity in the second stage of this review process, that is, the work of the external referees who are asked to write reports on the basis of the submitted application materials. While the official requirements regarding format are quite unspecific, the CVs we collected were remarkably homogenous in their structure. They sequentially document educational and employment history, awards and successful grant applications, teaching experience, community service, and a comprehensive publication list. The latter is subdivided into journal articles and other forms of output such as chapters in edited volumes. Almost all CVs contain a significant number of citation-based metrics, such as the applicant's h-index, raw total citation count, and the journal impact factors of individual venues (if available). We would assume that this high degree of uniformity has something to do with the selectivity and prestigious nature of the fellowship. The applicants who make it to the external review phase generally have impressive formal career trajectories, and thus constitute a subset of scientists who have learned to adhere to the CV conventions of the most prestigious institutions in Western Europe, the US, and Asia.

While organizational aspects such as specific review modalities and the format of application materials should be taken into account in interpreting our empirical analysis (cf. Lamont, 2012), review processes should not be understood as bounded events to be studied in isolation (Gläser, 2006). Instead, the very activity of evaluation for peer review is itself a distributed practice that is learned through repetition and socialization. Although our interviews typically started off with questions about the review process for this specific fellowship, our respondents usually contextualized their statements by drawing on rich experience from previous evaluation situations in various, often international settings (other fellowship programs, grant frameworks on national and European levels, competitive tenure processes). All respondents are relatively mature scientists from the assistant professor level upwards, and as such have experience in the roles of both evaluator and applicant for fellowships and grants. When speaking about the interpretation of CVs, their perspective often tended to oscillate between the two roles.

Following the basic premises of grounded theory (Charmaz, 2006), we began coding our transcripts according to an emergent and iteratively refined set of themes. It quickly turned out that the mutually constitutive relation between the lived experience of scientists and their evaluative use of CVs would become a particularly prominent topic. This also prompted us to make slight revisions to our interview guide as we went along in the data collection, thus allowing us to pose more specific questions about the intricacies of interpreting biobibliographic information. Naturally, this interest has tended to emphasize the interpretive reflexivity of referees, thus creating a certain analytical contrast to studies that were designed to highlight the pragmatic constraints of evaluative situations (Hammarfelt and Rushforth, 2017; Lamont, 2009; Musselin, 2009).

Empirical analysis

Calibrating expectations towards the CV

A rather common stance among our respondents is that the interpretation of CVs is a matter of proper contextualization. Biobibliographic information, many believe, can

provide a meaningful basis for judging scientific potential, but needs to be assessed by standards suitable for the respective field and career stage of a researcher. Such considerations were often brought up when we asked our respondents how they interpret publication lists, citation scores, and grant application track records, and what role such evidence plays in evaluative decisions. For example, a referee in the field of robotics indicated that there is significant diversity in publication rhythms and citation-based metrics across fields that he feels must be taken into account to ensure fair assessment:

Of course, I can judge people in my field but if sometimes I look at people that are doing robotics and something different, you see that it's much easier to publish in other fields or much harder. If you work in neuroscience, you're happy if you publish one journal paper in your PhD. If robotics, maybe you expect a bit more. That changes a lot. Also, if you look at the impact factor of the journals, it varies a lot. (Interview, assistant professor of robotics, France)

Many interviewees explained that the interpretation of CVs requires calibrating expectations, in the sense that referees draw on their own experience to set suitable standards of productivity and success. When interpreting CVs, these researchers operate with an experience-based understanding in which they relate organizational and epistemic features of particular forms of research to aspects such as publication rhythm and funding modalities. For example, a senior theoretical physicist indicated that research in her field ideally involves developing innovative theoretical ideas and exploring them through complex numerical simulations. As she can tell from personal experience, such work often takes significant preparation, and hence results in a relatively slow publication turnover when compared to 'high throughput' fields such as biomedicine. This, she argued, must be taken into account when interpreting output figures:

[S]o what people are looking for are, uhm, good ideas and new ... theoretical ideas, theoretical innovation. Or, ah, impressive numerical simulation, so taking ideas and turning them into a physical result through numerical simulation and that typically takes, I would say, in the order of years rather than months. So it's quite a different field to many others because the time to publication can be quite long and that is something that actually is difficult when you're judged against people in other fields where publishing is a lot faster. (Interview, professor of theoretical physics, Ireland)

But calibration should not be understood only in quantitative terms. Referees do not draw on their own experience just to derive standards of publication productivity, citation metrics or acquired grant sums, but also to define aspects such as the relative prestige of particular journals, funding bodies, and institutions. This does not mean that referees will necessarily calibrate their expectations in identical ways, even if they work on identical research problems – researchers may still disagree when it comes to identifying the most prestigious publication venues in a given field, or whether candidates should publish two or three papers per year to be considered excellent. The unifying feature of this practice of calibration lies elsewhere, namely in the enactment of a principally unproblematic relation between scientific practice and the abstract categories through which it is represented in a CV.

In fact, when drawing on their individual practical experience to calibrate their expectations towards applicants' CVs, referees also buy into the conceptual assumptions that underpin biobibliographic categories – the idea that the individual scientist is a useful level for assessment and comparison, the idea that publications are an expression of individual intellectual abilities, the notion that citations are a direct expression of scientific relevance, etc. The foundational abstraction that enables the representation of unique career biographies in terms of standardized CV categories is thereby reconciled with the referees' personal understandings of practice. Of particular importance to the continued success of this interpretive maneuver, we suggest, are situations where an applicant's achievements clearly coincide with the referees' substantive judgment of the underlying research problems. Every once in a while, there are applicants who have successfully tackled what referees deem particularly difficult questions or methodological challenges, and who also have succeeded in producing high-impact publications on that basis. The coincidence of an applicant's choice of challenging topics and subsequent publication success is then taken to confirm the viability of publication track record as a basis for evaluation. Below is a characteristic example from a review report:

[The applicant's] pedigree shows that not only is she scientifically very productive, she also has the uncanny ability to realize her (large and ambitious) circuits and make them work using unfamiliar new principles. ... The considerable number of patents and best paper awards testify of the great abilities of [the applicant]. (Review report, electrical engineering)

In other words, the practice of calibration can best be conceptualized as a rationalization process. Referees draw on their experience to set expectations, but in the process also reinterpret and update their understanding of scientific practice according to the assumptions that underpin CV categories. Having achieved such mutual alignment, information such as publications, citations, and successful grant applications can be treated as direct expressions of the intellectual capabilities of an applicant and thus as a framework for comparison. A typical feature of this way of reading CVs is moreover that they are commonly interpreted according to a temporal logic (Hammarfelt et al., forthcoming; Musselin, 2009). This means that achievements – in particular publications and citations – are seen as milestones in a career from which referees make inferences about the (gradually unfolding) intellectual potential of a researcher. Such temporal interpretation is of particular importance in the assessment of younger researchers. Referees often try to discern significant positive or negative trends on the basis of the first few years of academic employment:

[The applicant] does not belong to the 'worldwide top in her field'. This is clearly apparent from the fact that no significant publication resulted from four years as a postdoctoral fellow in the laboratory of [a reputed scientist] at [an elite biomedical laboratory in the US]. (Review report, biomedicine)

So I would argue that [the applicant] already is a top researcher in her field, but as the strong gradient of her research recognition and the relatively few years of her academic research indicate, she has the clear potential to rise even further – and I have no doubt that she will do so. (Review report, mathematics)

The principle of calibrating expectations regarding productivity and success presupposes that referees assume a reasonable degree of coincidence between their own research experience and that of the applicant. However, in practice, referees are often asked to judge CVs that are not in the particular area of research they are themselves working on, but in areas that are imagined to be ‘overlapping’, ‘related’, or ‘adjacent’. Many of the referees we interviewed stated that they try to take into account differences in the epistemic and material organization of different specialties by adjusting their expectations regarding such factors as publication rhythm, citation figures, and grant sums according to their idea of this partly understood, but also somewhat unfamiliar, research practice. In the following, a referee specialized in algebraic geometry explains his expectations towards the publication productivity of applicants across the vast area of mathematics.

R: Well, even within mathematics it’s a little bit different from field to field, but I would say, as a general guideline, if somebody wants to be active, one to two good journal papers a year. ... in statistics, mathematical statistics, people publish much more than, for example, than in pure algebra. ... [M]aybe statistics is simply easier. In applied statistics you can count I don’t know what ... and you publish a paper about it.

WK: Okay, and in pure algebra?

R: Well, you have to produce something better than 300 years of mankind before tried to do in difficult questions. Then it takes a little bit more time. (Interview, professor of mathematics, Switzerland)

This exchange suggests that referees take normatively inflected choices in the process of calibration. The proposed comparison of the relative difficulty of research problems reveals a value-laden judgment: Researchers in pure algebra tackle the really fundamental mathematical questions people have been struggling with, whereas statistics is a relatively ‘easy’ form of research by comparison. This serves to justify a relatively low bar regarding the sheer publication output referees should expect from their colleagues in algebra.

In addition, the exchange highlights a subtle performative effect of the calibration approach. The informant first makes a rough observation about the typical publication rhythm of mathematicians, based on his own longstanding experience in the area of pure algebra. This move establishes the general legitimacy of publication output as a relevant basis for assessment, as well as a standard of ‘normal’ productivity. The referee then points to mathematical statistics as an exceptional case – although he does not have any personal experience in that area, he has observed statisticians publishing at a fast pace. His tentative conclusion is that it may ‘simply [be] easier’ to produce a paper in statistics. The suggested causal explanation that this is perhaps due to the more data-driven character of the field (‘you can count I don’t know what’) is offered more as an afterthought that rationalizes the observed differences in publication rhythm. This again implies that the calibration approach is not just a one-way process of deriving expectations from personal experience of scientific work. There is also a movement in the inverse direction, in which scientific practice is (re)interpreted through the lens of biobibliographic categories. Referees applying the calibration approach are thus able to treat substantive

differences in the organization of fields as mere differences of scale, which allows them to extend the reach of their supposedly experience-based judgment to other areas of study.

Branching out

In the calibration approach presented above, the comparison of abstract career accounts and the referee's experience of how research is organized always ends up confirming the foundational conceptual assumptions that underpin the CV (i.e. the individual scientist is considered an unproblematic level of comparison, citations are seen as a direct reflection of scientific relevance, etc.). The implicit comparative script built into the CV is imported into the situated review process and directly followed to assess applicants. Another interpretive approach can be characterized by its very suspicion towards the conceptual assumptions that enable the straightforward comparison of biobibliographic information. Referees draw on their own experience to critically examine the 'purification' that is performed as the applicants' work lives are translated into CV categories, and to reconstruct some of that original richness to create an alternative conceptual basis for assessment.

A number of astrophysicists in our interview sample illustrated the practice when they touched on the difficulty of using CV information to assess the achievements and potential of researchers in Big Science. As is well known, research in some fields has become so resource-intensive that contributions can only be made through the collaborative use of large-scale instrumentation for collecting and analyzing data. Such collaboration is often underpinned by international contractual arrangements that commit participating institutions to certain kinds of investments in shared infrastructure. In return, those institutions are free to send a certain contingent of scientists to become part of the joint projects. The research results of Big Science are circulated in the shape of articles published by dozens, hundreds, or even thousands of co-authors, who are often listed alphabetically. For the case of high-energy physics, Knorr-Cetina (1999) has argued that this goes along with a uniquely collective organization of knowledge production in which reputational competition between individual scientists is much less pronounced than in organizationally smaller-scale and more fragmented fields such as molecular biology. However, more critical accounts (e.g. Birnholtz, 2006) suggest that the degree of competition in Big Science is not simply lower in absolute terms, but rather takes a more implicit form. Our own material would seem to complicate both stances.

A pervasive opinion among the astrophysicists we interviewed is that traditional authorship-related information provided on CVs is frequently not of much use for assessing applicants for individual funding programs and fellowships in their field. One established astrophysicist framed this in especially vivid terms. The publication list of a scientist, he said, may contain many highly cited papers just because that individual has been allowed 'to push a button' in telescope-based data collection or a particular collaborative experiment, given the contractual obligation that comes with the jointly funded astrophysical research infrastructure:

[S]o the very fact that you're part of the team, and you're part of the team because your organization pays some money ... so if somebody pays for you, you have the right to push a

button and ... and put your name on a paper and then if you count the citations of this paper, there is no link between the person and the citation ... indicators have no value in terms of selecting the right people. If you just select people coming from a collaboration, if you have 300 people that you don't know, you ... you can't distinguish between them. (Interview, senior researcher in astrophysics, Italy)

Far from being a useful basis for comparing scientists, the foundational abstraction that underpins publication and citation data on CVs here poses a challenge for 'good' assessment, given the mismatch between the collaborative character of much astrophysics research and the focus of bibliometric evidence on the individual. At the same time, our interview partners expressed diverging opinions on what specific kinds of information are actually lost through pervasive 'hyper-authorship' (Cronin, 2001). The astrophysicist quoted just above specifically regrets the difficulty of discerning individually excellent scientists, given the tendency of the individual to disappear behind the collective of co-authors in big collaborations. To create an alternative basis for assessment, he draws on and triangulates CV information that is not based on journal publications. A particular important evaluative criterion for him is whether applicants for funding have a track record of conference presentations. According to his experience in many large-scale scientific undertakings, researchers who introduce and defend collective work in public are often the ones who also provide important ideas and leadership in the underlying projects. Contributing to the proceedings of a big conference, he suggested, is a 'better indicator ... than a full [journal] paper', because it actually provides a useful proxy for judging the grit and individual abilities of a scientist. Aside from this, the astrophysicist tries to simply avoid using biobibliographic evidence as a basis for assessing the quality of applicants, unless he already knows them or gets the chance to interview them personally.

However, another astrophysicist framed the problem of abstraction in CVs in a subtly different way. The respondent similarly explained that journal publications are not of much use for the purpose of selecting worthwhile applicants in the context of very large collaborative formats – while it would be odd if a candidate had no such publications, the sheer fact of being a coauthor is an insufficient basis for assessing his or her potential. However, while the previous respondent deemed very long lists of authors problematic because they make it impossible to assess individual intellectual capabilities, this second researcher was primarily worried about the difficulty of judging collaborative qualities in an applicant. Big science, he explained, means not just the possibility of gratuitous publications, but also – and perhaps more importantly – that scientists are formally included in a collective regardless of whether they actually work well as part of a team:

Also, in a subtle way you try to, without being explicit, figure out the way he relates to other people, and in these big collaborations this is a relevant fact. You know, a theoretical physicist can be completely obnoxious, but he's in his own corner and it's okay. But if you work in collaboration with other people being obnoxious is not a good quality. (Interview professor of astrophysics, Brazil)

As in the previous example, this respondent tries to draw on other information to form judgment on the relevant criteria for review. However, in line with his interest in social skills and the collaborative spirit of applicants, he is interested in different types of

sources. One part of his review routine is to scan CVs for evidence that an applicant has previously been entrusted with significant administrative tasks and community responsibilities, since this can be read as a testimony to reliability and altruism. Moreover, this astrophysicist actively draws on personal networks for peer review purposes, that is, trusted colleagues who might know the candidate personally and are able to comment on his or her ability to fit into a team. A crucial aspect thus is not necessarily the individual ‘brilliance’ of a candidate, but rather his or her ability to contribute to the research collective of the particular institution at hand:

There is always this personal issue, wherever you can you should ask people who work close-by, ‘How is this person?’ This is essential. Sometimes you have people who are not great scientists, in a sense, but they are very strong in ... depending on the institution, and that is very important, too. (Interview, professor of astrophysics, Brazil)

The examples discussed in this section show that the role of CVs in peer review for Big Science is distinct from organizationally smaller-scale fields. In the previous mode of interpreting CVs, biobibliographic categories seem to exert a certain pull on the work of the referees, in the sense that the calibration of expectations reifies the underlying conceptual assumptions. The specific characteristics of Big Science, by contrast, mean that referees habitually need to problematize a central element of these assumptions, namely the idea that journal publications and citation data indicate the creativity and abilities of a scientist. Referees here cannot simply follow the comparative script built into the CV, but instead tend to ‘branch out’ to create alternative conceptual bases for comparing applicants. The analysis also shows that referees take different directions in the process, depending on how they interpret the nature of the gap created as concrete research practice in Big Science is translated into bibliographic evidence – it can either be seen as a problem for assessing the intellectual potential of individuals, or for assessing their collaborative qualities (cf. Galison, 2003).

Repairing the CV

The previous two forms of drawing on CVs’ information to assess the potential of a researcher have in common that they are perceived as relatively commonsensical. To be sure, the case of astrophysics showed a practice where the administrative notion of authorship is critically examined and worked around in diverging ways, depending on the emphasis that individual referees place on the individual versus the collective as a relevant organizational level. While this leaves the referees with an interesting form of conceptual discretion, the more basic notion that highly collaborative work makes the category of journal publications relatively meaningless for evaluation is generally taken for granted. In this section, we will discuss a more intentionally controversial variant of mobilizing biobibliographic information for review purposes. The specificity of this approach lies in its intention to ‘repair’ what referees perceive as bad practice in the use of CVs for assessment.

A recurrent concern among some of our respondents was that the informational value of CVs is becoming increasingly problematic in a principal way, insofar as *résumés* are

subject to attempts of researchers to optimize their chances in evaluative situations (Butler, 2004; Colwell et al., 2012). More specifically, a number of senior researchers observed longitudinal shifts in the length and composition of publication lists. One professor of civil and environmental engineering explained that early career paths in her field have become more differentiated in recent decades. Graduates need to make early decisions about whether they wish to opt for an academic as opposed to an industrial career, so that they can build up the necessary credentials. As a result of this increased competition, young researchers with academic ambitions tend to have many more publications than was the case when our interview partner was at the corresponding career stage. She ascribes this development at least partly to the increasing prevalence of questionable publication practices, such as splitting up results into artificially small units. Such ‘salami-slicing’ can actually make it more difficult to judge and compare the intellectual potential of applicants:

So my PhD students now are graduating with as many publications as I had when I went up for tenure. ... there tends to be a focus on quantity over quality, which is not so say that the publications are bad, but the contribution – you often have to read three or four papers before you see the real contribution [I]t seems to me ... that people are sort of dividing up their work into small slices to have more [T]he productivity and the numbers are the most important thing often... You know – I go to conferences sometimes where some of the senior faculty have 25 papers ... it’s a mark of their dynasty if you will, their students and their grants. But it’s also ridiculous. (Interview, professor of civil and environmental engineering, US)

Referees can deal with changes in publishing practices in different ways. One option – embraced by some of our interview partners – is to apply a variant of the above-described practice of calibrating expectations towards publication output figures. Three or four small contributions can, for example, be treated as an equivalent intellectual achievement of a single publication from 20 or 30 years earlier. Others, however, do not merely choose to adjust their expectations, but instead attempt to induce change in those very publishing practices themselves. The engineer quoted above is particularly explicit about this. She uses her role as referee and head of tenure committees in her institution to promote an approach of fewer publications with more substantive contributions over what she describes as artificially ‘inflated’ publication lists:

You know – by the time a CV comes to me, I can’t really comment too much on it, but I do have the opportunity as the chair to interact with new faculty as they start the process and so I can emphasize the – you know – we care more about the quality of the work and the contributions that you’re making to the field... so we rather see fewer publications in high quality journals than 25 publications – you know – in sort of marginal venues. Now, whether that can make a difference I don’t know, but I do feel that changing the field in this way is, it’s up to senior faculty ... it’s like grade inflation like you have, someone has to put the brakes or otherwise it’s just going to escalate. (Interview, professor of civil and environmental engineering, US)

More generally, this researcher tries to contribute to altering values usually placed on authorship of journal publications in peer review for both tenure and funding decisions. This includes acknowledging mentorship of graduate students as a relevant part of

academic work. Another aspect is to more highly value other forms of publications, in particular chapters in edited volumes. The latter may be particularly useful for judging academic potential of candidates precisely because they are not heavily refereed and normally do not count much in formal evaluation settings. This, the respondent reasons, makes them a venue where academics actually pursue their real intellectual interests, and without being driven by the rationales of career development:

Yeah, so the edited volumes are one of these formats where you do have the latitude to say what you really think [amusement] Those don't count very much in, as a measure of productivity, but they are the kind of thing that gets circulated more and read, I think.

The attempt to 'repair' the use of bibliographic information in evaluative decisions is thus spurred by the ability of this particular referee to observe profound systemic developments in the career system and publishing conventions in civil engineering over a long period of time. Looking back at 30 years of experience, she posits that publication practices in the 1980s were comparatively more aligned with the epistemic organization of research, in the sense of amounting to a desirable partition between individual contributions that suits human reading habits. In a neat illustration of Goodhart's law (e.g. Strathern, 1997), the ever-tighter alignment of peer review and recruitment processes with the assumptions that underpin publication and citation data in academic CVs is perceived to have diminished the value of publication lists for judging scientific potential. The comparative script implicit in the CV is thus seen as a distorting force if allowed to uncritically inform evaluative decision-making. The proposed solution is to realign peer review around alternative bases of comparison, for example more collectively oriented forms of achievement and forms of output that have remained 'uncorrupted' by researchers' strategic considerations.

Discussion

We have argued that CVs and the information they contain do not enter peer review processes as some kind of monolithic ahistorical entity. Instead, the representational function of biographical evidence is effectively renegotiated in every review situation. Under the label of 'calibration', we first described a convention according to which referees ascribe the CV an unproblematic relation to actual scientific practice. Particular categories and indicators here are used to straightforwardly compare researchers, for example in the sense of defining 'gradients' of success on the basis of publications and acquired funding. The result is similar to the empirical argument of the judgment devices literature (Hammarfelt and Rushforth, 2017; Musselin, 2009), insofar as the CV is used to reduce the complexity of the evaluative situation. Our analysis draws attention to the subtle performative mechanisms through which the perceived fit between scientific practice and biobibliographic categories is achieved in the first place. When referees mobilize their personal experience of scientific work to set expectations towards a CV, they also update their understandings of that work. Lived experience is reinterpreted through the lens of publication lists and citation data, thereby reaffirming the legitimacy of the CV as a technology for representing academic careers.

However, there is also a diametrically opposed possibility – what we call the ‘repair’ approach. Here, referees simply do not manage to reconcile their lived experience of practice with central biobibliographic assumptions, for example that authorship of journal articles indicates originality and that the number of publications is a useful proxy for productivity. The reason is not that these referees have somehow more time for review work or funding to distribute, and can therefore afford to be more considerate in assessing biographical evidence. The case rather drives home a point made by Stengers (2011), who draws attention to the fact that comparisons must also be perceived as viable in a normative sense. Stengers argues that the legitimacy of a comparison, especially in the context of scientific work, often rests on the impression that it comes easily – the juxtaposed entities should lend themselves to the comparison. If this condition is not met, the comparison risks being rejected as problematic and contrived. In the empirical case we discuss, a key source of irritation is the longitudinal dimension of the comparison, which makes the referee trip over questionable publication practices that were not common during the early stages of her own career. Importantly, the referee also tries to use her institutional authority to actively induce change in how young researchers go about building up their résumé. The conduct of peer review here has thus prompted an intentional attempt to reshape CV conventions, to hinder the creeping reorganization of scientific practice that comes with the widespread ‘salami-slicing’ of publications.

One might object that our analysis of the ‘repair’ approach gives disproportionate room to what is perhaps no more than a minority of senior scientists who are particularly concerned about the future of their fields. After all, there is no shortage of testimonies according to which the peer review system is increasingly dysfunctional, given the difficulty to find committed referees and a tidal wave of uncritically used indicators (Burrows, 2012; Wilsdon et al., 2015). But lest this finding be discounted as an outlier or an artifact of our own research interest, we propose a parallel between the ‘repair’ approach to CVs and recent initiatives like the Declaration on Research Assessment (2018) and Science in Transition (2015). These have channeled an apparently widely felt concern about ‘unintended effects’ of evaluation and systemic problems surrounding the inflation of academic credentials, with the aim of combating problematic assessment practices on various organizational levels of the scientific enterprise. It is also worth noting that these interventions are perfectly compatible with a concern for efficiency in peer review (Lamont, 2009). As our own findings suggest, reification of CV categories can result in a surfeit of overly standardized or incremental research products which diminishes their informational value for evaluation. The long-term effect is that peer review becomes more laborious, because referees will find it more difficult to make meaningful comparisons between researchers.

Here it is instructive to consider the evaluative use of CVs in astrophysics. Due to the organizational characteristics of Big Science, researchers in this field have acknowledged that authorship-related categories often carry little informational value for comparing the scientific potential of peers. Astrophysicists resolve the mismatch between highly collaborative work and the administrative fiction of purely individual achievements through situated interpretive choices. Metaphorically speaking, the black box of authorship is habitually left open. Referees regularly branch out from the comparative script of the CV and draw on their personal experience to create diverse alternative bases for assessment.

Perhaps this should not be interpreted as a failure of the astrophysics community to close the gap between a collaborative research culture and the focus of the CV on the individual. Instead, it could be seen as the basis for a productive coexistence of evaluative registers, one register that focuses on the intellectual capabilities of particular scientists, and another that emphasizes more the collective as the relevant site of innovation (Galison, 2003). This could point to an important precondition for effective peer review also in other fields. What if we thought of peer review not so much as a way of ensuring a ‘fit’ between evaluated phenomena and scientific quality criteria, but rather as an inherently generative inquiry (cf. Fochler and de Rijcke, 2017) that must regularly problematize and reshape evaluative categories to maintain its ability to select original contributions?

In any case, our findings emphasize the need to take into account a more detailed understanding of the CV in future studies of peer review. While the assessment of CVs is typically only one part of review processes, our empirical analysis shows that the use of biobibliographical information can significantly influence review decisions on a fundamental conceptual level. It does so by creating a need for referees to choose between different ways of drawing on the representational assumptions built into a CV. Essentially, they can (1) choose to import the comparative script of the CV into the situated review process and compare individual scientists according to standardized categories, (2) use the script in a selective fashion and branch out where categories appear problematic or not useful, or (3) decide that the CV script is flawed and constitutes a potential source of distortion in peer review. This fundamental choice precedes the possibility of comparing applicants on the basis of particular indicators (cf. Schinkel, 2016), and it arguably has important implications for how referees approach other parts of the review process, such as interviews and the assessment of research proposals.

Acknowledgments

We wish to thank all of our interview partners for their time and their support of this study. Moreover, we are grateful to Isabel Burner-Fritsch for assisting us in gathering data and coding the transcripts. The empirical material for the article was collected while both of the authors were affiliated with the Munich Center for Technology in Society, as a postdoctoral researcher and as a TUM-IAS Anna Boyksen fellow, respectively. We especially acknowledge the support of Ruth Müller during this time. Moreover, we thank a number of individuals for their comments on earlier versions of the article. This includes, first, our colleagues from the Center for Science and Technology Studies at Leiden University, Guus Dix, Jochem Zuiderwijk, Paul Wouters, Anne Beaulieu, and Clifford Tatum. We have also benefited from feedback we received from the attendants of a workshop in September 2018, organized by Thomas Hellström and Merle Jacob in the framework of the KNOWSCIENCE project. Finally, we thank Sergio Sismondo and the anonymous referees for their very helpful feedback.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Wolfgang Kaltenbrunner  <https://orcid.org/0000-0003-4311-8212>

References

- Becher T and Trowler PR (2001) *Academic Tribes and Territories: Intellectual Inquiry and the Culture of Disciplines*. Buckingham: Open University Press.
- Biagioli M (2000) Rights or rewards? Changing contexts and definitions of scientific authorship. *Journal of College and University Law* 27(1): 83–108.
- Biagioli M and Galison P (2003) *Scientific Authorship: Credit and Intellectual Property in Science*. New York: Routledge.
- Birnholz JP (2006) What does it mean to be an author? The intersection of credit, contribution, and collaboration in science. *Journal of the American Society for Information Science and Technology* 7(13): 1758–1770.
- Boltanski L and Thévenot L (2006) *On Justification: Economies of Worth*. Princeton: Princeton University Press.
- Bornmann L (2008) Scientific peer review: An analysis of the peer review process from the perspective of sociology of science theories. *Human Architecture* 6(2): 23–38.
- Bornmann L (2011) Scientific peer review. *Annual Review of Information Science and Technology* 45(1): 199–245.
- Burrows R (2012) Living with the h-index? Metric assemblages in the contemporary academy. *The Sociological Review* 60(2): 355–372.
- Butler L (2004) What happens when funding is linked to publication counts? In: Moed H, Glänzel W and USchmoch U (eds) *Handbook of Quantitative Science and Technology Research*. Dordrecht: Springer, 389–405.
- Cañibano C, Otamendi J and Andújar I (2009) An assessment of selection processes among candidates for public research grants: The case of the Ramón y Cajal Programme in Spain. *Research Evaluation* 18(2): 153–161.
- Charmaz K (2006) *Constructing Grounded Theory: A Practical Guide Through Qualitative Analysis*. Thousand Oaks: SAGE.
- Cole JR (2000) The role of journals in the growth of scientific knowledge. In: Cronin B and Atkins HB (eds) *The Web of Knowledge: A Festschrift in Honor of Eugene Garfield*. Medford, NJ: Information Today, 109–142.
- Colwell R, Blouw M, Butler L, et al. (2012) *Informing Research Choices: Indicators and Judgment. The Expert Panel on Science Performance and Research Funding*. Ottawa: Council of Canadian Academics.
- Cronin B (2001) Hyperauthorship: A postmodern perversion or evidence of a structural shift in scholarly communication practices? *Journal of the American Society for Information Science and Technology* 52(7): 558–569.
- Csiszar A (2017) How lives became lists and scientific papers became data: Cataloguing authorship during the nineteenth century. *British Journal for the History of Science* 50(1): 23–60.
- Declaration on Research Assessment (DORA) (2018). Home page. Available at: <https://sfedora.org/> (accessed 18 June 2019).
- Deville J, Guggenheim M and Hrdličková Z (2016) Same, same but different: Provoking relations, assembling the comparator. In: Deville J, Guggenheim M and Hrdličková Z (eds) *Practising Comparison: Logics, Relations, Collaborations*. Manchester: Mattering Press, 99–129.
- Dewey J (1939) *Theory of Valuation*. Chicago: University of Chicago Press.
- Fochler M and de Rijcke S (2017) Implicated in the indicator game? An experimental debate. *Engaging Science, Technology, and Society* 3: 21–40.
- Fuller S (2000) *Thomas Kuhn: A Philosophical History for Our Times*. Chicago: University of Chicago Press.

- Gad C and Jensen CB (2016) Lateral comparisons. In: Deville J, Guggenheim M and Hrdlickova Z (eds) *Practising Comparisons: Logics, Relations, Collaborations*. Manchester: Mattering Press, 189–220.
- Galison P and Stump D (1996) *The Disunity of Science: Boundaries, Contexts, and Power*. Stanford: Stanford University Press.
- Galison P (2003) The collective author. In: Biagioli M and Galison P (eds) *Scientific Authorship: Credit and Intellectual Property in Science*. New York: Routledge, 325–353.
- Gläser J (2006) *Wissenschaftliche Produktionsgemeinschaften. Die soziale Ordnung der Forschung*. Frankfurt am Main: Campus.
- Guetzkow J, Lamont M and Mallard G (2004) What is originality in the humanities and the social sciences? *American Sociological Review* 69(2): 190–212.
- Hammarfelt B and Rushforth AD (2017) Indicators as judgment devices: An empirical study of citizen bibliometrics in research evaluation. *Research Evaluation* 3(1): 169–180.
- Hammarfelt B, Rushforth A and de Rijcke S (forthcoming) Temporality in academic evaluation: ‘Trajectorial thinking’ in the assessment of biomedical researchers. *Valuation Studies*.
- Karpik L (1996) Dispositifs de confiance et engagements crédibles. *Sociologie du Travail* 38(4): 527–550.
- Karpik L (2010) *Valuing the Unique: The Economics of Singularities*. Princeton: Princeton University Press.
- Knorr-Cetina K (1999) *Epistemic Cultures: How the Sciences Make Knowledge*. Cambridge: Harvard University Press.
- Krause M (2016) Comparative research: Beyond linear-casual explanation. In: Deville J, Guggenheim M and Hrdličková Z (eds) *Practising Comparison: Logics, Relations, Collaborations*. Manchester: Mattering Press, 45–67.
- Krüger A and Reinhart M (2017) Theories of valuation: Building blocks for conceptualizing valuation between practice and structure. *Historical Social Research* 42(1): 263–285.
- Kuhn TS (1962) *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
- Lamont M (2009) *How Professors Think: Inside the Curious World of Academic Judgment*. Cambridge: Harvard University Press.
- Lamont M (2012) Toward a comparative sociology of valuation and evaluation. *Annual Review of Sociology* 38(1): 201–221.
- Lampland M and Star SL (eds) (2009) *Standards and Their Stories: How Quantifying, Classifying, and Formalizing Practices Shape Everyday Life*. Ithaca, NY: Cornell University Press.
- Langfeldt L (2001) The decision-making constraints and processes of grant peer review, and their effects on the review outcome. *Social Studies of Science* 31(6): 820–841.
- Mayernik MS, Wallis JC and Borgman CL (2013) Unearthing the infrastructure: Humans and sensors in field-based research. *Computer Supported Cooperative Work* 22(1): 65–101.
- Merton RK (1973 [1942]) The normative structure of science. In: Merton RK (ed) *The Sociology of Science: Theoretical and Empirical Investigations*. Chicago: University of Chicago Press.
- Morita A (2014) The ethnographic machine: Experimenting with context and comparison in Strathernian ethnography. *Science, Technology, and Human Values* 39(2): 214–235.
- Müller R and de Rijcke S (2017) Exploring the epistemic impacts of academic performance indicators in the life sciences. *Research Evaluation* 26(3): 157–168.
- Musselin C (2009) *The Markets for Academics*. New York: Routledge.
- Nowotny H (2014) Excellence attracts excellence – and what about the rest? Reflections on excellence and inclusion. Lecture presented at the EMBO–EMBL Anniversary Science and Policy Meeting, Heidelberg. Available at: http://www.helga-nowotny.eu/downloads/helga_nowotny_b160.pdf (accessed 18 June 2019).

- Paradeise C and Filliatreau G (2016) The emergence of a metrics action field: From rankings to alt-metrics. In: Popp Berman E and Paradeise C (eds) *The University Under Pressure: Research in the Sociology of Organizations* (46). Bingley: Emerald, 87–128.
- Pinch T (2010) On making infrastructure visible: Putting the non-humans to rights. *Cambridge Journal of Economics* 34(1): 77–89.
- Polanyi M (1962) The republic of science: Its political and economic theory. *Minerva* 1 (Autumn): 54–73.
- Reinhart M (2010) Peer review practices: A content analysis of external reviews in science funding. *Research Evaluation* 19(5): 317–331.
- Rushforth A and de Rijcke S (2015) Accounting for impact? The journal impact factor and the making of biomedical research in the Netherlands. *Minerva* 53(2): 117–139.
- Schinkel W (2016) Making climates comparable: Comparison in paleoclimatology. *Social Studies of Science* 46(3): 374–395.
- Science in Transition (2015) About science in transition. Available at: <http://scienceintransition.nl/english> (accessed 18 June 2019).
- Serrano Velarde K (2018) The way we ask for money... The emergence and institutionalization of grant writing practices in academia. *Minerva* 56(1): 85–107.
- Sonnert G (1995) What makes a good scientist? Determinants of peer evaluation among biologists. *Social Studies of Science* 25(1): 35–55.
- Stengers I (2011) Comparison as a matter of concern. *Common Knowledge* 17(1): 48–63.
- Strathern M (1997) ‘Improving ratings’: Audit in the British university system. *European Review* 5(3): 305–321.
- Van den Besselaar P, Sandström U and Schiffbaenker H (2018) Studying grant decision-making: A linguistic analysis of review reports. *Scientometrics* 117(1): 313–329.
- Verran H (2011) Comparison as participant. *Common Knowledge* 17(1): 64–70.
- Whitley R (2000) *The Intellectual and Social Organization of the Sciences*. Oxford: Oxford University Press.
- Wilsdon J, Allen L, Belfiore E, et al. (2015) *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. London: HEFCE.
- Wouters P (1999) *The Citation Culture*. PhD thesis, University of Amsterdam.

Author biographies

Wolfgang Kaltenbrunner is a postdoctoral researcher at the Center for Science and Technology Studies at Leiden University. He has studied the epistemic and practical effects of diverse new forms of governing academic knowledge production, including the proliferation of research evaluation practices, novel modes of funding scientific work, as well as the development of digital research infrastructures in the humanities.

Sarah de Rijcke is Professor of Science, Technology and Innovation Studies and Scientific Director at the Center for Science and Technology Studies, Leiden University. She specializes in social studies of research evaluation, which she considers in relation to epistemic cultures, knowledge infrastructures, valuation processes, and roles of research in and for society.