

<https://helda.helsinki.fi>

A theoretical framework for explaining the paradox of university rankings

Kaidesoja, Tuukka

2022-03-01

Kaidesoja, T 2022, ' A theoretical framework for explaining the paradox of university rankings ', Social Science Information, vol. 61, no. 1, 05390184221079470, pp. 128-153. <https://doi.org/10.1177/>

<http://hdl.handle.net/10138/342376>

<https://doi.org/10.1177/05390184221079470>

cc_by

publishedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

A theoretical framework for explaining the paradox of university rankings

Social Science Information

2022, Vol. 61(1) 128–153

© The Author(s) 2022



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/05390184221079470

journals.sagepub.com/home/ssi

Tuukka Kaidesoja 
University of Helsinki, Finland

Abstract

University rankings have led to the following paradox. On one hand, global and national university rankings have an increasing impact on scientific research and higher education. On the other hand, a growing number of researchers have argued that university rankings are biased and methodologically flawed as well as documented their unintended consequences that are counterproductive to education and research activities in universities. In this article, I combine sociological and cognitive perspectives to develop a theoretical framework for explaining this paradox. The theoretical framework has four interrelated parts. The first is a distinction between three temporal stages through which university rankings commensurate universities. The second consists of an account of the social mechanisms through which university rankings generate reactive outcomes that tend to transform universities instead of just measuring their quality. The third is a league table metaphor that links the conceptual domain of team sports and the conceptual domain of universities and, I argue, provides a cognitive mechanism that shapes how many extra-academic actors, such as prospective students and policymakers, understand the results of university rankings. The fourth focuses on the affordances of the published league tables of university rankings that many extra-academic actors use for outsourcing part of their decision-making to the league tables. As a whole, this framework allows us to understand how the interrelated and materially mediated actions of different groups of actors give rise to and sustain the paradox of university rankings.

Keywords

affordance, league table, mechanism, metaphor, reactivity, university rankings

Corresponding author:

Tuukka Kaidesoja, Practical Philosophy, University of Helsinki, Unioninkatu 40, Uusimaa, 00014 Helsinki, Finland.

Email: tuukka.kaidesoja@helsinki.fi

Résumé

Les classements universitaires ont mené au paradoxe suivant. D'un côté, les classements internationaux et nationaux ont un effet de plus en plus fort sur la recherche scientifique et l'enseignement supérieur. D'un autre côté, un nombre croissant de chercheurs ont avancé que les classements universitaires sont biaisés et comportent des faiblesses méthodologiques, et ils ont documenté leurs conséquences imprévues, contre-productives pour l'éducation et la recherche dans les universités. Dans cet article, j'associe les perspectives sociologiques et cognitives pour développer un cadre théorique afin d'expliquer ce paradoxe. Le cadre théorique comporte quatre parties corrélées. La première est une distinction entre trois stades temporels à travers lesquels les classements mettent les universités en comparaison. La seconde consiste en un compte-rendu des mécanismes sociaux à travers lesquels les classements universitaires génèrent des résultats qui tendent à transformer les universités au lieu de simplement mesurer leur qualité. La troisième est une métaphore du tableau de classement qui lie le domaine conceptuel des sports d'équipe et le domaine conceptuel des universités et, selon moi, fournit un mécanisme cognitif qui modèle la façon dont nombre d'acteurs extra-universitaires, tels que les futurs étudiants et les décideurs, comprennent les classements universitaires. La quatrième se concentre sur les affordances des classements universitaires publiés, que de nombreux acteurs extra-académiques utilisent pour externaliser une partie de leur prise de décision vers les classements universitaires. Dans son ensemble, ce cadre nous permet de comprendre comment les actions interdépendantes et matériellement médiatisées de différents groupes d'acteurs engendrent et entretiennent le paradoxe des classements universitaires.

Mots-clés

affordance, classements universitaires, mécanismes, métaphores, réactivité, tableau de classement

Introduction

University rankings are comparative tools for public evaluation of universities that are based on ordinal metrics. They arrange universities into hierarchical order and create competitive relations between them. The impact of university rankings on research and teaching activities in universities has grown significantly since the turn of the century (e.g. Erkkilä, 2013; Erkkilä and Piironen, 2018; Espeland and Sauder, 2007, 2009, 2016; Fowles et al., 2016; Hallonsten, 2021; Hazelkorn, 2015; Salmi and Saroyan, 2007, 2009; Sauder and Espeland, 2009). Simultaneously, a large number of scholars in higher education and other social sciences have not only identified serious biases and methodological problems in university rankings, but also documented unintended consequences of rankings that are counterproductive to university-based research and education (e.g. Bougnol and Dulá, 2014; Brink, 2018; Erkkilä, 2013; Erkkilä and Piironen, 2018; Espeland and Sauder, 2007, 2016; Fernandez-Cano et al., 2018; Gadd, 2020; Hallonsten, 2021; Hazelkorn, 2015; Kauppi, 2018; Marginson, 2014; Münch, 2014; Münch and Schäfer,

2014; Salmi and Saroyan, 2007). This situation is paradoxical in the sense that the impact of university rankings has increased at the same time as a growing number of academic researchers have converged to the conclusion that the methodological problems and the unintended negative consequences of rankings exceed their potential benefits, even though not all of them recommend abandonment of university rankings altogether. In this article, I will develop a theoretical framework for explaining this paradox by utilizing the idea of mechanistic explanation (e.g. Hedström and Ylikoski, 2010).

The theoretical framework has four complementary parts. First, I address the social process where universities are commensurated by university rankings and distinguish between the stages of construction, implementation, and outcomes of rankings. Second, I provide an account of the semi-general social mechanisms through which university rankings may generate reactive outcomes that transform universities instead of just measuring their quality, focusing on the mechanisms of self-fulfilling prophecy, Matthew effect, data manipulation based on reverse engineering, and narrative. Third, I describe the league table metaphor that links the conceptual domain of team sports and the conceptual domain of universities and suggest how it shapes extra-academic actors' understanding of the results of university rankings. Fourth, by using the concepts of cognitive artifact and affordance, I show how many extra-academic actors partially outsource their decision-making to the published league tables of university rankings. I will develop this framework by critically synthesizing earlier theoretical and empirical research on university rankings – especially Wendy Espeland and her collaborators' ideas about commensuration and reactivity (e.g. Espeland and Sauder, 2007, 2009, 2016; Espeland and Stevens, 1998; Espeland and Yung, 2009, 2019; Sauder and Espeland, 2009) – and by complementing it with cognitive perspectives on university rankings that draw on George Lakoff and Mark Johnson's (1999, 2003) conceptual theory of metaphor, and the concepts of cognitive artifact (e.g. Norman, 1991) and affordance (e.g. Gibson, 1979). Before developing these ideas and arguments, I will outline the conceptual background for the ensuing arguments by characterizing the concepts of commensuration, status, and reactivity.

Conceptual background

The concepts of commensuration, status, and reactivity are useful for analyzing the social dynamics of university rankings and specifying phenomena for explanatory purposes. This section provides brief characterizations of these concepts that are used in the following arguments.

Commensuration

Commensuration is a social process through which qualitatively different phenomena are made comparable by constructing a common metric for them and implementing it in specific contexts in a way that creates new relations between the phenomena (Espeland and Stevens, 1998: 314–318). For example, global university rankings, such as the *Shanghai Academic Ranking of World Universities* (ARWU) and the *Times Higher Education* rankings, commensurate diverse universities all over the world by making

their ‘quality’ comparable in terms of the ranking metric that its designers use to arrange universities into a hierarchical order, which in turn creates new competitive relations between them. As Espeland and Stevens (1998) contend, commensuration provides ‘a way to reduce and simplify disparate information into numbers that can easily be compared’ (p. 316). This type of quantified information is often considered more objective and rational than the qualitative descriptions of particular phenomena (see Daston, 1992; Espeland, 2002; Espeland and Yung, 2019; Porter, 1995).

The concept of commensuration is different from the concepts of classification, standardization, measurement, and quantification, although it closely relates to them all (see Espeland and Stevens, 1998, 2008; Timmermans and Epstein, 2010). Without providing a detailed analysis of the relations between these concepts, it can be said that commensuration processes require qualitatively different phenomena to be classified by using a common category that is quantified by means of building a common metric for the category. However, this does not imply that all classification, quantification, measurement, and standardization processes involve commensuration. In the next section, I will provide a more detailed characterization of the social process through which universities (or their parts) are commensurated by regularly published rankings.

Status

The concept of status is rooted in social interactions between individuals in groups (e.g. Brown, 1988: 73–77). Its basic meaning refers to a rank (or reputation) that an individual is perceived to hold in a group in relation to its other members. This means that the other members’ perceptions and judgments constitute the status of a particular group member. Hence, the concept of status does not mean the same as the concept of social role, which refers to behavioral expectations and norms pertaining to a social position of an individual in a group. Although the class position of an individual in a society may influence the status that she enjoys or suffers in a particular group, the referents of these concepts are also analytically distinct. This is because the status hierarchy of a particular group is more local, intersubjective, and dynamical phenomenon than the class positions of individuals that reside on large-scale social structures (e.g. the prevailing relations of production or occupational structures). In addition, cultural stereotypes about gender, race, age, and other social categories as well as the personal skills and knowledge of individuals tend to affect status judgments of interacting individuals but they are also analytically distinct from the concept of status.

According to Joe Podolny and Freda Lynn (2009), the following four patterns comprise the status dynamics that is common to many human groups:

First, actors look to others’ status as a signal of their underlying quality. [. . .] Second, an actor’s status influences the rewards that she receives. [. . .] Third, an individual’s status position is not fixed, but arises from the exchange relations between individuals. [. . .] [Fourth,] actors are especially likely to rely on status queues to make inferences about quality when there is considerable uncertainty about that underlying quality. (p. 545)

What is important for the purposes of this article is to understand that, while group members tend to evaluate other members’ actions by using the status of an actor as a

signal of quality, these two do not necessarily go hand in hand. This is because status hierarchies in groups may be influenced by the factors (e.g. gender and racial stereotypes) that do not have anything to do with the members' competences to contribute to group activities. In what follows, I will provide an account of those social and cognitive mechanisms through which university rankings transfer the idea of a status hierarchy to the relations between universities.

Reactivity

The concept of reactivity can be used to explain how university rankings affect universities over time. Reactivity refers to 'the idea that people change their behavior in reaction to being evaluated, observed, or measured' (Espeland and Sauder, 2007: 1). The concept of reactivity includes an assumption that people are reflexive cognitive actors with the ability to shape their actions and beliefs in accordance with external expectations and incentives (but it does not include a stronger assumption that people are only motivated by external rewards and sanctions). Thus, the mere awareness of being an object of evaluation, observation, or measurement can alter the way in which a person thinks and behaves. In the methodology of psychology, reactivity is considered as a problem that should be controlled for since it is believed to compromise the research results (e.g. Campbell, 1957; French and Sutton, 2010). By contrast, in policymaking and management, reactivity is often associated with the attitude that Jeremy Z. Muller (2018) calls 'metric fixation' (p.15). It refers to an uncritical belief that the uses of quantitative performance and accountability metrics whose results are made public provide the most efficient ways to transform social activities and organizations in the desired direction, especially when combined with rewards and sanctions. It is interesting to contrast this attitude with 'Campbell's law' that is named after its formulator Donald T. Campbell. It says that '[t]he more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor' (Campbell, 1979: 85). University rankings lie at the intersection of the context of methodology and the context of policymaking and management since they often rely on apparently 'scientific methodologies', while also generating reactive outcomes that may (or may not) serve policymakers' and managers' aims (e.g. Espeland and Sauder, 2007).

Three stages of commensuration by university rankings

Let us now move on to the question of how university rankings are used to commensurate universities (or their parts). I will answer this question by distinguishing between three temporal stages of this commensuration process. The first stage is the construction of a new ranking metric. The second is the implementation of the ranking metric in specific contexts. The third consists of the publication of rankings results, which may trigger reactive mechanisms if the rankings are regularly implemented. Figure 1 represents these stages and their relations. In what follows, I will describe each stage in some detail by using Espeland and Sauder's (2007, 2016) case study of *the U.S. News & World Report* (shortly: USN) rankings of law schools as my primary illustrative example.¹ The USN

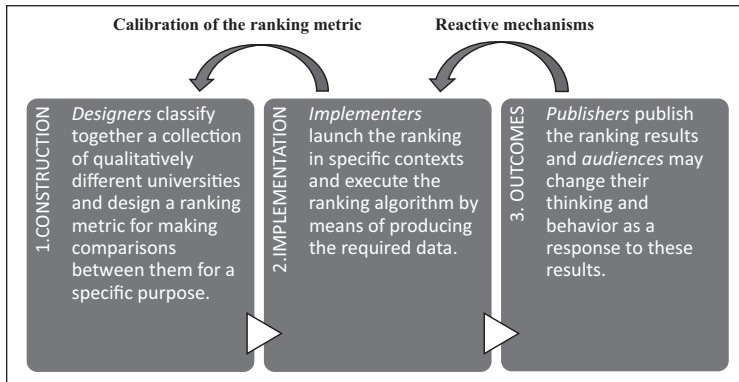


Figure 1. Three stages of commensuration by university rankings.

rankings are important because they were one of the first systematic and regularly published efforts to rank universities in the form of league tables that provided a model for many other university rankings (Salmi and Saroyan, 2007: 81). Although my focus is on the USN rankings, I suggest that the general descriptions of the three stages also apply to the other cases of university rankings.

Construction

In the construction stage, designers of a new ranking metric classify those phenomena they want to rank under a unifying category and create an ordinal metric for the category that selects, simplifies, and organizes information (e.g. Espeland and Sauder, 2007: 16, 2016: 10–17, 28–30; also Hazelkorn, 2015: chapter 2). There are three reasons why it would be problematic to describe this construction process in terms of a selection of a valid metric for an objectively existing ‘quality of universities’. First, the ideas regarding the quality of universities in different academic and extra-academic groups are context-dependent and include conflicting value judgments about higher education and scientific research, meaning that there is no consensus what quality means in this context (e.g. Brink, 2018; Espeland and Sauder, 2016; Hallonsten, 2021: chapter 2; Hazelkorn, 2015: xiv; Münch, 2014: chapter 3; Salmi and Saroyan, 2007: 85–86).² Second, a universal metric for the quality of universities by definition excludes all context-dependent measures, such as universities’ unique missions and ties to local communities, and values that are impossible to measure by means of simple quantitative indicators, such as truth, wisdom, and freedom. These two points do not deny the existence of scientific values that are used for epistemic evaluation in scientific research and that are constitutive of modern science.³ Third, the assumption that the ranking metrics for the quality of universities can be objective ignores the ways in which the published ranking results may transform their audiences’ perceptions of the relative quality of universities (Espeland and Sauder, 2007; Muller, 2018: chapter 7). Hence, it can be said that a new ranking metric creates a single norm for quality that not only tends to replace the context-dependent and conflicting conceptions that different academic and extra-academic groups have

about the relative quality of universities (cf. Sauder and Espeland, 2009: 73). It may also influence the standards of scientific evaluation via the reactive mechanisms that I will discuss in the next section.

Espeland and Sauder's case study of the USN rankings of law schools illustrates these points. In the 1990s, the editors of USN magazine set the goal of building a new ranking of law schools that would provide prospective law students and their parents comparable, unbiased, and consumer-friendly information about the relative quality of law schools in the United States (Espeland and Sauder, 2016: 10–12, 49). The editors hoped, too, that the publication of the ranking results would generate additional revenue for their magazine. In this goal setting, USN framed prospective law students as consumers who invest a large amount of money for the educational services offered by the law schools and who, therefore, need objective information about the relative quality of these services for their choices as rational consumers. Later on, USN additionally motivated their law school rankings by claiming that the ranking results provide useful information to law firms employing graduated lawyers, and that broader audiences may also use the ranking results as the means for holding law schools accountable for their customers and funders. Similar rhetoric emphasizing the instrumental value of higher education and science (at the expense of their intrinsic value) and the accountability of universities for their funders and customers has been used to market global university rankings (e.g. Hazelkorn, 2015: 133–135) in a way that resonates with Michael Power's (1997) notion of 'audit society'.

The basic structure of the metric used in the USN law school rankings has remained the same since its introduction in the 1990s (Espeland and Sauder, 2016: 15, 218). In 2014, the magazine classified quantitative indicators used in calculating rankings into the following four categories (weighting of each category is mentioned within parentheses): reputation (40%), selectivity (25%), placement success (20%), and faculty resources (15%). It is noteworthy that these four categories and their indicators are built upon a narrow and somewhat arbitrary conception of 'the relative quality of law schools'. For example, no indicators for the quality of teaching, student well-being, educational equity, or collaboration between law schools are included in the ranking metric. Nor does it take into account differences in educational missions of different law schools or differences in their ties to local communities. In addition, the concept of quality of law schools is assumed to be one-dimensional: it is measured by a single score that is calculated by adding the weighted scores of the above four indicators standardized and re-scaled to a maximum of 100, while the weightings of these indicators are based on subjective decisions by the designer of the ranking metric. Hence, the composite multi-indicator used in the USN rankings of law schools is the result of many subjective, value-laden, and arbitrary choices. As has been recognized by many scholars in higher education (e.g. Brink, 2018: chapter 2; Fernández-Cano et al., 2018; Hazelkorn, 2015: chapter 2; Marginson, 2014; Münch, 2014: 23–37; Salmi and Saroyan, 2007), this feature is common to popular university rankings that rely on composite multi-indicators. However, the rankings metrics of global university rankings are mostly comprised of the indicators for research outputs, such as the number of peer-reviewed publications in high-impact journals, the citation indices, and the number of Nobel Prizes and Fields Medals (see, for example, Marginson, 2014), and include none or few indicators for other type of publications and the quality of higher education.⁴

Implementation

Once designers have built a new ranking metric, implementers (who may or may not be the same people or organized groups as designers) execute the ranking in specific contexts by means of producing the required data for the ranking algorithm. The first attempt to launch the new ranking may lead to calibration of the ranking metric, especially in cases where the relevant data are difficult or impossible to access. In addition, the new ranking has to be legitimated to its implementers (if they are not the same people as its designers) and its potential audiences. Designers typically legitimize their rankings to implementers and audiences by claiming that they provide information to prospective students for making consumer choices as well as information to taxpayers and other stakeholders for holding universities accountable. It seems to me that these legitimations often appeal to a common tendency of people in modern capitalist societies to pay attention to status hierarchies since rankings can be understood as a way of constructing a status hierarchy among universities that the extra-academic audiences take to signal quality differences between universities (cf. Bowman and Bastedo, 2009: 419). However, there is a difference between the spontaneously emerging status dynamics in groups and the status dynamics constructed by the university rankings since the latter phenomenon is created by the non-spontaneous construction and implementation of the rankings. Accordingly, the implementation of university rankings typically requires a large amount of institutional work and encourages university-based implementers to use different kinds of gaming strategies in order to secure the best possible ranking results (e.g. Espeland and Sauder, 2016: 176; Muller, 2018, 77–78). These issues are seldom discussed when the costs and benefits of rankings are considered.

To continue our illustrative example, the implementation of the USN rankings of law schools requires, among other things, data about acceptance rates, accepted students' scores in the standardized Law School Admission Test (i.e. LSAT scores), employment rates after graduation, expenditure rates per student, student-faculty ratios, and volumes in the library for each school. In addition, the USN magazine conducts two reputation surveys in which academics and practitioners rate all law schools in the United States, basing their ratings on their subjective judgments. Although the basic indicator categories have remained the same, the USN magazine has slightly calibrated how it calculates some indicators since the introduction of the ranking metric (see Espeland and Sauder, 2016: 14–15).

Espeland and Sauder (2016: 15–17, 176–178; also Sauder and Espeland, 2009) interestingly describe how USN sought to delegate implementation of the new ranking mostly to the employees of the law schools and how the employees responded to this requirement. Initially, some law schools refused to produce the ranking data requested by the magazine since their employees considered the USN rankings to be methodologically flawed and counterproductive to the goals of legal education. Opponents of rankings also protested that producing ranking data for a commercial magazine steals time from more important work. However, the resistance faded away mainly for two reasons. First, USN replaced the missing ranking data with its own numbers, underestimating the ranking scores for those law schools that refused to provide the requested data. Second, prospective law students, law firms, and external funders began to pay increasing attention to the

published ranking results of law schools. As a resultant effect of these two causes, almost all law schools that initially protested against the USN rankings eventually ended up submitting the requested data to USN because their refusal to submit would have risked their reputation (or perceived status) among the audiences of the rankings and, subsequently, their future resources. This episode illustrates the disciplinary power that university rankings may have with respect to those working at universities, making the impact of rankings difficult to buffer by symbolic actions that do not affect the basic functions of universities (Espeland and Sauder 2016: 178; Sauder and Espeland, 2009; see also Münch, 2014: chapter 1). It also forms a bridge to the next stage of commensuration by rankings, which is the process through which the published ranking results generate reactive outcomes.

Outcomes

University rankings are typically published in the form of league tables that are nowadays accessible through the web and widely reported in the news media (e.g. Espeland and Sauder, 2016: 181; Hazelkorn, 2015; Salmi and Saroyan, 2007). For example, USN publishes its rankings of law schools as a part of its annual issue devoted to the rankings of graduate schools in the United States. The issue represents the ranking results in the league table format that organizes law schools (and other university colleges) into a hierarchical order according to their ranking scores (see Table 1 below). The recognition of the fact that the USN rankings of law schools are published annually is important for understanding how they have transformed the field of legal education in the country. This is because the increasing attention gained by the ranking results on the part of their extra-academic audiences, such as prospective law students, legal employers, and external funders, was the triggering condition for the social mechanisms through which the USN rankings generated reactive outcomes that were mostly unintended by their designers (Espeland and Sauder, 2009, 2016, 2009). In the next section, I will identify four reactive mechanisms and suggest that extra-academic actors' responses to ranking results tend to trigger similar mechanisms in other cases of popular university rankings.

Social mechanisms generating reactive outcomes

In order to understand how university rankings generate reactive outcomes, it is essential to realize that competition for ranking positions is a zero-sum game. This means that if the rank of one university rises, then the rank of some other university/ies must fall. This feature of university rankings together with the wide attention that repeatedly published rankings receive from their audiences are constitutive features of competition between universities for high ranking positions, which may be understood as a kind of scarce resource that is socially constructed by the rankings (cf. Brankovic et al., 2018; Münch, 2014 chapter 1). High rankings are important for universities mostly because they affect (or may even constitute) the status that a university enjoys or suffers among the extra-academic audiences of rankings, which, as will be argued below, affects its opportunities to gain new resources. From this perspective, it is not surprising that many deans and other employees of law schools interviewed by Espeland and Sauder (2016) found the

Table 1. The top 10 league table of best law schools in 2020 that is provided on the USN website (<https://www.usnews.com/best-graduate-schools/top-law-schools/law-rankings>, read 1 February 2020).

Rank 2020	Law school	Location	Tuition (full time)	Enrollment (full time)
1	Yale University	New Haven, CT	\$64,267	621
2	Stanford University	Stanford, CA	\$62,373	565
3	Harvard University	Cambridge, MA	\$64,978	1737
4	University of Chicago	Chicago, IL	\$64,089	600
5	Columbia University	New York, NY	\$69,916	1268
6	New York University	New York, NY	\$66,422	1380
7	University of Pennsylvania	Philadelphia, PA	\$65,804	755
8	University of Virginia	Charlottesville, VA	\$60,700 (in-state) \$63,700 (out-of-state)	923
9	University of Michigan – Ann arbor	Ann Arbor, MI	\$59,762 (in-state) \$62,762 (out-of-state)	1012
10	Duke University	Durham, NC	\$64,722	682
10	Northwestern University (Pritzker)	Chicago, IL	\$64,402	752
10	University of California – Berkeley	Berkeley, CA	\$49,325 (in-state) \$53,276 (out-of-state)	976

time preceding the publication of law schools rankings very distressing and spent a significant amount of their working hours on activities that aim to improve their ranking.

In what follows, I will describe four interrelated social mechanisms through which university rankings tend to generate reactive outcomes (for mechanism-based theorizing and explanations, see Hedström and Ylikoski, 2010; Ylikoski, 2019). In particular, I will provide theoretical accounts of the generalizable social mechanisms (i.e. mechanism schemas) together with some illustrative examples. The following mechanism schemas pertain to the feedback loop between the outcomes and implementation stages in Figure 1. The idea is these abstract schemas help researchers to identify concrete social mechanisms that operate in diverse cases of university rankings where their effects may be slightly different due to the contextual differences (cf. Espeland and Sauder, 2016: 28, 180–190; see also Münch, 2014: 79–85; Ylikoski, 2019).⁵ In the context of case studies, these mechanism schemas can be used to develop case-based causal narratives that explain outcomes in a particular case in terms of interacting mechanisms and causally relevant contextual factors.

Self-fulfilling prophecy

Perhaps the most important social mechanism through which university rankings generate reactive outcomes is self-fulfilling prophecy. In Espeland and Sauder's (2007: 11)

study, this term refers to processes ‘by which reactions to social measures confirm the expectations or predictions that are embedded in measures or which increase the validity of the measure by encouraging behavior that conforms to it’ (cf. Merton, 1968a: chapter 13). A subspecies of this mechanism is a process whereby expert respondents to the USN reputation survey used the previous ranking results (that were easily available to them) to rate those law schools that they did not have detailed knowledge of – note that there are over 200 law schools in the United States meaning that no one can have detailed knowledge of them all (Espeland and Sauder, 2016: 31). In this way, they ended up ‘validating’ the results of earlier rankings. This type of reputation surveys have a central role in some global university rankings, such as the *Times Higher Education* rankings and the *Quacquarelli Symonds (QS)* rankings, where they also function as self-fulfilling prophecies (e.g. Marginson, 2014). Another variant of self-fulfilling prophecy is the process through which deans and administrators of law schools allocate more resources to those functions that have potential to enhance their school’s standing in the next ranking (Espeland and Sauder, 2016: 32). There is evidence that responses of policymakers, university leaders, and administrators to other popular university rankings have given rise to similar self-fulfilling prophecies in which ranking metrics influence on the allocation of funding (e.g. Fowles et al., 2016; Hazelkorn, 2015: chapter 3). In the case of global university rankings, this has resulted in allocating more funding to the natural sciences, mathematics and medicine at the expense of the social sciences and humanities (e.g. Münch, 2014: chapter 1).

Matthew effect

Richard Münch (2014: 82–83, chapter 4) has argued that university rankings have contributed to the growing inequality between universities, departments, and scientific disciplines by drawing on Robert K. Merton’s notion of Matthew effect in science. According to Merton’s (1968b) original characterization, ‘the Matthew effect consists in the accruing of greater increments of recognition for particular scientific contributions to scientists of considerable repute and the withholding of such recognition from scientists who have not yet made their mark’ (p. 58). Münch’s (2014: chapter 4) idea is that university rankings (alongside with some other performance metrics that are used to evaluate universities) produce Matthew effects at the level of organizations that compete for socially constructed excellence (or status). For example, the metrics of the global university rankings tend to favor private, wealthy, English-speaking, and research-intensive universities. And if the published ranking results gain wide public attention, then a high ranking of the university affects its opportunities to acquire new resources – such as bright students, renowned researchers, external research funding, patent income, and high tuition fees – more effectively than the universities with a low ranking. When the ranking results are regularly published, this tends to trigger the self-reinforcing mechanism that widens the gap between highly and lowly ranked universities, which may result in ‘over-equipment beyond the efficiently and effectively usable level at the top’ and under-equipment at the middle and bottom (Münch, 2014: 251; see also Hazelkorn, 2015: 16–18; Kauppi, 2018; Marginson, 2014). Although this outcome benefits elite universities, it does not benefit higher education and scientific research as a whole. These self-reinforcing

mechanisms not only deepen inequality between universities but may also lead to the non-optimal allocation of resources unless they are constrained by counteracting social mechanisms or institutional restrictions. According to Münch (2014: 82–85), Matthew effects between highly and lowly ranked universities are generated and sustained by the previously described self-fulfilling prophecies combined with the exclusiveness and social closure of elite universities (for evidence of the stability of the ranking positions of universities over time, see Fowles et al., 2016).

Data manipulation based on reverse engineering

By reverse engineering, Espeland and Sauder (2016: 33) refer to ‘the process of working backward through the construction of a complicated object or artifact to gain knowledge about how it works’. In the case of university rankings, this complicated object is the composite multi-indicator used as a ranking metric. University administrators try to deconstruct the ranking metric into its component indicators with the aim of finding the most efficient ways to improve their ranking scores in the next round. Since the reverse engineering mechanism in Espeland and Sauder’s work also includes university administrators’ manipulation of data submitted to USN, I suggest renaming this mechanism as ‘data manipulation based on reverse engineering’. For example, Espeland and Sauder (2016: chapter 6) describe how career services personnel identified the importance of placement success for their ranking scores through reverse engineering and ended up manipulating the placement statistics of their law schools that they reported to the USN magazine. Their actions included both direct manipulations of the submitted data – including plain cheating – and short-term recruitments of newly graduated lawyers just before the measurement of employment rate of graduates. Since the employees of law schools are mostly responsible for the implementation of the USN ranking, they have plenty of opportunities for reverse engineering and data manipulation. There is also evidence that university administrators use similar reverse engineering techniques and data manipulations in their responses to other popular university rankings, such as setting up targeted advertising campaigns to boost the ratings in reputation surveys, which are important in some university rankings (e.g. Muller, 2018: 75–81). In addition, Mario Biagioli and Alexandra Lippmann’s (2020) edited volume maps and documents examples of gaming and data manipulation pertaining to academic publication processes and discussed their relation to global university rankings. They include excessive self-citations, citation rings, peer-reviewing rings, extending the number of co-authors, universities’ incentives for researchers to publish in high-impact factor journals, and the uses of article-generating software (e.g. SCIfgen).

Narrative

In the context of university rankings, narrative can be understood as a culturally embedded cognitive mechanism that university employees use to make sense of the ranking results by means of re-contextualizing them (Espeland and Sauder, 2016: 36–37). For example, they may create and tell narratives to each other in order to identify the sequences of events through which the ranking results were generated in their university

and, thereby, to come up with suggestions for actions to improve their standing. In the case of the USN rankings of law schools, this kind of re-contextualization is badly needed because the ranking metric excludes nearly all contextual information about law schools. Espeland and Sauder (2016: 36–38) describe, for example, how the deans of law schools told either celebratory or defensive narratives after the publication of ranking results, depending on whether the position of their law school had improved or deteriorated. These narratives identified heroes or villains deemed responsible for the ranking results and, in the case of poor ranking, they typically included blaming the USN ranking methods. Because the published results of other university rankings also provide highly decontextualized information (e.g. Brankovic et al., 2018), it is likely that similar narratives are used to re-contextualize their results.

In conclusion, the second part of my theoretical framework for explaining the paradox of university rankings consists of the accounts of the reactive mechanisms of self-fulfilling prophecy, Mathew effect, data manipulation based on reverse engineering, and narrative through which university rankings may transform universities. It is the task of empirical case studies and comparative studies to find out how the instances of these generic mechanisms play out in particular cases and how different contexts shape their outcomes. For example, it is likely that the impact of particular global university rankings on universities is more limited than the impact of the USN law school rankings on law schools since there are many competing global rankings that gain wide publicity and none of them ranks all universities in the world (e.g. Hazelkorn, 2015: chapter 2).

The league table metaphor

I suggested above that the reactive mechanisms are triggered when the extra-academic audiences of university rankings begin to take the published ranking results seriously. However, I did not answer in detail why extra-academic audiences take them seriously and how they use the published ranking results in their decision-making. In the next three sections, I will focus on these issues. I will introduce a couple of conceptual tools from cognitive linguistics and anthropology to answer the above questions. The first is the cognitive mechanism of embodied metaphor and, more specifically, the league table metaphor that I will develop in this section. The others are the concepts of cognitive artifact, affordance, and outsourcing that I will discuss in the next section. To illustrate these concepts, I will again use the USN ranking of law schools as my primary example.

The league table metaphor draws on George Lakoff and Mark Johnson's (1999, 2003) conceptual theory of metaphor, according to which conceptual metaphors shape our cognitive processes and actions in profound ways. Roughly speaking, we use metaphors to understand and experience 'one kind of thing in terms of another' (Lakoff and Johnson, 2003: 5, italics removed). Lakoff and Johnson (1999, 2003) emphasize that metaphors ultimately have their basis in our bodily experiences in our natural and cultural environments, such that we understand abstract concepts by relying on the concepts that are closer to our embodied experiences. This allows us to use our embodied experiences to create and make sense of abstract concepts by means of analogical reasoning. For example, we often understand time in terms of the moving objects that we can perceive in our

everyday environments and, accordingly, use the terms ‘in front of’, ‘behind’, ‘follow’, ‘precede’, ‘close’, ‘far away’, ‘fast’, and ‘slow’ to make assumptions and inferences about time (Lakoff and Johnson, 1999: chapter 10). In addition, Lakoff and Johnson (2003) emphasize that metaphors have a dual nature: while a particular metaphor highlights some aspects of an abstract concept to us, it hides some of its other aspects (pp. 10–13). This means that different metaphors may provide us with different understandings of the same abstract concept. For example, an alternative way of understanding time is to rely on the ‘time is money’ metaphor that allows us to think of time in terms of ‘spending’, ‘saving’, ‘investing’, and ‘wasting’ (Lakoff and Johnson, 1999: 163–164). It is obvious that the ‘time is money’ metaphor provides a quite different understanding of time than the ‘time is moving object’ metaphor and presupposes a cultural context where money is used in economic transactions.

Building upon these concepts and ideas that are supported by experimental and other evidence (for a review, see Lakoff and Johnson, 2003: 243–276), I suggest that the league table metaphor employs the conceptual domain of team sports to make sense of the results of university rankings and, thereby, transfers the idea of status competition to the relations between universities.⁶ Although it juxtaposes sport teams and universities, the league table metaphor does not include an assumption that universities are in all respects identical with sport teams nor the assumptions that competition between universities is in all respects identical with competition between sport teams. Instead, the league table metaphor suggests that the results of university rankings are *partially* structured, understood and talked about in terms of the league tables of team sports. Accordingly, the league table metaphor does not deny that an actor’s reasons for accepting the idea that university rankings measure the relative quality of universities may be different from her reasons for accepting that the league tables of sport teams represent the relative quality of sport teams. Still, my claim is that the league table metaphor provides a plausible mechanistic explanation that shows how many extra-academic actors consider university rankings as unproblematic representations of the relative quality of universities.

More specifically, I propose that those who use the league table metaphor to make sense of university rankings are prone to make the following assumptions and inferences about the ranked universities that are (mostly unconsciously) guided by the ways in which they have used the league tables of sport teams to make assumptions and inferences about sport teams:

- Universities compete with each other in annual rankings whose results are regularly published as league tables.
- There is an unambiguous and objective metric for the relative quality of universities.
- The position of a particular university in the league table displays the status of the university among the ranked universities and, thereby, provides a reliable signal of its relative quality since the position is based on the unambiguous and objective ranking metric.

- Excellent quality is a scarce resource among universities, meaning that there can be a very limited number of top universities.
- When the standing of a university rises in the subsequent league tables, its quality improves.
- When the standing of a university falls in the subsequent league tables, its quality declines.

By relying on Lakoff and Johnson's (1999, 2003) ideas, it could be further argued that all league tables rely on more basic metaphors and embodied image schemas, such as 'groups/organizations are individuals', 'competition is zero-sum game', 'more is up', 'better is up', 'more is better', and 'quality is quantity', but I will not pursue this argument here. As such, this argument would be compatible with the view that the league table metaphor, built upon these basic metaphors and image schemas, is transferred to the conceptual domain of universities via the conceptual domain of team sports where the basic metaphors are institutionally well established and uncontroversial.

My suggestion is that the league table metaphor allows us to analyze the ways in which especially extra-academic audiences are prone to understand the results of university rankings although they are not necessarily aware how this metaphor guides their thinking. The following considerations support this hypothesis. First, the results of university rankings are published in the league table format that closely resembles the league tables of sport teams (I will analyze the USN league table of law schools in detail in the next section). Second, there are terminological similarities in the ways in which sport leagues and university rankings are discussed, exemplified by the terms 'league table', 'competition', 'score', 'standing', 'winner', 'loser', and 'top team/university'. Third, it is plausible to assume that the source domain of the league table metaphor is familiar to the extra-academic audiences of university rankings since the league tables of sport teams are widely published in the news media and their uses are uncontroversial in the domain of team sports (e.g. Ringel and Werron, 2020: 153–156).⁷ Fourth, the league table metaphor that emphasizes status competition between universities resonates with New Public Management and academic capitalism that have been the driving forces of science and higher education policy all over the world in recent decades and affected how universities are discussed in the media (e.g. Cantwell and Kauppinen, 2014; Münch, 2014).

One might object to the above hypothesis by the following claim: instead of relying on the league table metaphor, extra-academic actors tend to understand university rankings in terms of market competition. To support this view, it could be argued that extra-academic actors are surely aware that universities compete for students in higher education markets and that the main purpose of rankings is to provide information to prospective students for making consumer choices between universities. However, even if one conceptualizes competition between universities for students in terms of market competition and prospective students as consumers, this understanding alone does not seem to provide conceptual resources for making sense of the results of university rankings. The main reason is that market competition is mediated by the price mechanism that is entirely absent from university rankings. According to the ideal

model of competitive markets, price mechanism operates by balancing the supply of many producers and the demand of many consumers in a decentralized manner. By contrast, university rankings require there to be a central designer who articulates ‘demand’ for all universities by constructing the ranking metric (Münch, 2014: 2–3). In addition, unlike competition between universities for ranking positions, market competition between companies for revenues is not a zero-sum game, since profits of many companies can increase at the same time if the market as a whole grows. For these reasons, the market metaphor does not allow one to make sense of the results of university rankings – it is also debatable to what extent it can be applied to the competition between universities for students despite the fact that the market imagery is commonly used in policy rhetoric (see Marginson, 2013).

Given the above idea that metaphors both highlight and hide aspects of abstract concepts, we can ask what aspects of universities the league table metaphor hides. Since it takes for granted that rankings provide objective evaluative information about universities, it hides the arbitrariness and value-ladenness of the ranking metrics. Since it emphasizes status competition between atomistic universities, it hides collaboration between universities and their relations to their environments. Since it assumes that competition between universities is a zero-sum game, it hides the fact that many universities can be good at the same time when evaluated by qualitative criteria and that universities can be good in different ways. Since it highlights the relative movements of universities in the ranking hierarchy over time, it hides the fact that these movements may reflect meaningless statistical fluctuations in ranking scores or shifts in ranking metrics rather than any substantial changes in universities.

How do prospective students outsource part of their decision-making to the results of university rankings?

Nicholas Bowman and Michael Bastedo (2009: 418–419) have identified three paths through which college rankings may influence decision-making of prospective students:

First, rankings can affect perceptions of institutional quality directly. Second, they can create subjective norms of which colleges and universities are considered to be the ‘best’, which may or may not align with an individual’s own perception of quality. Third, they can affect students’ perceptions of the probability that they will be accepted by the college of their choice.

Although this distinction is useful, Bowman and Bastedo (2009) do not explicate in detail the above processes and they also remain undertheorized in Espeland and Sauder’s (e.g. 2007, 2016) otherwise brilliant case study. This section aims to fill this gap by developing a hypothesis that the league tables of university rankings can be understood as cognitive artifacts that provide specific affordances to prospective students that many of them use to outsource part of their decision-making to the ranking results. By the mechanism of cognitive outsourcing, I refer to a process in which an actor begins to use a cognitive artifact to perform a cognitive task that she would otherwise have to perform herself. Before developing these ideas in detail, I will characterize the concepts of cognitive artifact and affordance.

According to Donald Norman (1991), a cognitive artifact is ‘an artificial device designed to maintain, display, or operate upon information in order to serve a representational function’ (p. 17). Maps, newspapers, books, and computers are all examples of cognitive artifacts that we use in our everyday lives. The notion of affordance was introduced in James J. Gibson’s (1979) ecological theory of visual perception that addressed the question how living organisms perceive their immediate natural environments and emphasized the action-relatedness of perceptual processes. Norman (1993: 106) extended this concept to the domain of human-made artifacts and technologies by arguing that ‘[d]ifferent technologies afford different operations’, thereby making ‘some things easy to do, others difficult or impossible’. It is important to understand that the affordances of a particular technology not only depend on the intrinsic properties of the technology but also on the specific bodily and cognitive features, abilities, and skills of its users. For example, a geographical map provides affordances for navigation only for those who can read cartographic symbols and compass points. In this sense, affordances are relational.

Since cognitive artifacts serve representational functions, the notion of external representation can be used to analyze how the affordances of a cognitive artifact affect the ways in which their users process information. According to David Kirsh (2010), external representations displayed in cognitive artifacts may transform the cognitive powers of their users in at least seven ways:

They change the cost structure of the inferential landscape; they provide a structure that can serve as a shareable object of thought; they create persistent referents; they facilitate re-representation; they are often a more natural representation of structure than mental representations; they facilitate the computation of more explicit encoding of information; they enable the construction of arbitrarily complex structure; and they lower the cost of controlling thought – they help coordinate thought. (p. 441)

Hence, cognitive artifacts are not just external aids to internal cognitive processes. Instead, they tend to transform cognitive processes of their users by enabling them to outsource cognitive tasks that they would otherwise have to (try to) perform internally.

Let us now apply these ideas to the USN law school rankings by focusing on the top 10 league table of the 2020 ranking published in the USN website (see Table 1). It forms a part of the comprehensive league table that includes approximately the top 150 law schools ordered according to their ranking scores. I suggest that the USN league table can be understood as a cognitive artifact that provides the following affordances to the prospective law students who, it is plausible to assume, are all literate and numerate:

- Affords them to perceive a hierarchical and transitive order represented by the spatial relations among the names of law schools such that highly ranked law schools are at the top;
- Affords them to make unequivocal, quick, and easy comparisons between any two law schools in terms of their rank;
- Affords them to coordinate information about the rank, location, tuition, and enrollment for each school;

- Affords them to compare the rank of a university to its ranks in the previously published tables;
- Affords them to share the ranking results with others (e.g. through social media);
- Provides them with a stable object that affords joint attention and references in conversations (either in the web-mediated or in the face-to-face communication).

The first four affordances relate to the visual features of league tables, while the last two affordances pertain to the functional properties of league tables as parts of the socially distributed cognitive processes that involve more than one actor (for distributed cognition, see Hutchins, 1995). These aspects of university rankings are often overlooked in the research literature on rankings that, with some exceptions (e.g. Ringel and Werron, 2020), tends to focus only on quantification and numbers.

My claim is that the affordances of the USN league tables allow prospective students to outsource their reflections about the concept of relative quality of law schools, judgments about the relative quality of law schools, and information gathering about diverse factors pertaining to the relative quality of law schools (however defined) to the USN rankings of law schools. This type of cognitive outsourcing may be a reasonable thing to do for those prospective students who rely on the league table metaphor and, accordingly, equate high quality with high status, and who seek to maximize their monetary gains for their educational investments. This is because the self-fulfilling nature of the USN rankings guarantees that studying in a highly ranked elite law school provides one with increasing returns in the form of accelerated career development and higher salaries if the relevant groups of actors, especially the most prestigious law firms, take the ranking results as measures for the relative quality of legal education. However, it is good to remember that most applicants are not accepted to the elite law schools and that not all prospective students have only instrumental aims for their education although they may use the USN league tables in their decision-making (e.g. Espeland and Sauder, 2016: chapter 3). For these groups, that form the majority of prospective law students, the USN rankings may be more harmful than helpful. The reason is that the USN ranking contributes to the reproduction of unequal access to lower or higher ranked law schools and unequal returns based on the lower or higher investments of students and their parents (cf. Münch, 2014: 61).

The above claims, however, do not imply that all those prospective law students who rely on the league table metaphor in their interpretation of the USN ranking results would apply to Yale University to study law. The relative quality (or status) of a law school is seldom the only criterion that a prospective student uses in deciding between law schools. For example, law schools' distance to home, the financial resources of their parents, their career plans, and their own LSAT scores affect decisions of many prospective students (Bowman and Bastedo, 2009; Espeland and Sauder, 2016: chapter 3). However, my hypothesis is that at least those prospective students who have internalized the league table metaphor tend to use the USN league table as an overall frame for coordinating their decision-making. More specifically, they may use the ranking order as a way of shortlisting law schools or deciding between those law schools they have shortlisted using other criteria. They may also refer to the USN league table in order to justify their decisions to others, such as their parents and friends.

Although it is an empirical question to what extent prospective students outsource their decision-making to the USN league tables, there is some evidence that supports the above claims. Espeland and Sauder's (2016: chapter 3) interviews of prospective law students indicate that they actually use the USN league tables to make comparative judgments about the relative quality (or status) of law schools and to shortlist law schools quite uncritically. They also cite their own and other researchers' survey results showing that most applicants admitted they consulted rankings in their decision-making between law schools (Espeland and Sauder, 2016: 55). Bowman and Bastedo (2009) in turn provide evidence that the USN rankings have the strongest impact on the decisions of highly qualified prospective students (i.e. those who have high LSAT scores in the case of law schools). In addition, Ellen Hazelkorn (2015: 134, 150) reports survey results according to which students had the most favorable view on university rankings of all their users and that over 80% of them had a high interest in rankings. She also refers to the data from USN showing that 80% of the visitors in its website 'directly enter the ranking section rather than go through the magazine's home page', suggesting that these visitors are only interested in the published league tables (Hazelkorn, 2015: 137). Espeland and Sauder (2016: 52) also found that web-based chats on the pre-law blogs and discussion boards of prospective law students (e.g. *top-law-schools.com*) make frequent references to the USN league tables of law schools.

Furthermore, there are also specific motivating reasons why many prospective law students are prone to outsource part of their decision-making to the published league tables: the affordances of the league tables help them to tackle the problems of information overload and information (non-) reliability (cf. Espeland and Sauder, 2016: 13, 48, 50–53). The information overload pertains to the fact that there is too much (rather than too little) information available about different law schools, including the marketing information provided in the websites and advising practices of law schools, which makes it time consuming to acquire information about different law schools and cognitively demanding to remember the acquired information. The lack of reliability of available information pertains to the fact that it is largely provided by the law schools themselves for marketing purposes. By contrast, more valuable inside information about law schools, especially information that could be harmful to their reputations, is hard to obtain to prospective law students unless they have reliable contacts within law schools. This creates a situation of asymmetrically distributed information. Both of these factors make the affordances of the USN league table of law schools appealing to prospective law students and their parents since they apparently provide a means to avoid both of these problems. Similar considerations seem to apply to the ways in which prospective international students use league tables of global university rankings that closely resemble to the USN league table of law schools (cf. Hazelkorn, 2015: chapter 4).

How do policymakers use the league tables of global university rankings?

In addition to prospective students, there are also other extra-academic groups who increasingly use university rankings in their decision-making, such as employers, funders, and policymakers (e.g. Erkkilä, 2013; Espeland and Sauder, 2016; Hazelkorn, 2015). In this section, I will provide some reasons why I think that the above concepts

and ideas can be extended to analyze how policymakers tend to outsource part of their decision-making about universities to the global university rankings.

It is useful to begin by reminding four differences between the case of USN rankings of law schools and the case of global university rankings. First, as I already suggested above, while the USN rankings of law schools claim to provide information about the relative quality of legal education, many global university rankings claim to provide information about the relative quality of research. Second, while there are about 200 law schools in the United States of which nearly all are ranked by the USN rankings, there are over 18,000 universities in the world of which global university rankings rank about 500 institutions or fewer (Hazelkorn, 2015: 87). Third, while the USN ranking of law schools is undeniably the most influential law school ranking in the United States, there are at least three popular global university rankings – the *QS* ranking, the *ARWU*, and the *Times Higher Education* ranking – whose ranking metrics are slightly different (e.g. Marginson, 2014). Finally, policymaker's motivations for relying on the results of global university rankings are different from those of prospective law students: instead of choosing which university to apply to, policymakers are typically interested in increasing the transparency, efficiency, and competitiveness of universities as well as making them accountable for taxpayers and other stakeholders (e.g. Espeland and Sauder, 2007, 2016: chapter 7; Hazelkorn, 2015). These demands for universities have emerged from the growing distrust on university-based science and higher education, fueled by the New Public Management ideology, and from the attempts to democratize knowledge and to economize universities to boost the growth and competitiveness of national economies (e.g. Berman, 2014; Erkkilä, 2013; Erkkilä and Piironen, 2018; Hazelkorn, 2015: chapter 5; Hallonsten, 2021; Münch, 2014).

Hazelkorn (2015: 87) argues that global university rankings 'have generated a perception among the public, policymakers and stakeholders that only those within the top 20, 50 or 100 are worthy of being called excellent'. This in turn has intensified 'the battle for world-class excellence' since governments in many countries have explicitly set the goal for their 'top universities' to enter the small group of excellent world-class universities defined by the global university rankings (Hazelkorn, 2015: chapter 5; Münch, 2014, chapter 1). Given the policymakers' growing interest in international comparisons of universities and the huge number of universities in the world, they also face the previously discussed problems of information overload and information (non-) reliability. Hence, the league tables of global university rankings provide them with a tool for making comparative judgments about 'the relative quality of universities', especially if they are ready to equate relative quality with ranking status by relying on the league table metaphor. In addition, the affordances of the league tables of university rankings may also provide appealing shortcuts for policymakers when they are forced to make fast decisions and to provide 'objective evidence' to legitimate their policy decisions. For example, policymakers may take the declining rankings of the universities in their country as objective evidence for the weakening of the quality of scientific research and higher education. However, since policymakers in democratic societies are supposed to make political choices that they are able to justify to citizens, the outsourcing of policy decisions to global university rankings is highly problematic since the ranking metrics include hidden value judgments that are not justified by anyone.

There also seems to be an indirect path through which university rankings affect policy decisions regarding universities. As demonstrated by Richard Münch (2014: chapter 5), governments in many countries have restructured the public funding of universities in recent decades, such that the share of external funds allocated on a competitive basis has increased at the expense of block grants (see also Cantwell and Kauppinen, 2014). This, in turn, seems to have increased the impact of university rankings on publicly funded universities. For example, the increased share of competitive funding has allowed policymakers – often with the help of Organisation for Economic Co-operation and Development (OECD) – to set the standards for the distribution of competitive funding in a way that imitates the metrics of global university rankings, typically emphasizing the number of peer-reviewed publications in highly ranked scientific journals (Münch, 2014: 23, 35–37; see also Erkkilä and Piironen, 2018; Hazelkorn, 2015: chapter 5). Due to the previously described Matthew effects between countries, universities, and disciplines, these policy decisions tend to lead to narrowing down the diversity of scientific research which is a necessary condition for scientific renewal and progress of scientific knowledge in the long run (Münch, 2014: chapter 5; see also Brink, 2018; Espeland and Sauder, 2009; Halfman and Radder, 2015; Hallonsten, 2021). However, it is clear that there are also many other factors in addition to university rankings that affect policy decisions regarding universities and that the impact of global university rankings varies between countries.

Conclusion

In her comprehensive book of university rankings, Hazelkorn (2015: x) claims that ‘[d]espite the ongoing criticisms about the appropriateness or otherwise of the methodology, [global university] rankings are widely perceived as *the* international measure of quality’. She also reviews diverse empirical studies that have documented how university rankings have reshaped universities and higher education systems all over the world while contending at the same time that ‘there is no such thing as objective ranking’ (Hazelkorn, 2015: 53). Although Hazelkorn mostly focuses on global university rankings, these views apply to some national university rankings, such as the USN ranking of law schools, as was argued above.

At the beginning of this article, I termed this situation as the paradox of university rankings. Then I argued that the concepts of commensuration, status, and reactivity are crucially important for understanding the social dynamics of university rankings. By using these and some other concepts from sociology, cognitive linguistics, and cognitive anthropology, I developed a theoretical framework for explaining the paradox of university rankings. It consisted of the four complementary parts: the first part distinguished three temporal stages in the social process of commensuration of universities by rankings. This commensuration process constructs a status hierarchy between universities that begins to affect public perception of the relative quality of universities insofar as the repeatedly published ranking results gain wide attention among their extra-academic audiences. The second part described four social mechanisms – self-fulfilling prophecy, Matthew effect, data manipulation based on reverse engineering, and narrative – through which university rankings generate reactive outcomes. I argued that the responses of

extra-academic audiences to the published ranking results trigger the reactive mechanisms through which rankings tend to transform the core activities of universities instead of just measuring them. The third part provided an account of the league table metaphor that, I hypothesized, is a cognitive mechanism that shapes how many extra-academic audiences understand the results of university rankings. The league table metaphor that links the conceptual domain of team sports and the conceptual domain of universities also explains how the phenomenon of status dynamics is metaphorically transferred to the relations between universities. Finally, the fourth part identified affordances of the published league tables that enable extra-academic actors to outsource part of their decision-making to the university rankings. In particular, I argued that prospective students and policymakers are prone to rely on these affordances in their decision-making, especially if they have internalized the league table metaphor. While the first two parts of this theoretical framework critically synthesized earlier research on university rankings, the latter two parts provided new cognitive perspectives on university rankings.

In general terms, a plausible explanation for the paradox of university rankings is that the interrelated operations of the previously described mechanisms and processes have partially shifted the power to control the evaluation of research and education practices in universities from the research and education personnel to the various extra-academic groups, such as media, policymakers, politicians, administrators, and professional managers. This shift makes the paradox of university rankings understandable since it has created a situation in which scientific research and higher education are increasingly evaluated by using quantitative metrics that are counterproductive to these practices (for similar arguments, see Brink, 2018; Halffman and Radder, 2015; Hallonsten, 2021; Münch, 2014). As I already hinted above, the above processes and mechanisms could be contextualized by means of analyzing the broader societal processes pertaining to economization, commodification, and democratization of science. These broader processes were recently addressed in the special issue of this journal that included a number of responses to Olof Hallonsten's (2021) article, which analyzed and criticized simplistic quantitative metrics that are increasingly used in external evaluation of science. An interesting issue pertaining to this debate concerns the question of whether, or to what extent, the contributions of policymakers and university managers to the paradox of university rankings relate to their attempts to make universities accountable and to economize universities by means of introducing academic capitalist practices or whether, or to what extent, it results from their insufficient understanding of how higher education and scientific research work.

However, more empirical research is needed to apply, evaluate, and elaborate the theoretical framework developed above. It is a task of empirical research to find out how the reactive mechanisms play out in different cases of university rankings. Accordingly, it may be possible to identify other reactive mechanisms in addition to those discussed above. Another task for empirical research would be to analyze and compare the decision-making of different audiences of university rankings in a more detailed manner, for example, by using ethnographic methods and interviews. This would allow empirical testing and elaboration of the ideas pertaining to the league table metaphor and cognitive outsourcing, since it is possible that the league table metaphor is not the only metaphor that extra-academic actors use to make sense of the results of university rankings. It would also be interesting to study in detail whether and how university rankings have

reshaped the practices of epistemic evaluation in different sciences. Finally, one could study why many policymakers rely on the biased and methodologically flawed university rankings in their decision-making about universities while claiming that policies should be based on scientific evidence. The above theoretical framework provides conceptual resources and theoretical ideas for conducting empirical studies on these issues.

Acknowledgements

The author would like to thank two anonymous reviewers for their excellent comments. Marja Alastalo, Mikko Hyyryläinen, Omar Lizardo, Ronny Puustinen, Matti Sarkia, Soile Veijola, and Petri Ylikoski also provided valuable comments on the earlier versions of this article for which the author is thankful.

Funding

The author disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The author received funding for this publication from the Kone Foundation.

ORCID iD

Tuukka Kaidesoja  <https://orcid.org/0000-0002-9633-2944>

Notes

1. Espeland and Sauder's (2007, 2009, 2016, 2009; Sauder and Espeland, 2009) case study traced the reactive outcomes of the *U.S. News & World Report* (USN) rankings of law schools over a 10-year period by using a wide range of data from diverse sources. The USN rankings of law schools is an extreme case among university rankings because the ranking is comprehensive, highly influential, and lacks serious competitors (Espeland and Sauder, 2016: 5). Hence, Espeland and Sauder's case is different from the case of global university rankings since there are many popular global rankings that use slightly different sets of indicators and that are far from comprehensive, which limits the influence of each ranking (e.g. Brink, 2018; Erkkilä, 2013; Hazelkorn, 2015; Marginson, 2014). However, Espeland and Sauder's case provided them with a good opportunity to identify those theoretically generalizable social mechanisms through which rankings generate reactive outcomes (see Ylikoski, 2019).
2. It could be claimed that the inclusion of reputation surveys to ranking metrics allow them to take into account differences in perceptions of quality of those who answer these surveys. However, reputation surveys are not included in all university rankings and their weighing is less than 50% in those ranking metrics that include them (e.g. Marginson, 2014). In addition, their respondents represent only those groups who are considered as experts and the criteria that they use in their ratings cannot be automatically assumed to reflect their independent value commitments regarding quality (see my discussion on self-fulfilling prophecies).
3. There are differences in scholarly opinions regarding the extent to which scientific values vary between – or are weighted differently in – different disciplines and about the legitimate roles that (non-epistemic) social, cultural, and moral values may play in research process (for a good review, see Douglas, 2011). Despite these differences, there is a broad consensus among those scholars who care about the epistemic capacities of science that the excessive influence of extra-academic commercial and political interests on research process tend to corrupt its results. The excessive influence of this kind may involve replacement of empirical evidence

with judgments based on political or commercial interests. For example, the Lysenko case in the Soviet Union and the case of tobacco-industry-funded research on the health effects of smoking in the United States are well-known historical examples of this type of corruption of research process.

4. Recent attempts to develop new university rankings that do not rely on composite multi-indicators – such as EU’s U-multirank – address some of the discussed methodological problems, but they include their own problems and weaknesses. For example, U-multirank not only relies on reputation surveys whose reliability and validity are suspect but it also lacks transparency (e.g. Marginson, 2014: 56).
5. Espeland and Sauder (2016: 28) propose that commensuration – that they understand as ‘the process of turning qualities into quantities on a shared metric’ – is one of the mechanisms through which the USN rankings generate reactive outcomes. By contrast, I have considered commensuration as a more fundamental process to the mechanisms that generate reactive outcomes since the operation of these mechanisms requires that the ranking metric has already been constructed and put into use. In addition, I have extended commensuration processes to include three stages, whereas Espeland and Sauder (2016: 28–30) mostly focus on the first stage in their characterization of the commensuration mechanism. Nevertheless, in their early work, Espeland and Stevens (1998; also Espeland, 2002) seem to assume that commensuration comprises the implementation of the metric and the outcomes of this implementation in addition to the mere construction of the metric.
6. Although some authors have previously discussed similarities between university rankings and the league tables of sport teams (e.g. Marginson, 2014: 45; Ringel and Werron, 2020; Tight, 2000), they have not developed the idea of league table metaphor nor analyzed the affordances of the concrete league tables.
7. It is also possible that the league tables of sports teams provided a historical model for the first designers of university rankings. This is because the first league tables in team sports were developed in the league systems of the US-American Baseball League founded in the 1870s and the British Football League founded in 1888 (Ringel and Werron, 2020: 153), while the first league table of universities was published in 1906 by James Mckeen Cattell, editor of *Science* (Godin, 2007: 708). However, my arguments do not require the previous claim to be true since my focus is on how extra-academic actors understand and use the current university rankings.

References

- Berman EP (2014) Not just neoliberalism: Economization in US science and technology policy. *Science, Technology, & Human Values* 39(3): 397–431.
- Biagioli M, Lippman A (eds) (2020) *Gaming the Metrics: Misconduct and Manipulation in Academic Research*. Cambridge, MA: The MIT Press.
- Bougnol M, Dulá JL (2014) Technical pitfalls in university rankings. *Higher Education* 69: 859–866.
- Bowman NA, Bastedo MN (2009) Getting on the front page: Organizational reputation, status signals, and the impact of U.S. News and World Report on student decisions. *Research in Higher Education* 50: 415–436.
- Brankovic J, Ringel L, Werron T (2018) How rankings produce competition: The case of global university rankings. *Zeitschrift für Soziologie* 47(4): 270–288.
- Brink C (2018) *The Soul of a University: Why Excellence Is Not Enough*. Bristol: Bristol University Press.
- Brown R (1988) *Group Processes*. 2nd ed. Oxford: Blackwell Publishing.

- Campbell DT (1957) Factors relevant to the validity of experiments in social settings. *Psychological Bulletin* 54: 297–312.
- Campbell DT (1979) Assessing the impact of planned social change. *Evaluation and Program Planning* 2: 67–90.
- Cantwell B, Kauppinen I (eds) (2014) *Academic Capitalism in the Age of Globalization*. Baltimore, MD: Johns Hopkins University Press.
- Daston LJ (1992) Objectivity and the escape from perspective. *Social Studies of Science* 22(4): 597–618.
- Douglas H (2011) Facts, values and objectivity. In: Jarvie IC, Zamora-bonilla J (eds) *The SAGE Handbook of the Philosophy of the Social Sciences*. London: Sage, pp.513–529.
- Erkkilä T (ed.) (2013) *Global University Rankings: Challenges for European Higher Education*. Houndsmills: Palgrave Macmillan:
- Erkkilä T, Piironen O (2018) *Rankings and Global Knowledge Governance: Higher Education, Innovation and Competitiveness*. Cham: Palgrave Macmillan.
- Espeland WN (2002) Commensuration and cognition. In: Cerulo K (ed.) *Culture in Mind: Toward a Sociology of Culture and Cognition*. London: Routledge, pp.63–88.
- Espeland WN, Sauder M (2007) Rankings and reactivity. *American Journal of Sociology* 113(1): 1–40.
- Espeland WN, Sauder M (2009) Rankings and diversity. *Southern California Review of Law and Social Justice* 18(3): 587–608.
- Espeland WN, Sauder M (2016) *Engines of Anxiety: Academic Rankings, Reputation, and Accountability*. New York: Russell Sage Foundation.
- Espeland WN, Stevens ML (1998) Commensuration as social process. *Annual Review of Sociology* 24: 313–343.
- Espeland WN, Stevens ML (2008) A sociology of quantification. *Archives Européennes de Sociologie* XLIX: 401–436.
- Espeland WN, Yung V (2019) Ethical dimensions of quantification. *Social Science Information* 58(2): 238–260.
- Fernandez-Cano A, Curiel-Marin E, Torralbo-Rodriguez M, et al. (2018) Questioning the Shanghai ranking methodology as a tool for the evaluation of universities: An integrative review. *Scientometrics* 116: 2069–2083.
- Fowles J, Frederickson GH, Koppell JGS (2016) University rankings: Evidence and a conceptual framework. *Public Administration Review* 76(5): 790–803.
- French DP, Sutton S (2010) Reactivity of measurement in health psychology: How much of a problem is it? What can be done about it? *The British Journal of Health Psychology* 15: 453–468.
- Gadd E (2020) University rankings need a rethink. *Nature* 587: 523.
- Gibson JJ (1979) *The Ecological Approach to Visual Perception*. Boston, MA: Houghton Mifflin Harcourt.
- Godin B (2007) From eugenics to scientometrics: Galton, Cattell, and men of science. *Social Studies of Science* 37(5): 691–728.
- Halfman W, Radder H (2015) The academic manifesto: From an occupied to a public university. *Minerva* 53(2): 165–187.
- Hallonsten O (2021) Stop evaluating science: A historical-sociological argument. *Social Science Information* 60(1): 7–26.
- Hazelkorn E (2015) *Rankings and the Reshaping of Higher Education: The Battle for World-class Excellence*. 2nd ed. Houndsmills: Palgrave Macmillan.
- Hedström P, Ylikoski P (2010) Causal mechanisms in the social sciences. *Annual Review of Sociology* 39: 46–67.

- Hutchins E (1995) *Cognition in the Wild*. Cambridge, MA: The MIT Press.
- Kauppi N (2018) The global ranking game: Narrowing academic excellence through numerical objectification. *Studies in Higher Education* 43(10): 1750–1762.
- Kirsh D (2010) Thinking with external representations. *AI & Society* 25: 441–454.
- Lakoff G, Johnson M (1999) *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*. New York: Basic Books.
- Lakoff G, Johnson M (2003) *Metaphors We Live by (With a New Afterword)*. Chicago, IL; London: The University of Chicago Press.
- Marginson S (2013) The impossibility of capitalist markets in higher education. *Journal of Education Policy* 28(3): 353–370.
- Marginson S (2014) University rankings and social science. *European Journal of Education* 49(1): 45–59.
- Merton RK (1968a) *Social Theory and Social Structure*. Extended ed. New York: The Free Press.
- Merton RK (1968b) The Matthew effect in science. *Science* 159(3810): 56–63.
- Muller JZ (2018) *The Tyranny of Metrics*. Princeton, NJ: Princeton University Press.
- Münch R (2014) *Academic Capitalism: Universities in the Global Struggle for Excellence*. London: Routledge.
- Münch R, Schäfer LO (2014) Rankings, diversity, and the power of renewal in science: A comparison between Germany, the UK and the US. *European Journal of Education* 49(1): 60–76.
- Norman DA (1991) Cognitive artifacts. In: Carroll JM (ed.) *Designing Interaction*. Cambridge: Cambridge University Press, pp.17–38.
- Norman DA (1993) *Things That Make Us Smart: Defending Human Attributes in the Age of the Machine*. Reading, MA: Addison–Wesley.
- Podolny J, Lynn F (2009) Status. In: Hedström P, Bearman P (eds) *The Oxford Handbook of Analytical Sociology*. Oxford: Oxford University Press, pp.544–565.
- Porter TM (1995) *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton, NJ: Princeton University Press.
- Power M (1997) *The Audit Society*. Oxford: Oxford University Press.
- Ringel L, Werron T (2020) Where do rankings come from? Historical-sociological perspective on the history of modern rankings. In: Epple A, Erhart W, Grave J (eds) *Practices of Comparing*. Bielefeld: Bielefeld University Press, pp.137–170.
- Salmi J, Saroyan A (2007) League tables as policy instruments. *Higher Education Management and Policy* 19(2): 1–38.
- Sauder M, Espeland WN (2009) The discipline of rankings: Tight coupling and organizational change. *American Sociological Review* 74: 63–82.
- Tight M (2000) Do league tables contribute to the development of a quality culture? *Higher Education Quarterly* 54(1): 22–42.
- Timmermans S, Epstein S (2010) A world of standards but not a standard world: Toward a sociology of standards and standardization. *The Annual Review of Sociology* 36: 69–89.
- Ylikoski P (2019) Mechanism-based theorizing and generalization from case studies. *Studies in History and Philosophy of Science Part A* 78: 14–22.

Author biography

Tuukka Kaidesoja is a researcher and a member of the Centre for Philosophy of Social Science (TINT) at the University of Helsinki. His research interests range from sociological theory and science policy to methodological and philosophical issues concerning the social sciences and the relations between the social and cognitive sciences. Currently, he leads a research project that addresses the relations between the social and cognitive sciences (for more information, see <https://blogs.helsinki.fi/perspectives-on-cognitive-social-sciences/>).