



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Von Performativität zu Generativität

Citation for published version:

Mennicken, A & Kornberger, M 2021, 'Von Performativität zu Generativität: Bewertung und ihre Folgen im Kontext der Digitalisierung', *Kolner Zeitschrift fur Soziologie und Sozialpsychologie*.
<https://doi.org/10.1007/s11577-021-00755-x>

Digital Object Identifier (DOI):

[10.1007/s11577-021-00755-x](https://doi.org/10.1007/s11577-021-00755-x)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

Kolner Zeitschrift fur Soziologie und Sozialpsychologie

Publisher Rights Statement:

Publisher Copyright: © 2021, The Author(s).

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.





Von Performativität zu Generativität: Bewertung und ihre Folgen im Kontext der Digitalisierung

Andrea Mennicken · Martin Kornberger

Eingegangen: 1. November 2020 / Angenommen: 5. Dezember 2020
© Der/die Autor(en) 2021

Zusammenfassung In diesem Beitrag gehen wir zwei Fragen nach. Erstens, inwiefern kann ein Fokus auf Praktiken der Bewertung zur Klärung des Zusammenhangs zwischen Vergleich, Kategorisierung und Quantifizierung beitragen? Und zweitens, inwieweit verhilft er dazu, die Voraussetzungen und Folgen neuer digitaler Formate, wie etwa Vergleichsportale, Empfehlungssysteme oder Screening- und Scoring-Verfahren, besser zu verstehen? Im Gegensatz zur Auffassung, dass Wert im Bewusstsein des Subjekts (als Präferenz) existiert oder eine objektive Eigenschaft eines Gutes ist, argumentieren wir, dass ein Gut erst durch Bewertungspraktiken und -technologien wie Ratings, Rankings und andere evaluative Infrastrukturen als wertvoll oder als nicht wertvoll bestimmt wird. Wir entwickeln einen analytischen Rahmen für die Untersuchung interaktiver, digitalisierter Bewertungstechnologien und schlagen vor, dass sich solche Technologien besser analysieren und verstehen lassen, wenn wir den Fokus auf drei Elemente legen: erstens auf evaluative Infrastrukturen, verstanden als *Bewertungsregime* – und nicht als einzelne Bewertungsinstrumente, zweitens auf das *Protokoll* als ein evaluativen Infrastrukturen eigenes Machtinstrument, wobei Macht in diesem Fall paradoxerweise (und im Unterschied zum Disziplinarregime) zugleich verteilt und konzentriert ist, und drittens auf das *generative* (statt nur performative) Potenzial solcher digitalisierter Bewertungsregime, also deren Fähigkeit, neue Werte und Kategorisierungen mittels evaluativer Infrastrukturen zu produzieren.

Übersetzung Florian Wüstholtz

A. Mennicken (✉)

Department of Accounting, London School of Economics and Political Science
Houghton Street, London, WC2A 2AE, Großbritannien
E-Mail: a.m.mennicken@lse.ac.uk

M. Kornberger

University of Edinburgh Business School
29 Buccleuch Place, Edinburgh, EH8 9JS, Großbritannien
E-Mail: martin.kornberger@ed.ac.uk

Schlüsselwörter Evaluative Infrastrukturen · Digitalisierte Bewertung · Kommensurierung · Macht · Pragmatismus

From Performativity to Generativity: Valuation and Its Consequences in the Context of Digitization

Abstract In this article we explore (1) how the study of valuation practices can illuminate the relationships between comparison, classification, and measurement, and (2) how a focus on valuation practices can be used to critically analyze the conditions and consequences of new digital formats, such as comparison portals, recommender systems, and screening and scoring techniques. Departing from the view that value either resides in the mind of the subject (as preference) or is an objective property of a good, we argue that it is through valuation practices and technologies such as ratings, rankings, and other evaluative infrastructures that a good can be estimated as valuable or not. We develop an analytical framework for the study of interactive, digitized valuation technologies and propose that an analysis of such technologies can be enhanced by focusing on 1) evaluative infrastructures as regimes of valuation (as opposed to a narrow focus on individual devices); 2) protocol as the specific power effect of evaluative infrastructures, where, paradoxically (and in opposition to disciplinary power regimes), power is at once distributed and concentrated; and 3) the generative potential of such valuation regimes, i.e., the production of new values and categorizations through digital evaluative infrastructures (rather than performativity).

Keywords Evaluative infrastructures · Digitized valuation · Commensuration · Power · Pragmatism

„Nur als Schaffende! – Diess hat mir die größte Mühe gemacht und macht mir noch immerfort die größte Mühe: einzusehen, dass unsäglich mehr darin liegt, *wie die Dinge heissen*, als was sie sind. Der Ruf, Name und Anschein, die Geltung, das übliche Maass und Gewicht eines Dinges – im Ursprunge zu allermeist ein Irrthum und eine Willkürlichkeit, den Dingen übergeworfen wie ein Kleid und seinem Wesen und selbst seiner Haut ganz fremd – ist durch den Glauben daran und sein Fortwachsen von Geschlecht zu Geschlecht dem Dinge allmählich gleichsam an- und eingewachsen und zu seinem Leibe selber geworden: der Schein von Anbeginn wird zuletzt fast immer zum Wesen und *wirkt* als Wesen! Was wäre das für ein Narr, der da meinte, es genüge, auf diesen Ursprung und diese Nebelhülle des Wahnes hinzuweisen, um die als wesenhaft geltende Welt, die sogenannte 'Wirklichkeit', *zu vernichten!* Nur als Schaffende können wir vernichten! – Aber vergessen wir auch dies nicht: es genügt, neue Namen und Schätzungen und Wahrscheinlichkeiten zu schaffen, um auf die Länge hin neue 'Dinge' zu schaffen“ (Friedrich Nietzsche, *Die fröhliche Wissenschaft* (1999 [1882], S. 423)).

1 Einleitung

Unser Beitrag verfolgt zwei Fragen. Erstens, inwiefern kann ein Fokus auf Praktiken der Bewertung zur Klärung des Zusammenhangs zwischen Vergleich, Kategorisierung und Quantifizierung beitragen? Und zweitens, inwieweit verhilft er dazu, die Voraussetzungen und Folgen neuer digitaler Formate, wie etwa Vergleichsportale, Empfehlungssysteme oder Screening- und Scoring-Verfahren, besser zu verstehen? Um diese beiden Fragen zu beantworten, werden wir zunächst nachzeichnen, was Soziologinnen, Anthropologinnen und andere Sozialwissenschaftler auf der einen und Ökonomen auf der anderen Seite unter Wert oder Werten verstehen. Während Sozialwissenschaftler von einer Vielfalt von Wertsphären ausgehen, begreifen Ökonomen Wert im Singular im Sinne eines Nutzens, der sich monetär messen und quantifizieren lässt.

Vor diesem intellektuellen Hintergrund – und inspiriert von Friedrich Nietzsche und John Dewey – werden wir dafür plädieren, Bewertung als einen Prozess zu begreifen, durch den Dinge wertvoll gemacht werden. Unser Hauptargument ist, dass Güter, wie z. B. ein Buch, ein Kunstwerk, Whiskey, Gold, grüne Energie, Biolebensmittel und viele andere mehr, nicht aufgrund ihrer intrinsischen, objektiven Qualitäten und auch nicht aufgrund individueller Präferenzen wertvoll sind. Vielmehr argumentieren wir, dass ein Gut durch Bewertungspraktiken und -technologien, wie Ratings, Rankings und andere evaluative Infrastrukturen, als wertvoll oder als nicht wertvoll eingestuft werden kann.

Solche Bewertungspraktiken vermitteln zwischen dem Bewusstsein (von Subjekten) und Gütern (Objekten) und wirken dabei als Filter, die kategorisieren, systematisieren und Vergleiche ermöglichen. Mit hoher Wahrscheinlichkeit beurteilen wir ein Buch nicht nach dem Cover und schon gar nicht nach dem Preis – aber wir orientieren uns an den Rezensionen, dem Rang auf Bestsellerlisten oder den Preisen, die eine Autorin erhalten hat. Wir wären kaum in der Lage, Biogemüse von konventionellem Gemüse zu unterscheiden, wenn es keine Labels gäbe, die garantieren, dass bestimmte Standards während des Produktionsprozesses eingehalten wurden. Und die meisten Menschen wüssten kaum etwas über die wirtschaftlichen Aussichten eines Landes, wenn es keine Ratingagenturen gäbe, die Länder mit A's oder B's einstufen, oder wenn keine quantifizierten Vergleichsmaßstäbe zur Verfügung stünden wie etwa das Bruttoinlandsprodukt (BIP). Kurzum: Bewertungspraktiken versehen Dinge mit einem Wert, sie vermitteln zwischen „Individuum“ und „Welt“ (Konsument und Produzentin) und tragen damit zur Hervorbringung und Stabilisierung von Märkten bei.

Im Folgenden entwickeln wir einen analytischen Rahmen für die Untersuchung interaktiver, digitalisierter Bewertungstechnologien, wie sie von Plattformen in Form von Ratings, Rankings, Reviews und Kundenaudits eingesetzt werden. Wir argumentieren, dass sich solche Technologien besser analysieren und verstehen lassen, wenn wir den Fokus auf drei Elemente legen: 1. auf *evaluative Infrastrukturen*, verstanden als Bewertungsregime – und nicht als einzelne Bewertungsinstrumente, 2. auf das *Protokoll* als ein evaluativen Infrastrukturen eigenes Machtinstrument – wobei Macht in diesem Fall paradoxerweise (und im Unterschied zum Disziplinarregime) zugleich verteilt und konzentriert ist, und 3. auf das *generative* (statt bloß performa-

tive) Potenzial solcher digitalisierter Bewertungsregime – also deren Fähigkeit, neue Werte und Kategorisierungen mittels evaluativer Infrastrukturen zu produzieren.

Unser Beitrag gliedert sich in drei Teile. Zuerst umreißen wir den intellektuellen Hintergrund („Leinwand“) und machen die Leserinnen und Leser mit den Grundzügen der Wert- und Bewertungsdebatte vertraut. In einem zweiten Teil skizzieren wir unseren analytischen Ansatz für die Untersuchung von Bewertungspraktiken („Bild“) und in einem dritten Teil leiten wir aus diesem Bild eine neue Forschungsperspektive ab („Ausblick“).

2 Die Leinwand: Konzeptualisierungen von Wert

Die Frage, was wir unter „Werten“ verstehen sollen, ist eines der großen Probleme, mit denen sich Wissenschaft, Wirtschaft und Gesellschaft beschäftigen. Wir sprechen von ökonomischem Wert, den Werten von politischen Parteien und Bewegungen, kulturellen Werten sowie Werten, die wir einzelnen Individuen zuschreiben – persönliche Präferenzen, Überzeugungen oder Glaubensvorstellungen. Bereits diese vermeintlich einfache Liste wirft ein begriffliches Rätsel auf: Werte sind hier einerseits das, was wir teilen und was uns verbindet (kulturelle Werte), aber auch das, was uns unterscheidet und einzigartig macht (Persönlichkeit). Zudem ist Wert objektiviert in harter Währung (wie beim Unternehmenswert), aber auch das, was eine Gesellschaft zusammenhält (wie etwa ideologische und politische Werte).

Anstatt dieses Rätsel direkt anzugehen, hat die akademische Debatte eine arbeitsteilige Form angenommen – und damit Luhmann'sche „Zonen der Indifferenz“ miterfunden, in denen das Rätsel verschwindet. Stark (2009) spricht in diesem Zusammenhang von „Parsons' Pakt“: Soziologen und Ökonominen einigten sich darauf, dass erstere sich mit der Untersuchung gesellschaftlicher Werte befassen, während letztere sich auf die Frage des ökonomischen Werts konzentrieren sollen. Diese Arbeitsteilung setzte schon vor Parsons ein. Bereits Max Weber (2002 [1922]) differenzierte zwischen verschiedenen „Wertsphären“, die durch je eigene Prinzipien und Praktiken charakterisiert sind. Weber unterschied dabei sechs Sphären: Religion, Wirtschaft, Politik, Ästhetik, Eros und Intellektualismus. Diese Unterscheidung hat die sozialwissenschaftliche Debatte über Werte seitdem geprägt. Von Boltanskis und Thévenots „Wertordnungen“ (2006) bis hin zu Michael J. Sandels (2000) Versuch, die Grenze zwischen ökonomischen und moralischen Werten zu definieren, hat sich die Debatte vor allem darum gedreht, wie sich Wertsphären voneinander abgrenzen lassen und welche Werte welche Wertsphäre bestimmen. So hält z. B. Guy Oakes in Bezug auf Weber fest:

The antinomy of values – the doctrine that value positions, value axioms and value spheres can be constituted only in opposition to alternatives – is the source of some of Weber's more memorable images: the idea of an irresolvable axiological struggle and the metaphor of value spheres as kingdoms ruled by gods and demons locked in perpetual warfare (Oakes 2003, S. 30).

Der Monotheismus der Ökonomie bildet dabei die Antithese zu den polyzentrischen Wertsphären – von Boltanski und Thévenot „cités“ genannt. Seit Jeremy Bent-

ham haben Ökonomen das Problem der Werte auf die Frage nach dem „Wie viel“ reduziert. So betrachtete der Utilitarist Bentham „Lust“ („pleasure“) und „Schmerz“ („pain“) als die „zwei souveränen Gebieter“ über die Menschheit und leitete daraus sein ultimatives Leitprinzip ab: „Das größte Glück der größten Zahl ist der Maßstab für richtig und falsch“. Als Erläuterung führte er das Beispiel von zwei Personen an, die über die Vorzüge eines Gesetzes streiten. Ihre Argumente seien, so Bentham, nichts anderes als „Weiberschelte“ und

[...] childish altercation which is sure to irritate, and which can never persuade. – „I say, the legislature cannot do this – I say, that it *can*. I say, that to do this, exceeds the bounds of its *authority* – I say, it does *not*.“ It is evident, that a pair of disputants setting out in this manner, may go on irritating and perplexing one another for everlasting, without the smallest chance of ever coming to an agreement (Bentham 1988 [1776], S. 104).

Er schlug vor, das Rätsel dadurch zu lösen, dass Behauptungen in quantifizierbare Einheiten von Lust und Schmerz übersetzt werden: Lasst uns rechnen und lasst die Zahlen sprechen. Entsprechend nahm der Streit nun folgende Form an:

„I say, that the mischiefs of the measure in question are to such an amount – I say, not so, but to a less. – I say, the benefits of it are only to such an amount. – I say, not so, but to a greater.“ – This, we see, is a ground of controversy very different from the former (Bentham 1988 [1776], S. 104).

Bentham schlägt also eine Methode vor, um die Entscheidungsfindung aus rationalen Berechnungen abzuleiten. Die Subtilitäten und die Metaphysik von Jean-Jacques Rousseaus Gesellschaftsvertrag oder Thomas Hobbes' imaginärem Leviathan lehnte er ab. Sie sind für ihn nichts anderes als Vorstellungen (Leviathan) oder Repräsentationen (Bürgervereinigung) von etwas, das womöglich nicht jenseits des Bewusstseins existiert. In Benthams Welt ist es die Berechnung von Lust und Schmerz, die unsere Entscheidungsfindung anleitet – und nicht die politische Urteilskraft. Es ist eine Welt der Kosten-Nutzen-Analyse, nicht eine Welt der moralischen Debatte zwischen entgegengesetzten Überzeugungssystemen. Die nicht gerade triviale Frage, wie sich Lust und Schmerz messen lassen, versuchte er in seinen *Principles of Morals and Legislation* zu beantworten. Um den jeweiligen Wert von Lust und Schmerz zu bestimmen, schlug er verschiedene Maße vor, etwa Intensität, Dauer, Gewissheit und Unsicherheit. Sind diese erst einmal definiert, lassen sich die verschiedenen Lust- und Schmerzwerte aufsummieren:

Sum up all the values of all the pleasures on the one side, and those of all the pains on the other. The balance, if it to be on the side of pleasure, will give the good tendency of the act upon the whole [...] (Bentham 1823 [1789], S. 31).

Die Wirtschaftstheorie lässt sich dabei zumindest teilweise als Versuch verstehen, die Bentham'sche Gleichung mittels einer immer ausgefeilteren und generalisierbaren Wissensinfrastruktur zu lösen. Der Nobelpreisträger Gary Becker ist einer der wohl scharfsinnigsten Vertreter der Tendenz, sämtliche Lebenssphären unter ein einziges Bewertungssystem von ökonomischer Lust und Schmerz zu subsumieren. In der Praxis belegen jedoch die Probleme bei der finanziellen Entschädigung der Opfer

von 9/11 (Feinberg 2005) oder bei der Bewertung von Umweltschäden aus massiven Ölkatastrophen (Fourcade 2011) die Schwierigkeiten einer solchen Kommensurierung (Espeland und Stevens 1998).

Zusammenfassend lässt sich sagen, dass sich die verschiedenen Konzeptualisierungen von Wert – als ökonomischer Wert oder als soziale Wertsphären – gegenseitig widersprechen und ausschließen. Auf der einen Seite sprechen wir von Werten und Wertsphären in normativen Begriffen, die, wenn nicht objektiv, so doch dem individuellen Bewusstsein äußerlich sind. Werte in diesem Sinne sind nicht kommensurabel und folgen ihrer je eigenen Logik. Auf der anderen Seite ist Wert im Sinne von Nutzen subjektiv und hat seinen Sitz im menschlichen Bewusstsein (individuelle Präferenzen). Wert im Sinne von Nutzen folgt einer einheitlichen ökonomischen Logik, gemäß deren Wert quantifizierbar und vergleichbar ist.

Diese beiden Perspektiven kollidieren. Wenn Organisationen von „Wertschöpfung“ sprechen, meinen sie den monetären Ertrag des Unternehmenswerts. Aber sie meinen damit auch die Schaffung von ästhetischen, funktionalen, sozialen und anderen Werten, die den Kundinnen gefallen. Der Nutzen liegt im Auge des Betrachters, der Unternehmenswert im Tresor einer Bank. Auch wenn die beiden Konzepte kaum weiter auseinanderliegen könnten, verwenden wir das gleiche Wort dafür. Das Feld der *Valuation Studies* (im Folgenden: Bewertungsstudien), auf das wir im nächsten Abschnitt eingehen, teilt diese Diagnose und bietet eine diese Teilung überwindende alternative Forschungsperspektive, indem sie den Fokus weg von (vor-)etablierten Werten auf die *Praxis der Bewertung* legt.

3 Das Bild: Vom Wert zur Bewertung als Praxis

Auf der Grundlage des noch relativ jungen Forschungsfeldes der Bewertungsstudien (vgl. auch die Zeitschrift *Valuation Studies*) schlagen wir eine analytische Verschiebung vor, die sich auf Bewertungsprozesse und -praktiken konzentriert anstatt auf „Wert als etwas an sich“ (Muniesa 2011, S. 25; Adkins und Lury 2011; Beckert und Aspers 2011; Berthoin Antal et al. 2015; Heintz 2019; Kornberger et al. 2015; Krüger and Reinhart 2017). Obwohl John Dewey üblicherweise als Gründervater dieser Perspektive gilt, ist Friedrich Nietzsches Frage über den „Wert von Werten“ für den analytischen Fokus der Bewertungsstudien mindestens ebenso wichtig. In seinem genealogischen Projekt fasste Nietzsche Werte nicht als gegeben auf; stattdessen griff er auf die Geschichte zurück, um zu sehen, wann und wie ein spezifischer Wert begann, bestehende Werte zu unterlaufen und schließlich dominant zu werden. „Blut und Grausen ist auf dem Grunde aller ‚guten Dinge‘“, meinte er in seiner *Genealogie der Moral* (Nietzsche 2014 [1887]).

Dewey (1939) erörterte die Frage der Werte etwas systematischer. Für ihn liegt die Hauptschwierigkeit in der bereits angedeuteten ontologischen Ambiguität von Werten – als individuelle „emotionale Epitheta“ wie auch als rationale, geteilte Prinzipien, auf welchen das soziale Leben und seine Institutionen beruhen sollten. Dabei argumentierte er, dass eine Bewertung sowohl durch „prizing“ (im Sinne von „für kostbar halten“) als auch durch „appraising“ (im Sinne von „Wert zuweisen“) erfolgen kann. Letzteres ist dabei, so Dewey, eine Aktivität der Bewertung, ein Akt,

der einen Vergleich erfordert, wie z. B. bei der monetären Bewertung von Gütern und Leistungen (Dewey 1939, S. 5). Gemäß Fabian Muniesa (2011) kann Wert als etwas verstanden werden, „that something has by virtue of how people consider it (how they personally like it, in particular), but also as something that something has as a result of its own condition and of its relation to other things (for instance, in relation to work or to money, or to any sort of standard metric)“ (Muniesa 2011, S. 26). In diesem Kontext bezieht sich die Bewertung auf „something that happens to something, and this happening can be a matter of consideration or of relation, or of both at the same time“ (Muniesa 2011, S. 26). Die Frage, ob Wert objektiv oder subjektiv ist, ist für die am Pragmatismus orientierten Bewertungsstudien sekundär. Zentral ist, *wie* Wert(e) in der Praxis durch Bewertungen („appraising“), Wertschätzungen („prizing“) oder Bepreisung („pricing“) dargestellt, aktiviert und ausgeführt, herausgefordert und verändert werden. Diese Aktivitäten können zwar analytisch unterschieden werden, sie sind in der Praxis jedoch oft miteinander verwoben.

Die Untersuchung von Bewertungspraktiken verschiebt den Fokus auf den „Abwägungsprozess“ und die Bedingungen, die einen Vergleich und eine Beurteilung ermöglichen (Muniesa 2011; Stark 2009): Wenn jemand x für wertvoller hält als y , ist die analytische Aufgabe weder herauszufinden, zu welchen Wertsphären x und y gehören, noch deren Nutzen einzuschätzen und sie entsprechend zu hierarchisieren. Stattdessen liegt die Aufgabe darin, das Instrumentarium zu untersuchen, das Vergleich, Klassifikation und Messung von x und y überhaupt erst möglich macht. Marion Fourcade (2011) argumentiert in ihrer vergleichenden Studie zweier Ölkatastrophen in Frankreich und den USA, dass sich Bewertungspraktiken nur verstehen lassen, wenn man danach fragt, über welche konkreten Mechanismen Werte kalkuliert und miteinander verglichen werden und inwiefern Beurteilungen dadurch erleichtert werden. So beschädigten die Ölkatastrophen in Frankreich und den USA zwar die „gleiche“ Umwelt. Aber der Wert dieser Umwelt wurde in Frankreich und den USA sehr unterschiedlich konstruiert. Eine solche Fokussierung auf die Dynamik des Konstruktionsprozesses bringt mehr Einsichten zutage als eine statische Analyse, die a priori zwischen ökonomischem Wert und sozialen Werten unterscheidet. Dies ist die Hauptthese der Bewertungsstudien. Das bedeutet nicht, dass „Wertordnungen“ (oder Wertsphären) keine Rolle mehr spielen. Sie werden von den Bewertungsstudien nur anders interpretiert (siehe z. B. Boltanski und Thévenot 2006; Stark 2009). Wie Stark (2009, S. 22) schreibt: „For my French conventionalist colleagues [...] orders of worth are not values counterpoised to value but are constitutive of value. Orders of worth are the very fabric of calculation, of rationality, of value“. Wertordnungen lassen sich entsprechend nicht auf getrennte Bereiche abbilden. Vielmehr beziehen sie sich auf unterschiedliche Bewertungsprinzipien, die in (unklaren) Situationen aufgerufen werden.

Ein Fokus auf Bewertungspraktiken impliziert damit analytisch, dass man das Augenmerk auf die *Praktiken* der Kommensurierung, Klassifikation und Kategorisierung und auf die verschiedenen Instrumente legt, die Vergleiche und Einschätzungen ermöglichen. Solche Instrumente müssen nicht notwendigerweise quantifizierend sein. Während Espeland und Stevens (1998, S. 313) Kommensurierung als „Vergleich verschiedener Entitäten anhand einer gemeinsamen Metrik“ definieren, verstehen wir den Begriff breiter: als Praktiken, die darauf abzielen, Dinge

und Aktivitäten durch Quantifizierung und andere Formen der Kategorisierung, der Gewichtung und der Ordnung vergleichbar zu machen (vgl. auch Heintz 2016).

Ferner ist zu beachten, dass Kommensurierung die Welt tendenziell „glättet“ – insbesondere wenn sie durch Quantifizierung erreicht wird. Rankings glätten die Welt, indem sie wichtige Unterschiede unberücksichtigt lassen und ihre Zielobjekte stattdessen über ein homogenes Medium, ein gemeinsames Maß erfassen (Strathern 2000; vgl. auch Espeland und Sauder 2007). Publikationen in A-Zeitschriften, Kennzahlen, die die Einwerbung von Drittmitteln wiedergeben, Gehälter von Studienabgängerinnen usw. sind Maße, die ein spezifisches globales Ideal einer „guten“ Business oder Law School favorisieren. Ein Ideal, das den unterschiedlichen lokalen Realitäten von Schools, die eng mit der lokalen Bevölkerung verknüpft sind, auf Nischengebiete spezialisiert sind, keine großen Spenden anlocken, nicht bloß Elitestudenten eine juristische Ausbildung ermöglichen oder in Bereichen forschen, die sich an der Peripherie des akademischen Diskurses der A-Zeitschriften bewegen, konträr gegenübersteht.

Die aufpolierten, in Zeitungen und auf Webseiten veröffentlichten Rankings machen solche Unterschiede unsichtbar. Stattdessen führen sie uns eine übersichtliche, klare globale Ordnung vor Augen, die Pluralität in eine Hierarchie der Würdigen und weniger Würdigen überführt. Indem Rankings definieren, was zählt und was nicht zählt, tun sie das, was Bowker und Star (1999) als „sorting things out“ bezeichnet haben, ein Doppelspiel, das beleuchtet und gleichzeitig ausblendet (Espeland und Lom 2015). Das Begriffspaar „Beleuchten“ und „Ausblenden“ verrät dabei die ästhetische Dimension dieser Praxis. Denn Rankings und andere Bewertungen nehmen in Ligentabellen, Listen, Sternebewertungen, Akkreditierungssymbolen usw. eine konkrete, materielle Form an. Diese ästhetischen Qualitäten stellen sicher, dass sich Bewertungen einfach und schnell verbreiten und an sie in verschiedenen Bereichen angeschlossen werden kann. Dadurch werden sie in großem Stil geteilt und avancieren zu einem mächtigen Instrument für Beobachtungen zweiter Ordnung (Esposito und Stark 2019).

Die glättende Arbeit der Kommensurierung bringt in einem zweiten Schritt neue Kategorisierungen mit sich. Dies geschieht dadurch, dass innerhalb einer Kategorie die Ähnlichkeiten betont („lumping“), während zwischen Kategorien die Unterschiede hervorgehoben werden („splitting“) (Zerubavel 1996). So definiert in der Welt der Wirtschaftshochschulen die von der *Financial Times* veröffentlichte Top 45-Liste der wichtigsten wirtschaftswissenschaftlichen Zeitschriften, was unter Exzellenz verstanden wird. Auch die Top 10 der zu besuchenden Städte, die besten 50 Restaurants, die populärsten Orte auf TripAdvisor usw. verweisen darauf, dass Bewertungspraktiken neue Kategorien entstehen lassen (Kornberger und Carter 2010). Rankings, so Wendy Espeland (2001), erzeugen neue Objekte, neue Kategorien und neue Beziehungen zwischen Dingen und Menschen (Espeland 2001, S. 84).

Dies hat bedeutsame Auswirkungen: Kategorien definieren Mitgliedschaft und dadurch Identität, sie schließen ein und aus und sie verleihen Legitimität (Lamont 2012). Zudem bietet Kategorisierung einen interpretativen Rahmen (man denke etwa an Kategorien wie „Immigrantin“, „Geflüchteter“ oder „Asylsuchende“), der zur Etablierung kognitiver Schemata beitragen kann, die unsere Aufmerksamkeit steuern und unsere Entscheidungen beeinflussen. In diesem Sinne können Rankings als

Medien verstanden werden, die zwischen Subjekt und Welt vermitteln und unser Denken strukturieren (Douglas 1986). Dabei ist wichtig, dass solche Rahmen zum Zeitpunkt ihrer Konstruktion durchaus umstritten sein können. So schreibt Kornberger (2017, S. 1755) das Thomas-Theorem paraphrasierend: „Perceived as real, the league tables, ratings and indexes that come into existence as a corollary of valuation practices are real in their consequences“. Mit anderen Worten: Obwohl die meisten Forscherinnen Zeitschriftenrankings und das daraus resultierende „Veröffentlichungsspiel“ kritisch betrachten, spielen sie dennoch – vielleicht widerwillig – bei diesem Spiel mit.

Zusammenfassend lässt sich sagen, dass das Gebiet der Bewertungsstudien einen analytischen Ansatz vorschlägt, der zu rekonstruieren versucht, wie solche Spiele tatsächlich gespielt werden und was deren intendierte und nichtintendierte Folgen sind. Eine traditionellere sozialwissenschaftliche Perspektive endet bei der Diagnose, dass die Forschungswelt von der Marktlogik „kolonialisiert“ wird. Und eine ökonomische Analyse würde vermutlich untersuchen, auf welche Weise Nutzen, Motivation und Anreize das individuelle Verhalten steuern. Die Bewertungsstudien unterscheiden sich davon, indem sie die Entstehung, aber auch das Verschwinden der Werte, die wir als gegeben annehmen, kritisch unter die Lupe nehmen. Entsprechend schlagen sie ein alternatives analytisches Vokabular vor – nicht eines, das die bestehenden Begriffe ersetzt, aber eines, mit dem sich diese kritisch reflektieren und hinsichtlich ihrer Wahrheitsansprüche beurteilen lassen. Der springende Punkt ist nicht, dass Wert subjektiv ist, sondern dass er praktisch ist, wie Dewey (1915, S. 516) schreibt (zit. in Muniesa 2011, S. 25). Wert ist nicht etwas, das jemand oder etwas einfach so hat – „valuation is about considering a reality while provoking it“ (Muniesa 2011, S. 32).

Nun stellt sich die Frage, auf welche Weise Bewertungspraktiken durch die neuen, digitalen Technologien sowie Kategorisierungs- und Klassifikationsinstrumente herausgefordert und verändert werden. Wie sollen wir Bewertungspraktiken im Kontext von Digitalisierung, Big Data und maschinellem Lernen untersuchen? Im Folgenden breiten wir einige Begriffe und Vorschläge aus, die einen Rahmen für die Untersuchung von Bewertungspraktiken im Kontext von algorithmischer Governance und Plattformorganisation zu entwickeln helfen sollen.

4 Neue Perspektiven: Von Bewertungsinstrumenten zu evaluativen Infrastrukturen

4.1 Bewertung auf Plattformen

Immer mehr alltägliche Aktivitäten werden über digitale Plattformen organisiert (Srnicek 2017). Digitale Plattformen wie eBay, Deliveroo, Lieferando, Amazon Marketplace, Airbnb oder Uber verbinden Käuferinnen und Anbieter von Dienstleistungen, Unterkünften oder anderen Gütern. Solche Plattformen werden in der Betriebswirtschaftslehre oft als „zweiseitige“ oder „vielseitige“ Märkte bezeichnet, weil sie den Austausch zwischen unterschiedlichen Konsumententypen ermöglichen, die sonst kein Geschäft abschließen könnten (Yoffie et al. im Druck). So funktioniert z. B.

das Carsharing-Unternehmen Uber als Unterstützungsplattform, um Menschen, die Autos besitzen, mit Menschen, die Mobilität benötigen, zu verbinden. Airbnb ist ein US-amerikanisches Unternehmen für die Online-Vermietung von Ferienunterkünften, das Menschen hilft, Unterkünfte auf der ganzen Welt anzubieten und zu buchen. Und eBay ist eine Online-Auktionsplattform, auf der Menschen und Unternehmen weltweit eine Vielfalt von Waren und Dienstleistungen kaufen und verkaufen können.

Typischerweise besitzen Plattformorganisationen wie Airbnb, Uber oder eBay die Güter, die den Kern ihres Geschäfts ausmachen – Hotels oder Taxis –, nicht. Stattdessen organisieren sie mittels digitaler Technologien den Zugang zu diesen Gütern. Dabei vernetzen sie weltweit verstreute Nutzerinnen, die nicht vertraglich aneinander gebunden sind. Die Wertschöpfung geschieht nicht in einem klassischen Arbeitgeber-Arbeitnehmer-Verhältnis. Stattdessen hängt sie davon ab, wie und wie gut die Nutzer die durch die Plattform organisierten Tätigkeiten ausführen – sei es Taxifahren, Gäste bewirten oder Pizzas backen (Leoni und Parker 2019). Der Plattformbesitzer koordiniert und kontrolliert diese Aktivitäten, ohne direkt an deren Herstellung beteiligt zu sein. Koordination und Kontrolle werden nicht durch traditionelle Mittel organisatorischer Kontrolle wie Mitgliedschaft und Hierarchie erreicht. An ihre Stelle treten digitale evaluative Infrastrukturen (Kornberger et al. 2017), die das Rückgrat solcher Plattformorganisationen bilden.

Solche digitalen evaluativen Infrastrukturen basieren auf Praktiken des Rechnens, der Kalkulation und Bewertung, die eine plattformbasierte Organisation erst ermöglichen (Kornberger et al. 2017). Sie sind insofern „evaluativ“, als sie auf Rankings, Ratings, Reviews, Audits und anderen Formen einer oft quantitativen Beurteilung aufbauen. Diese Instrumente dienen dazu, die auf der Plattform ausgetauschten Produkte und Dienstleistungen wie auch die involvierten Nutzer zu bewerten (Lamont 2012). So zählt z. B. Uber die Anzahl der Fahrten eines Fahrers, die Stunden, während denen er online war, die Fahrpreise pro Stunde, die Anzahl angenommener Fahrten sowie die Ratings durch die Passagierinnen (Rosenblatt und Stark 2016). Diese Metriken werden in einem wöchentlichen „Leistungsbericht“ zusammengefasst, der jeder Fahrerin zugeschickt wird. Für die Ratings der Fahrer sind die Passagiere gebeten, jede Taxifahrt auf einer Skala von null bis fünf Sternen zu bewerten. Gemäß Rosenblatt und Stark (2016) müssen die Fahrerinnen mindestens einen Durchschnittswert von 4,6 aufweisen, um die App weiter zu benutzen, wobei dieser Wert je nach Stadt variieren kann. Falls sie bei den letzten 25 oder 50 Fahrten nur ein tiefes Rating erreicht haben, erhalten sie unter Umständen eine Deaktivierungsmeldung.

Bei Airbnb werden die Gäste und Gastgeber dazu angehalten, sich gegenseitig zu beurteilen und zu bewerten (Leoni and Parker 2019). Gäste können das allgemeine Erlebnis, die Sauberkeit der gemieteten Räumlichkeiten, die Genauigkeit der Angaben in der Anzeige, das Preis-Leistungs-Verhältnis, den Austausch mit der Gastgeberin vor und während des Aufenthalts, das Check-in-Prozedere, den Ort sowie die Verfügbarkeit und Qualität der Ausstattung bewerten.¹ Ähnlich wie bei Uber

¹ <https://www.airbnb.ch/help/article/1257/wie-funktionieren-die-sternebewertungen-f%C3%BCr-auf-enthalte> (29.01.2021).

werden die Ratings auf einer Likert-Skala von null bis fünf Sternen angegeben. Die Gastgeber können bei allen Kategorien sehen, wie oft sie fünf Sterne erhalten haben und wie Gäste andere Gastgeber in der Nähe bewertet haben. Sie erhalten darüber hinaus auch Tipps von der Plattform, wie sie ihre Anzeige verbessern können. Eine Gastgeberin muss von mindestens drei Gästen bewertet worden sein, bevor eine aggregierte Punktzahl erscheint.²

Zum Wohle der Gemeinschaft („for the good of the community“) sollen die Gastgeber auch ihre Gäste beurteilen. Solche Reviews bestehen aus schriftlichen Kommentaren mit bis zu 1000 Wörtern sowie Sternebewertungen für Sauberkeit (Haben die Gäste die gemieteten Räumlichkeiten sauber zurückgelassen?), Kommunikation (Wie klar haben die Gäste ihre Pläne, Fragen und Bedenken kommuniziert?) und die Befolgung der Hausregeln.³ Airbnb behält sich dabei vor, die Accounts von Gastgeberinnen temporär zu deaktivieren, wenn ihre Leistung von den Gästen schlecht bewertet wurde oder sie viele Stornierungen hatten. Airbnb kann die Accounts (von Gastgebern oder Gästen) auch permanent deaktivieren, falls deren Verhalten z. B. als diskriminierend beurteilt wurde (Leoni und Parker 2019).

Die Gastgeberinnen werden auch von Airbnb selbst eingestuft, indem ihnen detaillierte Statistiken zur Verfügung gestellt werden, die ihre Leistung mit der durchschnittlichen Leistung aller anderen Gastgeber und sogenannter „Superhosts“ vergleichen. Superhosts sind Gastgeber, die bestimmte, vom Plattformbesitzer festgelegte, Leistungskriterien erfüllt haben. Dazu gehören eine Antwortrate von 90 % oder mehr, eine Stornierungsrate von 1 % oder weniger und eine Gesamtbewertung von 4,8 Sternen – jeweils gemessen über die letzten zwölf Monate.⁴

Diese Plattformratings und Rankings funktionieren ähnlich wie nichtdigitale Ratings und Rankings (Brankovic et al. 2018; Espeland und Sauder 2007; Esposito und Stark 2019; Mehrpouya und Samiolo 2016; Pollock und D’Adderio 2012). Die Kundenbewertungen von Uber und Airbnb erleichtern z. B. die Vergleichbarmachung und das Vergleichen (Espeland und Stevens 1998; Heintz 2010). Sie quantifizieren und standardisieren die Leistung von Fahrerinnen und Gastgebern auf der Basis einer gemeinsamen Metrik (in diesem Fall einer fünfstufigen Likert-Skala). Dadurch machen sie die Erfahrungen der Kundinnen vergleichbar und die Leistung von Fahrern und Gastgeberinnen bewertbar und kontrollierbar. Die Ratings fungieren auch als wichtige „Statusmarkierungen“ (Espeland und Sauder 2016), denn sie ermöglichen die Identifikation von „Superhosts“ und „Top-Fahrern“.

Solche Benchmarking-Techniken spornen Menschen zu immer mehr Leistung an und führen zu dem, was Espeland und Sauder (2007) „Reaktivität“ genannt haben. Rosenblatt und Stark (2016) zeigen, dass einige Uber-Fahrerinnen ihre Passagiere über die Funktionsweise der Ratings aufklären und ihnen dabei klarzumachen versuchen, dass eine 4 eine ungenügende Note ist, die zu ihrer „Deaktivierung“ führen kann. Die Ratings beeinflussen auch den Umgang mit den Passagieren. Einige Fahrer bieten Wasser oder Smartphone-Ladegeräte an, um ihre Bewertung zu verbessern

² <https://www.airbnb.ch/help/article/1257/wie-funktionieren-die-sternebewertungen-f%C3%BCr-auf-enthalte> (29.01.2021).

³ <https://www.smartbnb.io/airbnb-host-reviews-how-to-automate-them/> (29.01.2021).

⁴ <https://www.airbnb.ch/help/article/829/wie-werde-ich-ein-superhost> (29.01.2021).

(Rosenblat und Stark 2016). Wie Rosenblat und Stark (2016, S. 3772) schreiben: „To achieve good ratings, drivers must modify their behavior to produce a homogenous Uber experience for riders“.

Ratings fungieren zudem als ein wichtiges „vertrauensbildendes Instrument“ (Jeacle und Carter 2011) in der Beziehung zwischen Käufer und Verkäuferinnen. In Abwesenheit anderer Kontrollmechanismen verhilft das Ratingsystem dazu, Verkäufer und Käuferin, Fahrerinnen und Passagiere, Gastgeber und Gäste wechselseitig rechenschaftspflichtig zu machen (Espeland und Sauder 2016; Rosenblat und Stark 2016; Orlikowski und Scott 2014). Weil die Nutzer räumlich getrennt sind und sich untereinander nicht kennen, sind Plattformorganisationen wie Uber und Airbnb auf Ratings und Reviews angewiesen, damit Vertrauen und Reputation im Marktplatz entstehen kann. Um Esposito und Stark (2019, S. 3) zu paraphrasieren: Ratings liefern einen hochgradig sichtbaren Bezugspunkt, an dem sich andere – ob Fahrerinnen, Passagiere, Gastgeberinnen oder Gäste – orientieren können. Wenn keine anderen qualitätssichernden Mechanismen wie Lizenzen oder formelle Zertifikate zur Verfügung stehen, um Unsicherheit zu reduzieren (z. B. hinsichtlich der Qualität des Fahrers oder der angebotenen Unterkunft), stellen Ratings eine wichtige Orientierungshilfe dar. Doch es gibt auch Unterschiede zu etablierten Rankings und Ratings, wie z. B. Universitätsrankings (Espeland und Sauder 2016; Brankovic et al. 2018) oder Bonitätsratings (Carruthers 2013).

Plattformratings dienen nicht bloß als wichtige Instrumente zur Bewertung und Kontrolle. Sie haben auch eine koordinierende Funktion, denn sie sind das Grundelement, das Käuferinnen und Verkäufer zusammenführt und bilden damit den eigentlichen Kern der Plattformorganisation. Entsprechend sollten sie auch nicht als einzelne Instrumente analysiert werden, deren Hauptanliegen darin besteht, einen möglichst genauen Bezug zwischen Objekten und ihrer Repräsentation herzustellen. Stattdessen, und das ist ein entscheidendes Moment digitaler evaluativer Infrastrukturen, verhelfen solche Bewertungen dazu, Beziehungen zwischen Dingen, Menschen und Ideen zu *generieren*, seien das Autos, Fahrerinnen, Passagiere oder Gastgeber, Wohneigentümer oder Mieter. Sie schaffen *Konnektivität* („connectivity“) und tragen damit dazu bei, genau jene Strukturen zu erzeugen, aus denen Plattformen bestehen. Dadurch unterscheiden sie sich von Universitäts- und anderen Rankings und Ratings, die normalerweise funktional äquivalent zu anderen Instrumenten der formalen Organisation und Kontrolle sind, wie etwa Verträge, formale Hierarchien und klare Definitionen von Mitgliedschaft.

Der Begriff der evaluativen Infrastruktur lenkt den Blick auf zwei Eigenschaften von Rankings und Ratings: ihre „infrastrukturellen“ und ihre „organisationsgenerierenden“ Qualitäten. Er verschiebt die Aufmerksamkeit von individuellen Bewertungsinstrumenten zu *Bewertungssystemen*, die ein Netzwerk von organisierten Transaktionen erschaffen, und er unterstreicht die Beziehung zwischen Bewertungsprozessen und deren Fähigkeit, Vernetzungen zu ermöglichen. Der Begriff der evaluativen Infrastruktur verdeutlicht auch, dass Bewertungsprozesse von einer Vielzahl verschiedener Elemente, „things, ideas and marks“, abhängig sind, wie Hacking (1992) es formulierte. Nimmt man nur die einzelnen Bewertungsinstrumente in den Blick, besteht die Gefahr, sie statisch zu begreifen. Im Gegensatz dazu macht der Begriff der Infrastruktur deutlich, dass es um Relationalität und das Zusammen-

spiel verschiedener Instrumente geht (Bowker et al. 2019). Wie Star und Ruhleder (2001) argumentieren, ist Infrastruktur im Kern eine Relation und niemals ein bloßes Objekt. Der Fokus auf Infrastrukturen verlagert den Schwerpunkt von Dingen und Menschen als Hauptursache für Veränderung auf die durch Infrastrukturen erzeugten *Beziehungen* – z. B. zwischen Messung, Klassifikation, Personen, Ideen, Informationstechnologien und Managementstrukturen.⁵

Bei eBay, Airbnb und Uber sind die Kundenbewertungen in eine größere evaluative Infrastruktur eingebettet. Diese organisiert die Plattform, die sich aus Ratings, Visualisierungen von Ratings, Feedbackschleifen, Empfehlungssystemen und Trackingtechniken zusammensetzt (Kornberger et al. 2019; Power 2019). Bereits Michel Foucault schlug vor, nicht bei den einzelnen Objekten (in unserem Fall Ratings, Rankings und anderen Bewertungsinstrumente) anzusetzen, sondern die Praktiken zu untersuchen, die solche Objekte gestalten und formen (Mennicken und Miller 2014, S. 19). Diese Perspektive stellt die Relationen in den analytischen Mittelpunkt (Veyne 1978, S. 236). Wie Mennicken und Miller (2014, S. 19) es formulieren: „This means attending not only to the devices that instrumentalize the real, but also analysing their interdependence with the multiple rationalities and codes that seek to prescribe how the real is to be programmed“. In unserem Fall heißt das, dass wir die Auswirkungen individueller digitaler Bewertungsinstrumente – wie z. B. die Kundenbewertungen von Airbnb oder Uber – allgemeiner aus der Perspektive von *Systemen* digitaler Informationstechnologie, Ideen und Programmen betrachten müssen. Digitale Bewertungsinstrumente sind Teil eines größeren Geflechts, welches sich aus Informationstechnologien sowie Ideen der „Sharing-Ökonomie“ und der algorithmischen Governance zusammensetzt.

Ein weiteres charakteristisches Merkmal digitaler evaluativer Infrastrukturen, das sie von traditionellen Ratings und Rankings unterscheidet, ist ihre Einbettung in Praktiken algorithmischer Governance. Evaluative Infrastrukturen sind auf automatisierte Systeme des Sammelns und der Analyse großer Datenmengen angewiesen – und füttern sie zugleich. So zeichnet Uber z. B. die genaue Fahrstrecke einer Fahrer:in auf und kann sie zur Einhaltung der effizientesten Route verpflichten (Rosenblat und Stark 2016). Die Plattform speichert auch Informationen über den verrechneten Preis jeder Fahrt, wann und wie oft die Fahrer eingeloggt sind, wie schnell diese auf Fahrtenanfragen reagieren usw. Wie Rosenblat und Stark (2016) betonen, fließen die Daten von Fahrerinnen und Passagieren über zahlreiche Kanäle zur Plattform, was neue Formen der Überwachung, Preisgestaltung und Wertschöpfung ermöglicht.

⁵ Der vorliegende Beitrag konzentriert sich auf digitale evaluative Infrastrukturen als Spezialfall evaluativer Infrastrukturen. Grundsätzlich bezieht sich der Begriff der evaluativen Infrastruktur sowohl auf digitale als auch auf nichtdigitale Bewertungspraktiken und -technologien. Auch nichtdigitale Ranking- und Ratingtechnologien, wie z. B. frühe Formen des Bonitätsratings (Carruthers 2013), sind auf die Vernetzung und ein Zusammenspiel verschiedener Instrumente angewiesen. Um die Operationsweise dieser Ratings zu verstehen, ist ein Blick auf die Beziehungen und Dynamiken, die sich zwischen Messung, Klassifikation, Menschen, Ideen, Technologie und Managementstrukturen entfalten, wesentlich. Die hier untersuchten digitalen evaluativen Infrastrukturen unterscheiden sich jedoch von nichtdigitalen evaluativen Infrastrukturen dadurch, dass sie sich auf Formen und Praktiken der algorithmischen Kategorisierung und Automatisierung beziehen, auf die wir im Folgenden näher eingehen. Darüber hinaus unterscheiden sie sich auch hinsichtlich ihres Umfangs und ihrer Reichweite.

Bei Uber werden die Preise algorithmisch festgelegt.⁶ Dank einer algorithmischen Auswertung von Angebot und Nachfrage kann das System die Fahrpreise für gewisse Gebiete für eine bestimmte Zeit erhöhen. Diese Praxis nennt sich „surge pricing“⁷ und sorgt dafür, dass sich bei hoher Nachfrage mehr Fahrerinnen auf den Weg machen, um diese Nachfrage zu bedienen. „Surge Pricing“ ist sowohl für die Fahrer als auch für die Passagiere erkenntlich; die Passagiere können wählen, ob sie den höheren Preis zahlen oder lieber auf eine Preissenkung in ein paar Minuten warten möchten. Auch bei Airbnb können sich Gastgeber für „smart pricing“ anmelden. Diese algorithmische Preisgestaltung erlaubt es ihnen, ihre Preise basierend auf den Nachfrageänderungen für ähnliche Angebote automatisch nach oben oder nach unten zu korrigieren.⁸ Auf diese Weise aktivieren und formen digitale evaluative Infrastrukturen die Praktiken (automatisierter) Markttransaktionen.

Auch viele Managementfunktionen sind automatisiert. Dies gilt für Deaktivierungsentscheidungen, das Setzen von Leistungsvorgaben oder für Tarifangebote (z. B. in der Form höherer Grundtarife für eine Fahrt). Digitale evaluative Infrastrukturen liefern ein skalierbares und weitgehend selbstständiges Instrument, um die Qualität einer fluktuierenden und weltweit verteilten Belegschaft zu sichern. In den nächsten beiden Abschnitten gehen wir zuerst auf die charakteristischen Machteffekte von digitalen evaluativen Infrastrukturen ein, um anschließend deren generatives Potenzial zu untersuchen. Damit unterscheidet sich unsere Zugangsweise von Ansätzen, die die Performativität und Reaktivität von Ratings und Rankings in den Mittelpunkt stellen.

4.2 Die Machteffekte von Protokollen

Unsere Analyse schlägt vor, genauer zu untersuchen, wie digitale evaluative Infrastrukturen die Modalitäten des Regierens sowie die Formen politischer Macht gestalten oder verändern (Deleuze 1992; Galloway 2004; Miller und Rose 1990, 2008). Digitale evaluative Infrastrukturen machen neue Formen der Intervention und Steuerung möglich (Bowker et al. 2019). Sie erlauben, eine Vielzahl unterschiedlicher Akteure und Einheiten miteinander zu verbinden und transformieren diese und die sie umgebende Welt. Sie beeinflussen die Fähigkeiten von Akteuren und Organisationen und die Beziehungen zwischen ihnen. Sie unterstützen neue Formen verteilter Intelligenz und Kontrolle, die unsere Aufmerksamkeit, unsere Entscheidungen und unsere Subjektivität prägen. Sie ermöglichen neue Formen, auf die Handlungen von Individuen einzuwirken, und sie etablieren eine bestimmte Auffassung der zu steuernden Objekte und Ziele.

In Plattformorganisationen ist Macht zugleich verteilt und konzentriert. Einerseits ist algorithmisches Management auf das Bewertungsfeedback weltweit verteilter

⁶ Uber legt Mindesttarife fest, die von den Fahrern verrechnet werden müssen. Zudem riskieren die Fahrerinnen eine Deaktivierung, falls sie unrentable Fahrten ablehnen. Die Fahrer dürfen zwar einen niedrigeren Fahrpreis aushandeln, aber keinen höheren (Rosenblatt und Stark 2016).

⁷ <https://www.uber.com/us/en/drive/driver-app/how-surge-works/> (29.01.2021).

⁸ <https://www.airbnb.ch/help/article/1168/wie-kann-ich-die-intelligente-preisgestaltung-aktivieren-oder-deaktivieren> (29.01.2021).

Nutzerinnen, Käufer und Verkäuferinnen angewiesen. In dieser Hinsicht ist Kontrolle hochgradig verteilt (Kornberger et al. 2017). Airbnb, eBay und Uber stützen sich auf die wechselseitigen Bewertungen der Nutzer und nutzen sie als „Prüfer“ der jeweils anderen (Jeacle und Carter 2011). Andererseits werden die Spielregeln jedoch von den Plattformbesitzern festgelegt. Wie Leoni und Parker (2019) in Bezug auf Airbnb zeigen, haben die Gastgeberinnen und Gäste zwar die Macht, Ratingentscheidungen zu fällen, aber es sind die Plattformbesitzer, die den Rahmen vorgeben. Sie entwerfen die evaluative Infrastruktur und entscheiden darüber, welche Informationen wo und wann gesammelt werden. Zudem haben die Plattformbesitzer die Macht, den Zugang zu den von ihnen erschaffenen Transaktionsnetzwerken einzuschränken und zu regulieren. So hat z. B. Airbnb das Recht, einen Account einzuschränken, zu sperren oder zu deaktivieren. Uber kann ohne Vorwarnung den Account von Fahrern deaktivieren – ohne Gründe dafür angeben zu müssen. Und eBay kann einschränken, was auf der Plattform gehandelt werden darf und was nicht (Kornberger et al. 2017). All diese Plattformen besitzen die Daten, die sie dort sammeln, und können sie ohne Zustimmung derjenigen, die an ihrer Erzeugung beteiligt waren, weiterverwenden und neu kombinieren.

Der Erfolg und die Macht von Plattformorganisationen hängen von der Fähigkeit ab, wechselseitige Überwachungskreisläufe zu schaffen und aufrechtzuerhalten. Scott Galloway (2004) spricht in diesem Kontext mit Bezug auf Deleuze (1992) von „protokollarischer Kontrolle“ (s. a. Galloway und Thacker 2004). Der Begriff des Protokolls stammt ursprünglich aus der Informatik und wurde von Galloway (2004) und Galloway und Thacker (2004) ausgeweitet, um die spezifischen Machtverhältnisse in Netzwerken zu beschreiben und zu analysieren. Unter einem Netzwerk verstehen sie „any system of interrelationality, whether biological or informatic, organic or inorganic, technical or natural“ (Galloway und Thacker 2004, S. 8). Für unsere Analyse ist der Begriff des Protokolls – und die damit verbundene Aufmerksamkeit für dessen Machteffekte – nützlich. Denn er verhilft dazu, die widersprüchliche Natur des Machtapparats digitaler evaluativer Infrastrukturen zu erfassen. Wie Galloway und Thacker (2004) schreiben: „Protocological control brings into existence a certain contradiction, at once distributing agencies in a complex manner, while at the same time concentrating rigid forms of management and control“ (Galloway und Thacker 2004, S. 8).

Gemäß Galloway und Thacker (2004, S. 8) umfassen Protokolle alle konventionalisierten Regeln und Standards, die die Beziehungen innerhalb von Netzwerken regeln. Für sie ist ein Protokoll eine Technologie, die den Informationsfluss reguliert, das Netzwerk steuert, Beziehungen kodifiziert und Lebensformen verbindet (Galloway und Thacker 2004, S. 10). In unserem Fall sind Protokolle die Regeln und Standards, die digitalen evaluativen Infrastrukturen zugrunde liegen. Sie sind die Regeln, die darüber bestimmen, wie Daten formatiert, übertragen und empfangen werden können, sodass Menschen und Maschinen im Netzwerk miteinander kommunizieren können. Die primäre Funktion eines Protokolls besteht darin, die Informationsflüsse zu steuern und Konnektivität zu schaffen. Solche Informationsflüsse werden durch bestimmte Aktivitäten ausgelöst (z. B. durch die Bewertung einer Taxifahrt). Verschiedene Computerprotokolle sorgen dafür, dass dieser Informationsfluss

übertragen, zusammengeführt, aggregiert und für andere Plattformnutzer dargestellt wird.

Gemäß Galloway und Thacker (2004) fordert uns der Begriff der protokollari-schen Kontrolle dazu auf, kritisches und politisches Handeln neu zu denken. Wie sie schreiben: „Protological control challenges us to rethink critical and political action around a newer framework, that of multi-agent, individuated nodes in a metastable network“ (Galloway und Thacker 2004, S. 10). Entsprechend üben Protokolle ihre Macht nicht „von oben herab“ aus (Galloway und Thacker 2004, S. 20), obwohl Plattformorganisationen oftmals auch Elemente hierarchischer Organisation und traditioneller Methoden der Kontrolle enthalten (Gandini 2019; Leoni und Parker 2019; Rosenblat und Stark 2016). Es handelt sich folglich nicht um eine „anarchic liberation of data from below“ (Galloway und Thacker 2004, S. 20). Vielmehr erfordert der Begriff der „protokollarischen Macht“ ein komplexeres Verständnis der Machtbeziehungen in Plattformorganisationen. Er verhilft dazu, das ihnen eigene Wechselspiel zwischen hierarchischen und heterarchischen Machtverhältnissen, das digitale evaluative Infrastrukturen charakterisiert, zu hinterfragen und zu entschlüsseln. Zudem lenkt er unsere Aufmerksamkeit auf die Entstehung neuer Subjektivitäten und Kollektive. Auf der einen Seite üben Plattformen wie eBay, Uber oder Airbnb eine strikte Kontrolle über die Infrastrukturen und die durch sie generierten Daten aus. Auf der anderen Seite übernehmen sie aber keine Verantwortung für die Behauptungen, die auf ihren Plattformen gemacht werden. Die Netzwerke der Plattformen sind partizipativ, aber daraus sollte man nicht auf individuelle oder kollektive Handlungsmacht schließen. Nimmt man die Machteffekte von Protokollen in den Blick, wird deutlich, dass man zwischen verschiedenen Subjektivierungsformen unterscheiden und die unterschiedlichen Machteffekte von Messung, Klassifikation und Vergleich analytisch voneinander trennen muss.

Im Falle digitaler evaluativer Infrastrukturen liegt Macht eher in der Anzahl der Nutzer und im Umfang der Daten als in der Fähigkeit, Individuen zu disziplinieren und zu kontrollieren. Wie Galloway und Thacker (2004, S. 10) unter Bezugnahme auf Deleuze (1992) schreiben: „Protocol has less to do with individually empowered human subjects than with manifold modes of individuation that arrange and remix both human and nonhuman elements.“ Solche Individuierungsformen haben wenig mit individuellen menschlichen Subjekten zu tun. Vielmehr geht es um den Prozess, durch den individuell generierte Daten aggregiert, gepflegt, gefiltert, neu zusammengesetzt und in Umlauf gebracht werden. Anders ausgedrückt können wir auch sagen, dass individuelle Plattformnutzerinnen „dividuiert“ werden (Deleuze 1992). Digitale evaluative Infrastrukturen machen es möglich, Individuen – seien es Fahrerinnen oder Passagiere bei Uber, Gastgeberinnen oder Gäste bei Airbnb – in formbare Datenbits zu „dividuierten“, die dann neu bemessen, rekombiniert und in Populationen von „dividuals“ aggregiert werden können (Whitson 2014, S. 554). Dadurch wird Kollektivität aufgelöst. Digitale evaluative Infrastrukturen ermöglichen Formen der algorithmischen Personalisierung (z. B. im Bereich der personalisierten Preisgestaltung oder bei Produkten und Dienstleistungsangeboten). Eine solche Personalisierung erschwert es Plattformnutzern, sich als Angehörige einer anerkannten Gruppe zu identifizieren oder als solche identifiziert zu werden (vgl. auch Lury und Day 2019; Moor und Lury 2018). Ein Beispiel dafür ist die Berufsgruppe der Ta-

xifahrer. Damit tragen digitale evaluative Infrastrukturen zur Aushöhlung etablierter Arbeitsbeziehungen bei.

Digitale evaluative Infrastrukturen verwandeln Angestellte (etwa die Taxifahrerinnen bei Uber) in Plattformnutzer und Plattformperformer. Auf diese Weise tragen sie zur Zersplitterung der Arbeitnehmerschaft bei, die nicht mehr von einem formalen Management überwacht und kontrolliert wird, sondern von den Konsumenten selbst (Rosenblat und Stark 2016; Stark und Levy 2018). Anders als Angestellte erhalten Fahrerinnen oder Gastgeber auf Plattformen wie Uber oder Airbnb keine Sozialleistungen und keinen Mindestlohn und es sind auch keine Mindestarbeitszeiten garantiert. Dies erlaubt den Plattformbetreibern, die Lohnkosten tief zu halten (Yoffie et al. *im Druck*). Wie Rosenblat und Stark (2016, S. 3771) für Uber aufzeigen, gibt es kein formales Management, das die Arbeitsqualität der einzelnen Fahrerinnen kontrolliert. Stattdessen wird diese Aufgabe an die Passagiere übertragen, die via Fahrerrating eine Art Managementbeurteilung durchführen. Die Fahrerinnen können zwar Nachfragen stellen, z. B. wie die Fahrpreise berechnet werden, aber in der Regel erhalten sie auf solche Fragen nur Musterantworten. Sie haben auch keine Möglichkeit, die Arbeitsbedingungen auszuhandeln oder Ratings zu löschen, die sie als ungerecht oder als falsch beurteilen. Dass Macht von Arbeit auf Kapital übertragen wird (in diesem Fall von den Plattformarbeitern zu den Plattformbesitzern), ist an sich nicht neu (vgl. dazu die Labour-Process-Debatte der 1970er-Jahre; Gandini 2019; Vallas und Schor 2020). Doch die hier untersuchten digitalen evaluativen Infrastrukturen ermöglichen die Einrichtung neuer Kontrollsysteme. Dabei handelt es sich um Systeme digital vermittelter Kontrolle, die jenseits formaler Organisation operieren und dadurch neue Formen von Governance und des „governing at a distance“ (Miller und Rose 1990) erzeugen.

Die neuen Kontrollsysteme sind auf neue Formen der Überwachung angewiesen (Zuboff 2019). Diese neuen Überwachungsformen stützen sich nicht mehr auf direkte Beobachtung wie die des Panoptikums (Foucault 1977). Ihre Kernelemente sind stattdessen digital vermittelte Quantifizierung, Berechnung und Bewertung – kurz: digitale evaluative Infrastrukturen –, die den Aufbau von Kreisläufen gegenseitiger Überwachung erst ermöglichen. Kontrolle ist in diesen Kreisläufen der wechselseitigen Bewertung oft nicht direkt wahrnehmbar – sie funktioniert subtiler. De facto ist die Kontrolle zwar ausgelagert und dezentralisiert, doch es sind die Plattformbesitzer, die die von den Nutzern generierten Daten zentral speichern, bearbeiten, analysieren, verarbeiten oder weiterverkaufen. Digitale evaluative Infrastrukturen funktionieren nicht als eine „mechanische Kraft“ („mechanical force“), sondern als ein „alles durchdringendes Medium“ („pervasive medium“), um Cristina Alaimo und Jannis Kallinikos (2020) zu paraphrasieren (siehe auch Kellogg et al. 2020). Sie „verführen“ Nutzer mit digitalen Benachrichtigungen und Eingabeaufforderungen, wöchentlichen Leistungsberichten oder personalisierten Empfehlungen. Mit solchen „nudges“ werden Menschen „geködert“ (Seaver 2019) und dazu verleitet, die Plattformen regelmäßig und dauerhaft zu nutzen. Auf diese Weise werden Präferenzen (und Werte) nicht nur verstärkt. Sie werden zudem in neue Richtungen gelenkt und damit unter Umständen auch transformiert.

Digitale evaluative Infrastrukturen unterwerfen Plattformnutzerinnen einem neuen Typus von Kontrolle, der gamifizierte Praktiken verwendet. So nutzt Uber z. B.

spielbasierte „psychologische Tricks“, damit Fahrer öfter und länger fahren, die App nicht abschalten und zu persönlichen Bestleistungen angeregt werden (Gandini 2019, S. 1050). Im Zentrum solcher gamifizierter Praktiken stehen digitale evaluative Infrastrukturen, die dank sofortigen und ständig verfügbaren personalisierten Ratings und Rankings die Nutzer zu bestimmten Verhaltensweisen verführen. Stark und Levy (2018) zufolge erzeugen solche gamifizierten Überwachungsinfrastrukturen eine neue Subjektposition, die sie den „überwachenden Konsumenten“ („surveillant consumer“) nennen. Nutzerinnen von Plattformen sind nicht bloß als „überwachte Personen“ involviert, sie fungieren selbst als Überwacher. Dadurch lösen sich die Unterschiede zwischen den Rollen der „Überwachenden“ und der „Überwachten“ auf, und entsprechend sind auch theoretische Modelle, die nur binär zwischen Top-down- und Bottom-up-Kontrolle unterscheiden, nicht mehr angemessen. Evaluative Infrastrukturen reduzieren Unsicherheit, indem sie aus unbekanntem Nutzern engagierte und lesbare Konsumenten machen, die jederzeit „mehr und weniger als eins“ sind. Wie Lury und Day (2019) schreiben: „The you that is addressed is both specific and a you, that is like everyone else“, only more or less so“ (Lury und Day 2019, S. 29). Das angesprochene Du ist sowohl spezifisch, d.h. über die Aufzeichnung des Verhaltens, der Ratings, der Klicks und Likes von Konsumentinnen auf den Einzelnen zugeschnitten. Gleichzeitig ist es „ein Du wie jeder andere“ und lenkt damit die Aufmerksamkeit auf Ähnlichkeit („Personen wie Du interessierten sich auch für ...“).

4.3 Von Performativität zu Generativität

Evaluative Infrastrukturen zeichnen Gewohnheiten, Präferenzen und Meinungen auf, bündeln sie und setzen sie neu zusammen. Dadurch übernehmen sie eine Schlüsselrolle bei der Produktion neuer Formen von (dividierter) Subjektivität und (gamifizierter) Kontrolle, die mit neuen Möglichkeiten der Entdeckung und Einflussnahme einhergehen. Sie lassen neue Opportunitäten und Subjektivitäten entstehen und eröffnen neue Welten. Unseres Erachtens lässt sich diese Entwicklung mit dem Begriff der Generativität besser erfassen als mit dem der Performativität (MacKenzie und Millo 2003; MacKenzie et al. 2007) oder der Reaktivität (Espeland und Sauder 2007, 2016).

Espeland und Sauder (2007) haben das Konzept der Reaktivität eingeführt, um die reflexiven Wechselwirkungen zwischen Personen und Messinstrumenten zu analysieren. Sie haben untersucht, wie Rankings dazu führen, dass Dekane, Studentinnen und Administratoren ihr Verhalten in Reaktion auf sie verändern. Ihr Argument ist, dass „Reaktivität“ die Unterscheidung zwischen dem Messakt und seinem Gegenstand verwischt (vgl. auch Espeland und Sauder 2016). Auch beim Begriff der Performativität geht es um das Verhältnis zwischen (oftmals ökonomischen) Modellen und Messinstrumenten einerseits und der Welt, die sie zu „repräsentieren“ vorgeben, andererseits (Callon 1998; Gond et al. 2016; MacKenzie und Millo 2003; MacKenzie et al. 2007). Studien zur Performativität ökonomischer Modelle, z. B. dem ‚Modell von Black, Scholes und Merton‘ zur Bewertung von Optionen (MacKenzie und Millo 2003), belegen, dass diese nicht einfach messen (erfassen, berechnen), was vorhanden ist – oder was ein Zentrum als vorhanden unterstellt. Vielmehr bewegt sich die

Welt gerade umgekehrt auf das Modell zu. Wie MacKenzie und Millo (2003) in ihrer heute als Klassiker geltenden Studie zur Performativität der Finanzökonomie gezeigt haben, veränderten sich die Finanzmärkte allmählich so, dass sie immer besser mit dem Modell (in diesem Fall dem Modell der Optionsbewertung) übereinstimmten – und nicht umgekehrt. Allerdings, und das betonen auch MacKenzie et al. (2007), ist der Begriff der Performativität ein komplexer Begriff, der in seinen verschiedenen Spielarten betrachtet werden muss.

Die Begriffe der Reaktivität und der Performativität zeigen auf, dass und auf welche Weise Quantifizierung und Bewertung Menschen und Organisationen verändern. Sie verhelfen dazu, die reflexiven Wechselwirkungen zwischen Messinstrumenten (und Modellen) und den sie umgebenden Welten kritisch zu reflektieren. Trotzdem reichen sie unserer Meinung nach nicht aus, wenn es darum geht, die struktur-generierenden Eigenschaften von Rankings und Ratings und anderen Formen der Quantifizierung zu erfassen. Studien zur Reaktivität oder Performativität machen ökonomische Modelle, Rankings und Ratings sowie andere Formen der Quantifizierung zum Ausgangspunkt – und nicht zum Endpunkt – ihrer Analyse und lenken damit die Aufmerksamkeit auf die Frage, wie sie auf andere einwirken. Die Messinstrumente und Modelle sowie deren Relationalität bleiben jedoch im Wesentlichen im Dunkeln. Der Begriff der Generativität stellt dagegen die *emergenten* Eigenschaften der Quantifizierung innerhalb digitaler evaluativer Infrastrukturen ins Zentrum. Damit ist nicht impliziert, dass Rankings und Ratings auf Online-Plattformen nicht auch performativ oder reaktiv sind (siehe hierzu die Studie von Rosenblat und Stark 2016 oder den umfassenden Überblick von Kellogg et al. 2020). Aber sie sind und tun mehr als das. Im Folgenden befassen wir uns insbesondere mit drei Aspekten von Generativität: der Konnektivität, der algorithmisch erzeugten Kategorisierung und der Wertschöpfung, die sie ermöglichen.

Digitale Ratings und Rankings sind eine Schlüsselkomponente evaluativer Infrastrukturen, sie machen den Kern einer Plattformorganisation aus. Aus diesem Grund haben wir argumentiert, dass eine Plattformorganisation ohne digitale evaluative Infrastrukturen nicht funktionieren kann. Denn digitale evaluative Infrastrukturen, einschließlich ihrer Ratings und Rankings, haben die Fähigkeit, Plattformnutzerinnen, die andernfalls nicht in einen Austausch miteinander treten könnten, zu vernetzen. Indem sie eine Schnittstelle zur Verfügung stellen, die Interaktionen über große Distanzen hinweg ermöglicht, tragen sie maßgeblich dazu bei, Nutzer zu gewinnen und zu erhalten. Damit liegen sie im Kern der Entstehung neuer Märkte und Organisationsformen.

Das Vernetzungspotenzial von Online-Ratings und Rankings ist auf digitale Technologien, IT-Netzwerke, Datenhubs und globale mobile Vernetzung angewiesen. Plattformunternehmen machen sich die Verhaltensgewohnheiten von Milliarden von Nutzerinnen zunutze, die sich täglich über ihre digitalen Endgeräte mit den Plattformen verbinden, um ihre Dienstleistungen zu nutzen (Yoffie et al. im Druck). Der Begriff der digitalen evaluativen Infrastruktur verweist auf diese Konnektivität. Er legt uns nahe, individuelle digitale Bewertungsinstrumente, wie z. B. Ratings und Rankings, nicht isoliert zu betrachten, sondern sie zur umfassenden soziomateriellen Infrastruktur in Beziehung zu setzen, deren Teil sie sind (Kornberger et al. 2019; Orlikowski und Scott 2014; Star 1999). Dazu gehören Endgeräte und Vi-

sualisierungstechniken, aber auch die Strukturen und Praktiken, einschließlich des maschinellen Lernens, mit denen sie die Daten, die durch die Bewertungen auf den Plattformen entstehen, sammeln, speichern und sortieren. Digitale evaluative Infrastrukturen berechnen nicht so sehr, was da ist, vielmehr erschaffen sie neue Welten. Dies geschieht durch die Herstellung von Objekten, die nicht das Ergebnis programmatischer Ansprüche oder Modelle sind, sondern aus einem Überschuss an Daten und Spuren entstehen, die wiederum neue Möglichkeiten der Entdeckung und Einflussnahme hervorbringen.

Besonders aufschlussreich ist die spezifische Kombination von Quantifizierung und Bewertung auf der einen und digitalem Tracking und algorithmischer Kategorisierung auf der anderen Seite (vgl. auch Barnard und Fourcade 2021 in diesem Band). Indem sie die Spuren, die Nutzer hinterlassen, aufzeichnen, bringen digitale evaluative Infrastrukturen neue Möglichkeiten, Klassifikationen und Subjektivitäten hervor (Bowker et al. 2019; Power 2019). Evaluative Infrastrukturen, wie sie z. B. Uber, Airbnb oder eBay zugrunde liegen, verfolgen die Objekte (und Subjekte) anhand ihrer Klicks oder ihrer Suchanfragen und machen sie dadurch sichtbar und zugänglich für weitere Interventionen (Power 2019).

Ratings und Ranking auf Online-Plattformen beruhen einerseits auf konventionellen Formen der Quantifizierung. Solche Quantifizierungen ähneln der Messung von Kundenzufriedenheit im Offline-Bereich, z. B. über eine Likert-Skala, die von 1 bis 5 reicht. Ein Beispiel dafür sind die Kundenratings bei Uber. Andererseits werden solche traditionellen Ratings mit algorithmisch produzierten Ratings und Rankings kombiniert, die auf enormen Mengen an nutzergenerierten Daten beruhen, die mittels Cookies und anderer Techniken aufgezeichnet werden. Diese Art der (oft versteckten) Datenextraktion ist für das Geschäftsmodell der Plattformen zentral. Denn sie ist die entscheidende Ressource, mit der sich bestehende Dienste vermarkten oder neue Dienste entwickeln lassen. Uber sammelt z. B. Daten über alle Fahrer und alle Fahrten, auch wenn sie nicht bezahlt wurden. Neben den Bewertungen der Passagiere zeichnet Uber auch den Standort der Fahrerinnen auf, die Anzahl angenommener Fahraufträge, die jeweiligen Fahrpreise, die Online-Zeiten und vieles andere mehr. Aus diesen Daten lassen sich weitere Größen ableiten, z. B. die beste Routenwahl oder die Leistung eines Fahrers im Vergleich zu anderen. Erst auf dieser Basis wird eine gezielte algorithmische Preisgestaltung (z. B. bei hoher Nachfrage) überhaupt möglich. Das automatisierte Aufzeichnen von Daten ermöglicht ein sofortiges Feedback, personalisierte Leistungsmessungen und Benchmarking. Auf diese Weise kann das Verhalten in Echtzeit verfolgt und ausgewertet werden, z. B. die Routenwahl oder wie schnell ein Fahrer auf einen Anruf reagiert und wie lange er online ist.

Mittels algorithmischen Trackings lassen sich Daten auf neue Weise sammeln und verarbeiten, desaggregieren, neu zusammenfügen und bewerten. Wie Fourcade und Healy (2017) schreiben, entstehen dadurch neue „classification situations“, wobei die Beziehung zwischen Klassifikation und Messung auf den Kopf gestellt wird (Mennicken und Espeland 2019). Zu einem gewissen Grad nimmt damit die Bedeutung der Klassifikation für das Messen ab. Denn im Gegensatz zu den Überlegungen von Desrosières (1998) ist die Kategorisierung ein Ergebnis der Messung und geht ihr nicht voraus (vgl. dazu auch Heintz 2021 in diesem Band). Die Kategorisierung

erfolgt bottom-up und es entstehen neue Beziehungen (z.B. Konkurrenz), die auf den Nutzerbewertungen und der Aufzeichnung ihres Verhaltens („clicks“) beruhen (Jeacle und Carter 2011; Kellogg et al. 2020; Lury und Day 2019; Orlikowski und Scott 2014).

Die digitalen evaluativen Infrastrukturen von Uber, Airbnb, eBay und anderen Plattformen umfassen weitaus mehr als Kundenratings und -rankings. Sie bestehen auch aus Systemen algorithmischer Personalisierung und Empfehlung (Alaimo und Kallinikos 2020; Lury und Day 2019; sowie Unternährer in diesem Band), die die Ratings und Rankings flankieren und die anhand von vielfältigen Interaktionsdaten – z. B. Kundenratings und Daten zu Such- und Bestellungsverläufen – entwickelt werden. Der ökonomische Erfolg von Plattformen wie Airbnb und eBay hängt zu einem großen Teil von der Entwicklung solcher Empfehlungssysteme ab. Personalisierte Empfehlungen – z. B. eines auf eBay angepriesenen Produkts oder einer auf Airbnb angebotenen Unterkunft – beruhen auf Messung und Kategorisierung mittels maschinellem Lernen. Sie stützen sich nicht auf externe kategoriale Merkmale, wie z. B. Klassenzugehörigkeit, Wohnort, Alter oder Geschlecht. Vielmehr werden sie intern generiert, z. B. über Collaborative Filtering, bei dem riesige Mengen an digitalen Daten über das Verhalten, die Aktivitäten oder die Präferenzen der Nutzer verarbeitet werden (dazu auch Heintz 2021; sowie Unternährer 2021 in diesem Band). Diese Berechnungen bilden die Grundlage dafür, um auf der Basis der Ähnlichkeit mit anderen Nutzern einzuschätzen, was einer spezifischen Nutzerin gefallen könnte (Alaimo und Kallinikos 2020; Lury und Day 2019). Diese Ähnlichkeit wird nicht extern abgeleitet, sondern beruht auf der Sortierung und dem Clustern von Daten, die von der Plattform gesammelt wurden. Alaimo und Kallinikos (2020, S. 15) formulieren dies treffend: „The category of similarity is established by computing, not by experience, cognition or knowledge ... on the basis of formal operations rather than via resemblances or other intrinsic or practice-related affinities.“

Digitale evaluative Infrastrukturen funktionieren ausschließlich über interne Prozesse, z. B., indem sie Datenobjekte nach ihrer Ähnlichkeit in Gruppen sortieren, wobei sich Ähnlichkeit anhand des Nutzerverhaltens bemisst. So werden z. B. Louis Armstrong und Ella Fitzgerald vom System als ähnlich beurteilt, weil sie von demselben Nutzer immer wieder gehört werden. Dies haben Alaimo und Kallinikos (2020, S. 14) für den Online-Musikdienst Last.fm gezeigt. Entitäten oder Aktivitäten werden nicht über ein externes Maß, wie z. B. Geld, vergleichbar gemacht, sondern über datengestützte algorithmische Kategorisierung – dem Zählen und Clustern von Likes und der Berechnung der Ähnlichkeit zwischen Nutzerpräferenzen (z. B. auf der Basis von Pearson-Korrelationen zwischen Ratingwerten oder der Cosinus-Similarität von Ratingvektoren). In diesem Kontext ist Quantifizierung, etwa in Form des Unterkunftsratings durch einen Gast, der Ausgangspunkt der Bewertung und nicht ihr Endpunkt. Qualitäten werden nicht in Quantitäten verwandelt, wie es Espeland und Sauders (2007) Begriff der Kommensurierung nahelegt, sondern Quantitäten werden in neue Formen der Qualifizierung übersetzt, z. B. einer Qualifizierung von Plattformnutzerinnen.

Indem sie unablässig verschiedene Datenformen verknüpfen, vergleichen, de- und rekontextualisieren, erzeugen evaluative Infrastrukturen neue Muster, Profile und damit über kurz oder lang, neue Werte. Sie sind das Kernelement bei der Formierung

neuer Marktstrukturen (Fourcade und Healy 2017), in denen die subjektive Einheit des Individuums (in diesem Fall des Käufers und/oder der Verkäuferin) erodiert. Wie Moor und Lury (2018) zeigen, führen neue digitale Technologien zu einer Verschiebung der Beziehung zwischen Preis und Person. Ursprünglich generisch oder soziodemografisch definierte Personen werden in Datenpunkte desaggregiert, die anschließend mithilfe von Algorithmen, die auf maschinellem Lernen beruhen, wieder zusammengesetzt werden. Dadurch werden Personen in Termini algorithmisch berechneter Ähnlichkeitsrelationen rekonstituiert.

Mit dem Begriff der digitalen evaluativen Infrastruktur machen wir auf die vielschichtige Gestalt der Quantifizierung in digitalen Kontexten aufmerksam, wo sie den Ausgangs- und Endpunkt rekursiv miteinander verbundener Bewertungen bildet. Digitalisierte Bewertungen sind mit neuen Formen des Kategorisierens und Klassifizierens verbunden, durch die, wie Lury und Day (2019, S. 19) zeigen, eine Vielzahl von Kontexten laufend eingeschlossen und wieder ausgeschlossen werden. Die algorithmisch berechneten Kategorisierungsschemata evaluativer Infrastrukturen können nicht im Voraus bestimmt werden; sie sind emergent, hochgradig flexibel und werden ständig neu konfiguriert und über eine Maschinerie miteinander verbundener Bewertungsinstrumente hergestellt (Tracking-Geräte, Kundenratings, Algorithmen für maschinelles Lernen und vieles mehr). Der Begriff der evaluativen Infrastruktur unterstreicht den relationalen Charakter der Bewertung. Er hebt hervor, dass die Analyseeinheiten nicht einzelne Endgeräte oder Akteure sind, sondern der Komplex der Interdependenzen zwischen materieller Praxis, Wissen und sozialer Organisation (Bowker et al. 2019). Darüber hinaus verweist der Begriff auf die verschiedenen Feedbackschleifen, die in digitalen Kontexten zwischen Quantifizierung, Bewertung, Klassifizierung und Messung bestehen. Deshalb beschränken wir uns nicht auf die Reaktivität von Rankings und Ratings, sondern untersuchen auch deren Konnektivität und generatives Potenzial.

Evaluative Infrastrukturen erzeugen neue Weisen des Sehens (Fourcade und Healy 2017). Sie verändern, was sichtbar und unsichtbar gemacht wird und schaffen dadurch neue Möglichkeiten der Einflussnahme, Wertschöpfung und „Assetisierung“ („assetization“), wie Birch und Muniesa (2020) es formulieren. Plattformrankings und -ratings spornen Nutzerinnen zur Leistung an. Dadurch ermöglichen sie eine Kontrolle und Standardisierung der Plattformaktivitäten. Doch als Teil maschinenlesbarer Big Data üben solche Ratings und Rankings mehr als nur Kontrolle aus. Sie schaffen Gelegenheiten, etwas zu kaufen oder zu verkaufen und erweitern dadurch die Netzwerke von eBay, Airbnb und Uber. Damit machen digitale evaluative Infrastrukturen die Aktivitäten auf den Plattformen nicht nur steuerbar, sondern auch skalierbar. Diese Dualität und Rekursivität verdienen unsere Aufmerksamkeit. Der Begriff der digitalen evaluativen Infrastruktur erfordert, Bewertung aus einer dynamischen Perspektive zu betrachten – einer Perspektive, die in der Lage ist, die vielfältigen Weisen sichtbar zu machen, über die Praktiken des Quantifizierens, Messens, Klassifizierens und Aggregierens miteinander verknüpft, herausgefordert und verändert werden, unter Einbezug alter und neu geschaffener Formen sozialer und politischer Ordnung.

5 Fazit

Der Beitrag ging der Frage nach, inwieweit bestehende Praktiken der Bewertung, Quantifizierung, Klassifizierung und Kategorisierung durch digitale Technologien herausgefordert und verändert werden. Um einen analytischen Rahmen für die Untersuchung interaktiver, digitaler Bewertungstechnologien zu entwickeln, wie sie von Plattformorganisationen in Form von Ratings, Rankings, Reviews und Kundenaudits genutzt werden, haben wir die These vertreten, dass man digitale evaluative Infrastrukturen als Bewertungs*regime* begreifen sollte, anstatt den Fokus auf einzelne Bewertungsinstrumente zu legen. Darüber hinaus haben wir die spezifischen *Machteffekte* von digitalen evaluativen Infrastrukturen in den Mittelpunkt gestellt und gezeigt, dass Macht sowohl verteilt als auch konzentriert ist: Auf der einen Seite werden Aufsicht und Macht vom mittleren Management auf die Kundinnen verlagert. Gleichzeitig behalten die Plattformorganisationen aber eine strikte Kontrolle über die evaluative Infrastruktur, deren Architektur und die Daten, die sie generiert. Um diese Machteffekte zu erfassen und um deutlich zu machen, dass sich Macht innerhalb von digitalen evaluativen Infrastrukturen weder aus einer einzelnen Quelle speist noch in Termini einer pyramidenförmigen Hierarchie verstehen lässt, haben wir unter Rückgriff auf Deleuze (1992) den Begriff der „protokollarischen Kontrolle“ eingeführt (vgl. auch Galloway und Thacker 2004). Und schließlich haben wir im Unterschied zu Reaktivitäts- oder Performativitätsansätzen das *generative Potenzial* digitaler evaluativer Infrastrukturen hervorgehoben.

Digitale evaluative Infrastrukturen begünstigen neue Formen verteilter Intelligenz und Kontrolle, die unsere Aufmerksamkeit, unsere Entscheidungen und unsere Subjektivität prägen. Sie ermöglichen neue Weisen, auf die Handlungen von Individuen einzuwirken. Sie erleichtern Konnektivität und helfen dadurch, die Strukturen zu bilden, aus denen Plattformorganisationen bestehen. Zu einem großen Teil operieren digitale evaluative Infrastrukturen endogen, also über rein interne Prozesse, und stellen dadurch die Beziehungen zwischen Messung, Quantifizierung und Klassifizierung auf den Kopf. Wo führt uns das hin? Welche neuen Forschungsperspektiven und -fragen eröffnen sich durch diesen Fokus?

Erstens müssen wir der Herstellung neuer Sichtbarkeiten mehr Aufmerksamkeit schenken – also dem, was in digitalen evaluativen Infrastrukturen sichtbar und unsichtbar gemacht wird. Über Online-Ratings und -Rankings und andere Formen des digitalen Trackings lassen sich die Aktivitäten von Plattformnutzern sichtbar machen und steuern. Gleichzeitig entstehen neue Informationsasymmetrien. Über zahlreiche Kanäle werden die Daten von den Plattformnutzern zur Plattformbetreiberin transportiert, die Nutzer selbst haben dagegen kaum eine Möglichkeit, zu Informationen zu gelangen (Rosenblat und Stark 2016). Die Rechenschaftspflicht ist oft einseitig verteilt. So können Uber-Fahrer dank digitaler evaluativer Infrastrukturen zu den effizientesten Routen verpflichtet oder für das Fahrerlebnis der Passagierinnen zur Rechenschaft gezogen werden. Aber sie brauchen eigene Daten, wenn sie diese Bewertungen anfechten oder die Plattformorganisation für die ihnen zustehende Löhne zur Rechenschaft ziehen wollen. Wie die algorithmisch ermittelten Leistungsbewertungen von der Plattformorganisation selbst verwendet werden – etwa unter welchen Umständen Accounts deaktiviert werden – ist nicht oder kaum transparent.

Digitale evaluative Infrastrukturen formen und verändern die Sichtbarkeitsverhältnisse. Um zu erkennen, was wahrgenommen wird und was nicht, müssen wir die neuen sozialen Beziehungen und Interventionsmöglichkeiten (inklusive Widerstand), die diese Sphären der Sichtbarkeit und Unsichtbarkeit mit sich bringen, besser verstehen. In diesem Zusammenhang muss auch beachtet werden, dass diese Sphären mit Kreisläufen wechselseitiger und dezentraler Überwachung verwoben sind, an denen die Plattformnutzer oft unbewusst mitwirken. Eine der Herausforderungen liegt deshalb darin, die digital vermittelte Überwachung zu verfolgen: Wer oder was wurde wann sichtbar gemacht. Mindestens ebenso anspruchsvoll (und wichtig) ist jedoch, das innere Funktionieren digitaler evaluativer Infrastrukturen zu begreifen: das Zusammenspiel und die Vernetzung verschiedener kalkulativer Instrumente, die unterschiedlichen Beziehungen zwischen Online-Ratings und -Rankings und deren Einbettung in Big Data sowie die algorithmischen Kategorisierungsprozesse, die durch maschinelles Lernen ermöglicht werden. Um solche Einsichten zu gewinnen, müssen neue interdisziplinäre methodologische Ansätze entwickelt werden, die sich nicht nur mit den sozialen Bedingungen und Folgen digitaler evaluativer Infrastrukturen auseinandersetzen, sondern auch in der Lage sind, ihre technische Struktur aufzudecken und zu erfassen, was diese möglich macht (zu einem Versuch, Algorithmen methodisch erkennbar zu machen vgl. Ziewitz 2017).

Zweitens muss die Temporalität digitaler evaluativer Infrastrukturen freigelegt werden. Digitale evaluative Infrastrukturen ermöglichen potenziell mehr proaktive und prognostische Zugänge zu Bewertung und Governance. Evaluative Daten können gesammelt, verarbeitet und (fast) in Echtzeit rückgekoppelt werden. Welche Auswirkungen haben diese unmittelbaren Feedbackschleifen? Welche neuen Möglichkeiten eröffnen sie dem Management (und der „spielerischen“ Ausbeutung) von Arbeit? Was könnten ihre (nichtintendierten) negativen Folgen sein? Zudem sollten wir auch nicht vergessen, dass digitale evaluative Infrastrukturen zu sperrigen und schwerfälligen Kontrollapparaten verkommen können. Sie können eigene Dynamiken und Pfadabhängigkeiten erzeugen und sie sind womöglich nicht leicht zu verändern, da sie beträchtliche politische, intellektuelle und finanzielle Investitionen für ihren Aufbau und Unterhalt erfordern (Kornberger et al. 2017). Wir brauchen mit anderen Worten ein besseres Verständnis der möglichen Lock-in-Effekte und ihrer Folgen für die Struktur der Plattformorganisationen, der Arbeitsbeziehungen und für das „making up“ (Hacking 2002) der Plattformnutzerinnen (sowohl als Käufer wie als Verkäuferinnen).

Drittens und *viertens* sollten wir uns damit beschäftigen, wie digitale evaluative Infrastrukturen unterschiedliche Welten erweitern und verbinden. Wir brauchen mehr Wissen darüber, was in und durch digitale evaluative Infrastrukturen vernetzt wird und was nicht. Dies erfordert einen genaueren Blick auf die vielschichtige Topologie digitaler evaluativer Infrastrukturen. Wir müssen die Dynamiken besser verstehen, die durch die Verknüpfung verschiedener Bewertungsinstrumente (z. B. Rankings und Ratings, algorithmische Kategorisierung) entstehen. Zudem sollten wir auch die wechselnden Beziehungen und Dynamiken, die sich zwischen Klassifizierung, Messung und Aggregieren entwickeln, genauer unter die Lupe nehmen. Denn dort ist die Quantifizierung oft zugleich Ausgangs- und Endpunkt der Bewertung, wobei sich Messung und Klassifizierung gegenseitig bedingen. Darüber hinaus

müssen wir uns mit den Prozessen von Inklusion und Exklusion befassen. Es braucht einen genaueren Blick auf die Bedingungen der Teilnahme an digitalen evaluativen Infrastrukturen und auf die Frage, was oder wer durch sie ausgeschlossen wird und warum. (Ein Beispiel sind die Bewertungen, die zur Deaktivierung eines Accounts führen und deren Folgen.)

Selbstverständlich kann dieser Beitrag diese vielfältigen Themen nur sehr holzschnittartig umreißen. Dennoch hoffen wir, dass er weitere Debatten, disziplinübergreifende Diskussionen und Untersuchungen anregt, die für ein Verständnis der vielfältigen Rollen der Bewertung in neuen Formen digitaler Organisation notwendig sind.

Funding Open access funding provided by London School of Economics and Political Science.

Literatur

- Adkins, Lisa, und Celia Lury. 2011. Introduction: Special measures. *The Sociological Review* 59:5–23.
- Alaimo, Cristina, und Jannis Kallinikos. 2020. Managing by data. Algorithmic categories and organizing. *Organization Studies* forthcoming:1–23.
- Barnard, Alex V., und Marion Fourcade. 2021. Das Unbehagen an der Ordinalisierung. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. <https://doi.org/10.1007/s11577-021-00743-1>.
- Beckert, Jens, und Patrik Aspers. Hrsg. 2011. *The worth of goods. Valuation and pricing in the economy*. Oxford: Oxford University Press.
- Bentham, Jeremy. 1823 [1789]. *Principles of morals and legislation*. Oxford: Clarendon Press.
- Bentham, Jeremy. 1988 [1776]. *A fragment on government*. Cambridge: Cambridge University Press.
- Berthoin Antal, Ariane, Michael Hutter und David Stark. Hrsg. 2015. *Moments of valuation. Exploring sites of dissonance*. Oxford: Oxford University Press.
- Birch, Kean, und Fabian Muniesa. Hrsg. 2020. *Assetization. Turning things into assets in technoscientific capitalism*. Cambridge, Mass.: MIT Press.
- Boltanski, Luc, und Laurent Thévenot. 2006. *On justification. Economies of worth*. Princeton: Princeton University Press.
- Bowker, Geoffrey C., und Susan Leigh Star. 1999. *Sorting things out. Classification and its consequences*. Cambridge, Mass.: MIT Press.
- Bowker, Geoffrey C., Julia Elyachar, Martin Kornberger, Andrea Mennicken, Peter Miller, Joanne Randa Nucho und Neil Pollock. 2019. Introduction to thinking infrastructures. In *Thinking Infrastructures. Research in the Sociology of Organizations* 62, Hrsg. Martin Kornberger et al., 1–13. Bingley: Emerald.
- Brankovic, Jelena, Leopold Ringel und Tobias Werron. 2018. How rankings produce competition. The case of global university rankings. *Zeitschrift für Soziologie* 47:270–288.
- Callon, Michel. 1998. Introduction: The embeddedness of economic markets in economics. In *The Laws of the Markets*, Hrsg. Michel Callon, 1–57. Oxford: Blackwell.
- Carruthers, Bruce G. 2013. From uncertainty toward risk. The case of credit ratings. *Socio-Economic Review* 11:525–551.
- Deleuze, Gilles. 1992. Postscript on societies of control. *October* 59:3–7.
- Desrosières, Alain. 1998. *The politics of large numbers. A history of statistical reasoning*. Cambridge, Mass.: Harvard University Press.
- Dewey, John. 1915. The Logic of Judgments of Practise. *The Journal of Philosophy, Psychology and Scientific Methods* 12:505–523.
- Dewey, John. 1939. *Theory of valuation*. International Encyclopaedia of Unified Science, Vol. 2, 4.
- Douglas, Mary. 1986. *How institutions think*. Syracuse: Syracuse University Press.
- Espeland, Wendy N. 2001. Commensuration and cognition. In *Cognition in Mind*, Hrsg. Karen Cerulo, 63–88. London: Routledge.
- Espeland, Wendy N., und Michael Sauder. 2007. Rankings and reactivity. How public measures recreate social worlds. *American Journal of Sociology* 113:1–40.
- Espeland, Wendy N., und Michael Sauder. 2016. *Engines of anxiety. Academic rankings, reputation, and accountability*. New York: Russell Sage Foundation.

- Espeland, Wendy N., und Mitchell L. Stevens. 1998. Commensuration as a social process. *Annual Review of Sociology* 24:313–343.
- Espeland, Wendy N., und Stacy E. Lom. 2015. Noticing numbers. How quantification changes what we see and what we don't. In *Making Things Valuable*, Hrsg. Martin Kornberger et al., 18–36. Oxford: Oxford University Press.
- Esposito, Elena, und David Stark. 2019. What's observed in a rating? Rankings as orientation in the face of uncertainty. *Theory, Culture and Society* 36:3–26.
- Feinberg, Kenneth R. 2005. *What is life worth? The inside story of the 9/11 fund and its effort to compensate the victims of September 11th*. New York: Public Affairs.
- Foucault, Michel. 1977. *Discipline and punish. The birth of the prison*. London: Penguin.
- Fourcade, Marion. 2011. Cents and sensibility: Economic valuation and the nature of "nature". *American Journal of Sociology* 116:1721–1777.
- Fourcade, Marion, und Kieran Healy. 2017. Seeing like a market. *Socio-Economic Review* 15:9–29.
- Galloway, Alexander R. 2004. *Protocol. How control exists after decentralization*. Cambridge, Mass.: MIT Press.
- Galloway, Alexander R., und Eugene Thacker. 2004. Protocol, control, and networks. *Grey Room* 17:6–29.
- Gandini, Alessandro. 2019. Labour process theory and the gig economy. *Human Relations* 72:1039–1056.
- Gond, Jean-Pascal, Laure Cabantous, Nancy Harding und Mark Learmonth. 2016. What do we mean by performativity in organizational and management theory? The uses and abuses of performativity. *International Journal of Management Reviews* 18:440–463.
- Hacking, Ian. 1992. The self-vindication of the laboratory sciences. In *Science as Practice and Culture*, Hrsg. Andrew Pickering, 29–64. Chicago: University of Chicago Press.
- Hacking, Ian. 2002. Making up people. In *Historical Ontology*, Ian Hacking, 99–114. Cambridge, Mass.: Harvard University Press.
- Heintz, Bettina. 2010. Numerische Differenz. Überlegungen zu einer Soziologie des (quantitativen) Vergleichs. *Zeitschrift für Soziologie* 39:162–181.
- Heintz, Bettina. 2016. „Wir leben im Zeitalter der Vergleichung.“ Perspektiven einer Soziologie des Vergleichs. *Zeitschrift für Soziologie* 45:305–323.
- Heintz, Bettina. 2019. Vom Komparativ zum Superlativ. Eine kleine Soziologie der Rangliste. In *(Be)Werten. Beiträge zur sozialen Konstruktion von Wertigkeit*, Hrsg. Stefan Nicolae et al., 45–79. Wiesbaden: VS Springer.
- Heintz, Bettina. 2021. Kategorisieren, Vergleichen, Bewerten und Quantifizieren im Spiegel sozialer Beobachtungsformate. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. <https://doi.org/10.1007/s11577-021-00741-3>.
- Jeacle, Ingrid, und Chris Carter. 2011. In TripAdvisor we trust. Rankings, calculative regimes and abstract systems. *Accounting, Organizations and Society* 36:293–309.
- Kellogg, Katherine C., Melissa A. Valentine und Angèle Christin. 2020. Algorithms at work: The new contested terrain of control. *Academy of Management Annals* 14:366–410.
- Kornberger, Martin. 2017. The values of strategy. Valuation practices, rivalry and strategic agency. *Organization Studies* 38:1753–1773.
- Kornberger, Martin, und Chris Carter. 2010. Manufacturing competition. How accounting practices shape strategy making in cities. *Accounting, Auditing and Accountability Journal* 23: 325–349.
- Kornberger, Martin, Geoffrey C. Bowker, Julia Elyachar, Andrea Mennicken, Peter Miller, Joanne Randa Nucho und Neil Pollock. Hrsg. 2019. *Thinking Infrastructures. Research in the Sociology of Organizations* 62. Bingley: Emerald.
- Kornberger, Martin, Lise Justesen, Anders Koed Madsen und Jan Mouritsen. Hrsg. 2015. *Making things valuable*. Oxford: Oxford University Press.
- Kornberger, Martin, Dane Pflueger und Jan Mouritsen. 2017. Evaluative infrastructures: Accounting for platform organization. *Accounting, Organizations and Society* 60:79–95.
- Krüger, Anne K., und Martin Reinhart. 2017. Theories of valuation. Building blocks for conceptualizing valuation between practice and structure. *Historical Social Research/Historische Sozialforschung* 42:263–285.
- Lamont, Michèle. 2012. Toward a comparative sociology of valuation and evaluation. *Annual Review of Sociology* 38:201–221.
- Leoni, Giulia, und Lee D. Parker. 2019. Governance and control of sharing economy platforms. Hosting on Airbnb. *British Accounting Review* 51:1–22.
- Lury, Celia, und Sophie Day. 2019. Algorithmic personalization as a mode of individuation. *Theory, Culture and Society* 36:17–37.

- MacKenzie, Donald, und Yuval Millo. 2003. Constructing a market, performing theory. The historical sociology of a financial derivatives exchange. *American Journal of Sociology* 109:107–145.
- MacKenzie, Donald, Fabian Muniesa und Lucia Siu. Hrsg. 2007. *Do economists make markets? On the performativity of economics*. Princeton: Princeton University Press.
- Mehrpouya, Afshin, und Rita Samiolo. 2016. Performance measurement in global governance. Ranking and the politics of variability. *Accounting, Organizations and Society* 55:12–31.
- Mennicken, Andrea, und Wendy N. Espeland. 2019. What's new with numbers? Sociological approaches to the study of quantification. *Annual Review of Sociology* 45:223–245.
- Mennicken, Andrea, und Peter Miller. 2014. Foucault and the administering of lives. In *The Oxford handbook of sociology, social theory, and organization studies: Contemporary currents*, Hrsg. Paul S. Adler et al., 11–38. Oxford: Oxford University Press.
- Miller, Peter, und Nikolas Rose. 1990. Governing economic life. *Economy and Society* 19:1–31.
- Miller, Peter, und Nikolas Rose. 2008. *Governing the present: Administering economic, social and personal life*. Cambridge: Polity Press.
- Moor, Liz, und Celia Lury. 2018. Price and the person. Markets, discrimination, and personhood. *Journal of Cultural Economy* 11:501–513.
- Muniesa, Fabian. 2011. A flank movement in the understanding of valuation. *The Sociological Review* 59:24–38.
- Nietzsche, Friedrich. 1999 [1882]. Die fröhliche Wissenschaft. In *Friedrich Nietzsche Kritische Studienausgabe*, Hrsg. Giorgio Colli und Mazzino Montinari, Bd. 3, 343–651. München: dtv.
- Nietzsche, Friedrich. 2014 [1887]. Genealogie der Moral. Band 656 von *Philosophische Bibliothek*, Hrsg. Claus-Artur Scheier. Hamburg: Meiner.
- Oakes, Guy. 2003. Max Weber on value rationality and value spheres. Critical remarks. *Journal of Classical Sociology* 3:27–45.
- Orlikowski, Wanda J., und Susan V. Scott. 2014. What happens when evaluation goes online? Exploring apparatuses of valuation in the travel sector. *Organization Science* 25:868–891.
- Pollock, Neil, und Luciana D'Adderio. 2012. Give me a two-by-two matrix and I will create the market. Rankings, graphic visualisations and sociomateriality. *Accounting, Organizations and Society* 37:565–586.
- Power, Michael. 2019. Infrastructures of traceability. In *Thinking Infrastructures. Research in the Sociology of Organizations* 62, Hrsg. Martin Kornberger et al., 115–130. Bingley: Emerald.
- Rosenblat, Alex, und Luke Stark. 2016. Algorithmic labor and information asymmetries. A case study of Uber's drivers. *International Journal of Communication* 10:3758–3784.
- Sandel, Michael J. 2000. What money can't buy: The moral limits of markets. *Tanner Lectures on Human Values* 21:87–122.
- Seaver, Nick. 2019. Captivating algorithms. Recommender systems as traps. *Journal of Material Culture* 24:421–436.
- Srnicek, Nick. 2017. *Platform capitalism*. Cambridge: Polity Press.
- Star, Susan Leigh. 1999. The Ethnography of Infrastructure. *American Behavioral Scientist* 43:377–391.
- Star, Susan Leigh, und Karen Ruhleder. 2001. Steps toward an ecology of infrastructure: Design and access for large information spaces. In *Information Technology and Organizational Transformation*, Hrsg. JoAnne Yates und John Van Maanen, 305–346. Thousand Oaks: Sage.
- Stark, David. 2009. *The sense of dissonance. Accounts of worth in economic life*. Princeton: Princeton University Press.
- Stark, Luke, und Karen Levy. 2018. The surveillant consumer. *Media, Culture and Society* 40:1202–1220.
- Strathern, Marilyn. 2000. The tyranny of transparency. *British Educational Research Journal* 26:309–321.
- Unternährer, Markus. 2021. Die Ordnung der Empfehlung. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*. <https://doi.org/10.1007/s11577-021-00753-z>.
- Vallas, Stephen, und Juliet B. Schor. 2020. What do platforms do? Understanding the Gig Economy. *Annual Review of Sociology* 46:273–294.
- Veyne, Paul. 1978. Foucault révolutionne l'histoire. In *Comment on écrit l'histoire*, Paul Veyne. Paris: Editions du Seuil.
- Weber, Max. 2002 [1922]. *Wirtschaft und Gesellschaft. Grundriss der verstehenden Soziologie*. Tübingen: Mohr Siebeck.
- Whitson, Jennifer R. 2014. Foucault's Fitbit. Governance and gamification. In *The gameful world*, Hrsg. Steffen P. Walz und Sebastian Deterding, 339–358. Cambridge, Mass.: MIT Press.
- Yoffie, David B., Annabelle Gawer und Michael A. Cusumano. (im Erscheinen). The platform organization. In *Designing modern organizations*, Hrsg. Charles Snow und Øystein D. Fjeldstad. Cambridge: Cambridge University Press.

- Zerubavel, Eviatar. 1996. Lumping and splitting. Notes on social classification. *Sociological Forum* 11:421–433.
- Ziewitz, Malte. 2017. A not quite random walk. Experimenting with the ethnomethods of the algorithm. *Big Data and Society* 4:1–13.
- Zuboff, Shoshana. 2019. *The age of surveillance capitalism. The fight for a human future at the new frontier of power*. New York: Public Affairs.

Andrea Mennicken. 1973, PhD, Associate Professor of Accounting, London School of Economics and Political Science. Forschungsgebiete: Soziologie der Bewertung und Quantifizierung, Regulierung, Accounting und Audit, Organisations- und Wirtschaftssoziologie. Veröffentlichungen: What's New with Numbers? Sociological Approaches to the Study of Quantification. *Annual Review of Sociology* 45, 2019 (mit W. N. Espeland); Zahlenwerk: Kalkulation, Organisation und Gesellschaft. Wiesbaden 2007 (als Hrsg., mit H. Vollmer).

Martin Kornberger. 1974, Dr., Professor für Strategie, University of Edinburgh. Forschungsgebiete: Strategie und Organisation kollektiven Handelns, Wirtschaftsphilosophie. Veröffentlichungen: Brand Society. Cambridge University Press, 2010; Making Things Valuable. Oxford 2015 (als Hrsg., mit L. Justesen, A. K. Madsen und J. Mouritsen); Thinking Infrastructures, Emerald 2019 (als Hrsg., mit G. C. Bowker, A. Mennicken, P. Miller, J. R. Nucho und N. Pollock); Figures of Thought for Distributed and Collective Action, Oxford University Press, im Erscheinen.