Kategorisieren, Vergleichen, Bewerten und Quantifizieren im Spiegel sozialer Beobachtungsformate

Bettina Heintz

Eingegangen: 25. Februar 2021 / Angenommen: 12. März 2021 / Online publiziert: 28. Mai 2021
© Der/die Autor(en) 2021

Zusammenfassung In den letzten Jahren haben sich rund um die Praktiken des Vergleichens, Bewertens, Kategorisierens und Quantifizierens eigenständige Forschungsfelder etabliert: die Soziologie der Bewertung und des Vergleichs, die Kategorisierungs- und die Quantifizierungsforschung. Das Sonderheft der *Kölner Zeitschrift für Soziologie und Sozialpsychologie* rückt die Verbindungen zwischen den vier Forschungsfeldern in den Vordergrund. Am Beispiel empirischer Fallstudien zeigt es auf, welchen Gewinn es bringt, wenn man die vier Praktiken nicht unabhängig voneinander untersucht, sondern sie in ihrem Zusammenwirken analysiert. Die Fallstudien befassen sich mit unterschiedlichen Formaten sozialer Beobachtung und decken ein breites thematisches Spektrum ab: Die Themen reichen von der Kolonialstatistik und dem frühen Staatenvergleich, der Kolonial- und der inter/nationalen Statistik, über Monitoringsysteme und Rankings bis hin zu Recommendersystemen, digitalen Plattformen und prädiktiven Algorithmen. Der einleitende Beitrag entwickelt den theoretischen Rahmen, der dem Band zugrunde liegt, und setzt die Beiträge zueinander in Beziehung.

Schlüsselwörter Digitale Technologien · Nationale und internationale Statistik · Monitoringsysteme · Rankings · Historischer Wandel

B. Heintz (✉)
Soziologisches Seminar, Universität Luzern
Froburgstr. 3, 6002 Luzern, Schweiz
E-Mail: bettina.heintz@unilu.ch

Categorizing, Comparing, Evaluating, and Quantifying in the Mirror of Social Observation Formats

Abstract So far, practices of classification, comparison, evaluation, and quantification have been studied in separate research fields. In contrast to this research practice, the special issue of the *Kölner Zeitschrift für Soziologie und Sozialpsychologie* brings their connections to the fore. The case studies collected in this special issue demonstrate the insights one gains by investigating how these four practices interrelate. The contributions deal with different formats of observation and cover a broad thematic spectrum: colonial statistics and early forms of comparing and grading states, inter/national statistics, monitoring systems, rankings and digital technologies such as recommender systems, digital platforms, and predictive algorithms. In the introductory article, I outline the theoretical framework of this issue and show how the contributions are connected.

Keywords Digital technologies · National and international statistics · Monitoring · Rankings · Historical change

1 Einleitung

Vergleichen, Bewerten, Kategorisieren und Quantifizieren stellen grundlegende kognitive und soziale Ordnungsverfahren dar. Über Kategorien werden singuläre und disparate Phänomene vereinheitlicht, über Vergleiche werden Sachverhalte hinsichtlich ihrer Ähnlichkeiten und Unterschiede beobachtet, über Bewertungen werden sie in eine Rangfolge gebracht und über Quantifizierung in numerische Größen übersetzt. Alle vier Verfahren reduzieren auf je eigene Art Komplexität. Kategorisierung schützt uns vor der unermesslichen Fülle der Eindrücke, indem sie den Strom der Ereignisse in kompakte „Dinge“ sortiert und die Welt in Gleiches und davon Verschiedenes unterteilt (Zerubavel 1996). Vergleiche erzeugen Ordnung, indem sie zwischen den verglichenen Einheiten einen Sinnzusammenhang herstellen (Heintz 2016). Bewertungen reduzieren Komplexität, indem sie Alternativen in eine ordinale Reihenfolge bringen (Aspers 2018), und Quantifizierung verdichtet Sachverhalte in Zahlen (Espeland und Stevens 2008).

Die Forschung zu diesen vier Ordnungsverfahren ist bis heute arbeitsteilig organisiert. Wer zu Kategorisierungsprozessen forscht, greift nur sporadisch auf vergleichssoziologische Überlegungen zurück, und wer sich mit Quantifizierung beschäftigt, interessiert sich nur selten für die Erkenntnisse der Bewertungssoziologie. Der vorliegende Band will dieser Fragmentierung entgegenwirken. Anstatt die vier Verfahren je für sich zu betrachten, belegt er, welchen Gewinn es bringt, wenn man sie in ihrem Zusammenwirken analysiert. Denn Vergleich, Bewertung, Kategorisierung und Quantifizierung hängen nicht nur begrifflich eng zusammen, sondern bedingen einander auch in ihrem praktischen Vollzug. Wie dies im Einzelnen geschieht, untersuchen die Beiträge anhand verschiedener Fallbeispiele, die sich drei thematischen Schwerpunkten zuordnen lassen: inter/nationale Statistiken (Teil 2), Rankings und Monitoringsysteme (Teil 3) und digitale Technologien (Teil 4). Vor-

angestellt sind vier Beiträge, die das historische Feld abstecken und für den Band Grundlagencharakter haben (Teil 1). Um die Verbindungen zwischen den Beiträgen sichtbar zu machen, beschreibe ich im Folgenden den theoretischen Rahmen, der dem Band zugrunde liegt. Ich werde zunächst in die allgemeine Thematik einführen und anschließend die einzelnen Beiträge ausführlicher vorstellen und sie in diesem Rahmen verorten und aufeinander beziehen.

Digitale Technologien sind ein besonders markantes Beispiel für die Verflochtenheit der vier Operationen, wie die im letzten Teil des Bandes versammelten Beiträge verdeutlichen. In detaillierten Analysen der algorithmischen Verfahren zeigen sie, dass die Besonderheit digitaler Technologien nicht bloß darin liegt, dass sie Algorithmen einsetzen und auf Massen von Daten – Big Data – zurückgreifen. Aus der Perspektive des Bandes betrachtet, besteht ihre Besonderheit vielmehr darin, dass sie Vergleich, Kategorisierung und Bewertung in einem rekursiven Prozess miteinander verknüpfen und sich diese Verknüpfung im Medium algorithmischer Berechnungen vollzieht: Die aus den Netzaktivitäten der Nutzer und Nutzerinnen gewonnenen Daten werden von Algorithmen verglichen, bewertet und in Kategorien sortiert, bis am Ende eine Rangliste resultiert – von Arbeitssuchenden, potenziellen Tälern, Produkten oder zukünftigen Partnern (ausführlicher Abschn. 5).

Nicht nur digitale Technologien kombinieren Vergleich, Bewertung und Kategorisierung. Dies geschieht auch bei jeder Entscheidung, nur zeitlich verschoben und nicht automatisiert. Ein alltägliches Beispiel sind Bewerbungsverfahren. Auf der Basis ihrer Unterlagen werden die Bewerber und Bewerberinnen hinsichtlich der Anstellungskriterien verglichen, gerankt und in Kategorien einsortiert. Das ist ein relativ mechanischer Prozess, der zunehmend auch von Algorithmen übernommen wird (Ajunwa und Greene 2019). Es geht um ein Matching zwischen dem Stellenprofil und dem Profil der Bewerber, bei dem alle drei Operationen involviert sind: Vergleich, Bewertung und Kategorisierung.

Ähnlich funktionieren auch Markttransaktionen – jeder Kauf erfordert vorgängige Kategorisierungs-, Vergleichs- und Bewertungsakte. Wie voraussetzungsvoll die Verbindung dieser drei Operationen ist, wird in der Marktsoziologie unter dem Begriff der „quality uncertainty“ diskutiert (als Überblick Beckert und Musselin 2013). *Erstens* braucht es Produktkategorien. Um ein neues Produkt zu lancieren, muss es von bereits bestehenden Produkten abgegrenzt werden – E-Book-Reader von gewöhnlichen Büchern, Hafermilch von Kuhmilch. Wie aus Untersuchungen zur Lancierung neuer Produkte hervorgeht, ist ihre Einführung vor allem mit dem Problem konfrontiert, ihnen eine kategoriale Identität zu verleihen (exemplarisch Ormrod 1994; Navis und Glynn 2010). Solange dies nicht gelingt, stehen auch keine Kriterien zur Verfügung, um ihre Funktion und Qualität zu beurteilen. Der Vergleich – und damit auch der Kauf – scheitert mit anderen Worten an ihrer kategorialen Unbestimmtheit (Zuckerman 1999). *Zweitens* müssen die Käufer in der Lage sein, die angebotenen Produkte der richtigen Produktkategorie zuzuordnen. Damit dies gelingt, werden sie durch Bezeichnungen, Siegel oder ihre Platzierung in den Verkaufsregalen als Exemplare einer bestimmten Produktkategorie kenntlich gemacht. Schließlich braucht es *drittens* Kriterien, um die verschiedenen Exemplare einer Produktkategorie miteinander zu vergleichen und sie bewerten zu können. Bei Standardgütern, für die es konsensuale und oft quantifizierte Beurteilungskri-

terien gibt, lässt sich ein solcher Vergleich relativ leicht bewerkstelligen. Anders sieht es bei Gütern aus, bei denen die Beurteilungskriterien vage sind oder keine standardisierten Verfahren zur Verfügung stehen, um sie auf einen konkreten Fall anzuwenden (Bühler 2019). In solchen Fällen braucht es Entscheidungshilfen – Lucien Karpik (2010) nennt sie „judgment devices“ –, die die Bewertung erleichtern. Das können Rankings sein, bibliometrische Indikatoren, Gütesiegel oder auch die Empfehlungslisten von Recommendersystemen.

Die drei Beispiele illustrieren auf unterschiedliche Weise, dass sich Bewerten nicht unabhängig von Vergleichen und Vergleichen und nicht separiert von Kategorisierungsakten vollzieht. Alle drei Operationen sind zwar analytisch unterscheidbar, empirisch sind sie aber eng miteinander verknüpft. Angesichts dieser faktischen Verwobenheit ist es erstaunlich, dass die ihnen zugeordneten Forschungsgebiete nur selten über die Mauern ihres eigenen Gevierts hinausblicken. Man pflegt den eigenen Begriff bis in seine kleinsten Verästelungen hinein und nimmt die Forschung in den anderen Gebieten höchstens am Rande zur Kenntnis. Bezeichnend für diesen Tunnelblick sind die Keywords, die den einschlägigen Texten mitgegeben werden. Michelle Lamont führt in ihrem Aufsatz „Toward a comparative sociology of valuation and evaluation“ (2012), der von vielen als Gründungstext der neuen Bewertungssoziologie angesehen wird, „worth, value, evaluative practices, social processes, heterarchies“ als Schlüsselwörter an. Andrea Mennickens und Wendy Espelands „What’s New with Numbers?“ (2019) figuriert unter „quantification, commensuration, economization, quantified self, politics of visibility“. Eviatar Zerubavels „Lumping and Splitting“-Aufsatz (1996) wird unter „classification, categories, clusters, language, cognitive sociology, topological thinking, similarities, differences“ aufgeführt und Stefan Hirschauers (2014) Aufsatz „Un/doing Differences“ unter „Soziale Differenzierung, Klassifikation, Gender, Ethnizität, Hybridität“. Querverweise finden sich keine, obschon die anderen Begriffe latent immer präsent sind.

Die Beiträge dieses Bandes untersuchen das Wechselspiel zwischen Kategorisierung, Vergleich, Quantifizierung und Bewertung an ganz unterschiedlichen Fällen, die von Kolonialstatistiken (Renard) und dem europäischen Staatenvergleich im 18. und 19. Jahrhundert (W. Steinmetz) über Kunst- und Hochschulrankings (Rügel und Werron) und den deutschen Integrationsmonitor (Petzke) bis hin zu Dating-Apps (Peetz), digitalen Plattformen (Mennicken und Kornberger) und dem Einsatz prädiktiver Algorithmen bei der Arbeitsvermittlung (Büchner und Dosdall) und in der Polizeiarbeit (Brayne) reichen. Eine erste Verbindung zwischen diesen auf den ersten Blick disparat anmutenden Beispielen wird über die Fragestellung des Bandes hergestellt: Wie hängen die vier Verfahren begrifflich und empirisch zusammen? Lässt sich auch ein Integrationsmonitor unter bewertungssoziologischen Gesichtspunkten untersuchen und was ist der Gewinn einer solchen Perspektive? Was sieht man anders und vielleicht auch mehr, wenn man Dating-Apps nicht nur aus der Perspektive der Bewertungssoziologie, sondern auch unter dem Aspekt der Kategorisierung und des Vergleichs analysiert?

Eine zweite Verbindungslinie ergibt sich aus dem Vorschlag, amtliche Statistiken, Rankings, Monitoringinstrumente und digitale Technologien als soziale *Beobachtungsformate* zu interpretieren (Heintz). Über das Konzept des Beobachtungsformats gewinnt man einen Bezugsrahmen, der es ermöglicht, die Fallbeispiele hinsichtlich

der Fragestellung des Bandes zu vergleichen und sie historisch zueinander in Beziehung zu setzen. Ihre Gemeinsamkeit liegt darin, dass es sich in allen Fällen um Instrumente handelt, die regelmäßig Daten erzeugen und in diesen Daten nach Mustern suchen, aus denen sich Planungs- und Steuerungswissen ableiten lässt. Sie sind insofern Teil des Quantifizierungsschubs, der die (westlichen) Gesellschaften seit dem 18. Jahrhundert erfasst hat (Porter 1995).

Das Konzept des Beobachtungsformats weist eine gewisse Nähe zum *Surveillance*-Begriff auf (Lyon 2007; aus historischer Perspektive Reichardt 2016). Im Unterschied zu den *Surveillance Studies* legt der Band den Akzent aber auf Beobachtung und Mustererkennung und nicht auf Überwachung und Kontrolle (zu diesem Unterschied auch Marx 2015). Diesen Fokus teilt er mit dem Buch von Armin Nassehi (2019), der die amtliche Statistik und digitale Technologien ebenfalls als Instrumente der Mustererkennung begreift, aber vor allem ihre Gemeinsamkeiten hervorhebt. Im Unterschied zu Nassehi beziehen die hier versammelten Aufsätze noch weitere Beobachtungsinstrumente mit ein und zeigen anhand detaillierter Fallstudien auf, wie die verschiedenen Beobachtungsformate Muster erkennen und worin sie sich unterscheiden. Aus den Gemeinsamkeiten und Unterschieden lässt sich daher nicht nur etwas über den Zusammenhang zwischen den vier Ordnungsverfahren lernen, sondern auch über den Wandel der gesellschaftlichen Beobachtungstechniken und ihrer kulturellen und sozialen Voraussetzungen. So gesehen sind Selbstvermessungsgeräte, Suchmaschinen, Dating-Apps oder die im Onlinehandel eingesetzten Recommendersysteme ein weiterer Schritt in der – allerdings noch zu schreibenden – Geschichte sozialer Beobachtungsformate, die mit der Entstehung der modernen Statistik im späten 18. Jahrhundert begann und über die Meinungs- und Marktforschung, Rankings und Monitoringsysteme bis hin zu den heutigen digitalen Technologien reicht. Das Aufzeigen historischer Kontinuitäten trägt zudem dazu bei, die etwas schrille Neuheitsrhetorik, die sich um diese Technologien rankt, zu relativieren.

Die in diesem Band angesprochenen Forschungsgebiete – die Quantifizierungs- und Kategorisierungsforschung, die Soziologie der Bewertung und des Vergleichs – haben sich in den letzten Jahren zu breit rezipierten Forschungsfeldern entwickelt. Während die Klassifikations- und die Bewertungssoziologie eine lange Tradition haben,¹ begann sich die Soziologie im Vergleich zu anderen Fächern erstaunlich spät für Quantifizierungsprozesse zu interessieren und die vielfältigen Formen des Zahlengebrauchs zueinander in Beziehung zu setzen (Espeland und Stevens 2008; Diaz-Bone und Didier 2016; Mennicken und Espeland 2019). Noch länger dauerte es, bis Vergleiche nicht mehr nur als wissenschaftliche Methode verstanden, sondern als ein soziales Phänomen eigenen Rechts behandelt wurden. Erst in den letzten Jahren und angestoßen durch Niklas Luhmanns (1995a) Arbeiten zum Kulturbegriff hat sich eine eigentliche Soziologie des Vergleichs entwickelt, die das Vergleichen als eine alltägliche soziale Praxis begreift und deren Ausdrucksformen und Folgen untersucht (Heintz 2016; Steinmetz 2019a; Bennani et al. 2020; Eppele et al. 2020).

¹ Vergleiche nur Durkheim und Mauss (1987/1902) sowie Dewey (1939). Für die neuere Forschung siehe exemplarisch Bowker und Star (2000), Hirschauer (2017), Lamont (2012) sowie Peetz et al. (2016).

Dass es sich um vergleichsweise junge Forschungsfelder handelt, mag mit ein Grund dafür sein, weshalb es bis jetzt noch kaum Bestrebungen gibt, über mögliche Zusammenhänge nachzudenken. Solche Zusammenhänge manifestieren sich nicht nur auf empirischer Ebene, sondern sie sind auch konzeptioneller Natur. Um die Verbindungen zwischen den vier Gebieten aufzuzeigen, bietet sich vor allem der Begriff des Vergleichs als ein Brückenkonzept an. Denn ob Kategorisierung, Vermessung oder Bewertung – alle setzen einen Vergleich voraus oder resultieren in einem Vergleich.

Vergleiche sind Ordnungsverfahren, die Sachverhalte auf ihre Ähnlichkeiten oder Unterschiede hin beobachten und sie dadurch zueinander in Beziehung setzen (Heintz 2016). Um etwas mit etwas anderem zu vergleichen, müssen zwei Bedingungen gegeben sein: Die zu vergleichenden Sachverhalte müssen erstens als vergleichbar eingestuft, d. h. derselben, wie auch immer abstrakten, *Kategorie* zugeordnet werden. Jeder Vergleich unterstellt mit anderen Worten eine minimale Gleichartigkeit des Verglichenen. Ob Sachverhalte als vergleichbar angesehen werden oder nicht, liegt nicht in der Sache selbst, sondern variiert je nach historischem und sozialem Kontext. Zweitens braucht es *Vergleichskriterien*, die die Unterschiede oder Ähnlichkeiten beobachtbar machen. Vergleiche können sich unterschiedlicher Kommunikationsmedien bedienen: Sie können sprachlich in Texten, visuell in Bildern oder numerisch in Zahlen mitgeteilt werden oder alle drei Medienarten kombinieren (Heintz 2010). Einige Vergleiche beschränken sich auf nur einen Vergleichsgesichtspunkt, andere haben eine komplexe Struktur, indem sie verschiedene Vergleichskriterien einsetzen, um Gleichheit oder Differenz festzustellen. Ein Beispiel für komplexe Vergleiche sind Universitätsrankings, die verschiedene Parameter verwenden und über einen Vergleich ihrer mutmaßlichen Relevanz ihr jeweiliges Gewicht bestimmen (Sauder und Espeland 2006, S. 208 f.). Und schließlich können sich Vergleiche auch hinsichtlich ihrer Zeitlichkeit unterscheiden: Es können synchrone Ereignisse miteinander verglichen werden oder Ereignisse, die zeitlich auseinanderliegen (zu verschiedenen Arten von Vergleichen und ihrer temporalen Struktur *Ringel* und *Werron* in diesem Band).

Bewertungen sind eine Sonderform von Vergleichen: Etwas ist besser, schöner, leistungsfähiger als etwas anderes. John Dewey (1939) spricht in diesem Zusammenhang von „appraising“ und definiert „appraising“ „as an act that involves comparison“ (S. 5). Den meisten Bewertungen geht mit anderen Worten ein Vergleich voraus – eine Abschätzung des relativen Wertes von Personen, Gütern, Personen oder Handlungen.² Der Unterschied zu nicht wertenden Vergleichen liegt darin, dass Bewertungen Objekte nicht nur hinsichtlich ihrer Ähnlichkeiten und Differenzen beobachten, wie das z. B. Statistiken tun, sondern sie zusätzlich in eine ordinale Reihenfolge bringen (z. B. Rankings).

² Ein Sonderfall sind Bewertungen, die sich als subjektive und absolute Urteile über ein Objekt zu erkennen geben: „Zazie ist großartig“. Dewey (1939) spricht hier von „prizing in the sense of holding precious, dear“ (Dewey 1939, S. 5). Es ist allerdings eine offene Frage, ob nicht auch solche absoluten Urteile auf einem komparativen Akt beruhen, ähnlich wie auch die Behauptung von Einzigartigkeit einen Vergleich voraussetzt (vgl. Abschn. 2).

Auch *Zahlen* sind auf Vergleiche hin angelegt oder setzen diese voraus.³ Jede Zahl, die man liest, wird automatisch in ein Verhältnis zu anderen Zahlen gesetzt. Nur für sich allein ergibt sie keinen Sinn. Wenn man erfährt, dass die Durchschnittstemperatur in der Schweiz im Januar 2020 $-2,3^{\circ}\text{C}$ betrug, weiß man noch nicht viel mehr. Die Zahl gewinnt erst dann eine Bedeutung, wenn man sie mit anderen Durchschnitts- oder früheren Temperaturen vergleicht. Aber nicht nur die Deutung, auch die Herstellung von Zahlen vollzieht sich über Vergleiche. Vergleiche sind bereits bei der Kodierung und Aggregation von Daten involviert, etwa wenn zu entscheiden ist, anhand welcher Schwellenwerte Kategorien zusammengefasst werden sollen, und bilden anschließend die Grundlage für die Erstellung von Tabellen (Bennani und Müller). Die Tabelle selbst ist ein Statistiken eigenes Kommunikationsmedium, das mit seinem Rasterformat den Vergleich auch visuell zur Anschauung bringt (Heintz).

Etwas komplexer ist das Verhältnis zwischen Vergleichen und *Kategorisieren*. Auf der einen Seite können Sachverhalte nur dann miteinander verglichen werden, wenn sie vorher als prinzipiell gleichartig eingestuft, d.h. der gleichen Kategorie zugeordnet wurden (s. o.). Um Menschen mit mentalen und organischen Einschränkungen hinsichtlich ihrer Beschäftigungsmöglichkeiten vergleichen zu können, müssen sie als Angehörige der gleichen Kategorie wahrgenommen werden, was erst in der zweiten Hälfte des 20. Jahrhunderts geschah. Vorher wurden „Blinde“, „Tauben“ oder „Idioten“ als unterschiedliche Menschensorten angesehen, die nicht in einer gemeinsamen Oberkategorie zusammengefasst werden können (Bennani und Müller 2018). Auf der anderen Seite beruht auch Kategorisierung auf einem vorgängigen Vergleich. Wer jemanden als behindert einstuft, vergleicht diese Person unwillkürlich mit seinem Bild eines nichtbehinderten Menschen. Solche Kategorisierungen vollziehen sich in vielen Fällen automatisch und nicht bewusst, aber dennoch liegt ihnen ein Vergleich zugrunde. Es handelt sich also um zwei Ordnungsverfahren, die im praktischen Vollzug aufs Engste zusammenhängen, aber trotzdem zu unterscheiden sind.

2 Kategorisieren, Vergleichen, Bewerten und Quantifizieren in historischer Perspektive

Die vier Beiträge des ersten Teils spannen einen weiten historischen Bogen und reichen zeitlich vom späten 17. Jahrhundert bis zur Gegenwartsgesellschaft. Zusammengefasst beschreiben sie den Gestaltwandel, den das Verhältnis von Vergleichen, Kategorisieren, Bewerten und Quantifizieren in diesem Zeitraum durchgemacht hat. Ich werde deshalb etwas ausführlicher auf den historischen Kontext eingehen, von dem dieser Gestaltwandel ausging.

Für Niklas Luhmann sind Vergleiche eine spezifisch moderne Ordnungstechnik, die sich erst im Laufe des 18. Jahrhunderts durchgesetzt und das Denken in asymmetrischen Unterscheidungen und Analogien bis zu einem gewissen Grade ver-

³ Unter „Zahlen“ (im Unterschied zu Ziffern) verstehe ich im Folgenden quantitative Werte, die auf Messen oder Zählen beruhen.

drängt habe (Luhmann 1995a, 1995b). Luhmann konzidiert zwar, dass es auch in ständischen Gesellschaften Vergleiche gab – Nationenvergleiche entlang von Adelskriterien oder Vergleiche zur Bestimmung der Rangverhältnisse –, aber diese waren aus seiner Sicht eher selten und auf Vergleiche innerhalb der Schichten beschränkt. Für das wissenschaftliche Vergleichen konstatiert Michel Foucault (1974, S. 84 ff.) einen ähnlichen Verlauf, er lässt die Hinwendung zum Vergleich aber schon im 17. Jahrhundert mit René Descartes beginnen. Mit dem Übergang zum „Klassischen Zeitalter“ sei das Vergleichen zur wissenschaftlichen Erkenntnismethode schlechthin avanciert und habe das analogisierende Denken der Renaissance durch eine Analyse in Termini von Identität und Differenz ersetzt. Neuere Untersuchungen belegen jedoch, dass die Geschichte des Vergleichens nicht so linear verlief, wie es Luhmann und Foucault annehmen. Dies gilt sowohl für den Vergleich als *wissenschaftliche Methode* (exemplarisch Eggers 2011, 2016) wie auch für den Vergleich als *soziale Praxis* (W. Steinmetz 2019a; Arlinghaus und Schuster 2020).

Zum einen waren schichtenübergreifende Vergleiche auch in der Vormoderne keineswegs unbekannt. Aufgrund der konstitutiven Bedeutung von Ehre waren Rangvergleiche geradezu endemisch für die vormoderne Gesellschaft, insbesondere an den Übergangszonen zwischen den einzelnen Schichten (Stollberg-Rilinger 2014). Ein Vergleich von Personen, die unterschiedlichen Schichten angehören – Willibald Steinmetz (2019b) nennt sie *Above/below*-Vergleiche –, galten zwar in einer Gesellschaft, die auf Rangdifferenzen beruhte, als unehrenhaft und als sozial gefährlich, dennoch wurden sie ständig praktiziert. Zum anderen ist es verkürzt, von *dem* Vergleich zu sprechen. Es gab (und gibt) verschiedene Formen des Vergleichs, die sich gegenseitig stimulierten und konkurrierten. Neben den erwähnten *Above/below*-Vergleichen unterscheidet Willibald Steinmetz (2019b) noch zwei weitere Varianten von Personenvergleichen, die er als *Better/worse*- und als *Simply-different*-Vergleiche bezeichnet.

Besser/schlechter-Vergleiche unterschieden sich von den früheren Oben/unten-Vergleichen dadurch, dass sich der Vergleich nun auf *individuelle* Merkmale bezog. Individuen wurden unabhängig von ihrem Stand anhand ihres Charakters, Benehmens oder ihrer Leistung verglichen und bewertet. „The most obvious way to distinguish oneself was now to become ‘better than’ or ‘less good than’ anyone else. Individual value and merit replaced traditional or attributed rank as the major criterion for social comparison“ (W. Steinmetz 2019b, S. 88). Diese Vergleichsform breitete sich seit Mitte des 18. Jahrhunderts aus, im Zusammenhang mit der Durchsetzung funktionaler Differenzierung und der sie begleitenden Gleichheitssemantik. Erst als man in jedem Menschen den „allgemeinen Menschen“ (Simmel) sah, wurde es möglich, Menschen unabhängig von ihrer sozialen Lage demselben Vergleichsraster zu unterziehen. „Modern“ wurde der Vergleich mit anderen Worten erst dann.⁴ Die mit *Besser/schlechter*-Vergleichen einhergehende Vorstellung einer grundsätzlichen Vergleichbarkeit blieb allerdings nicht unbestritten. Praktisch zeitgleich und in Reaktion darauf kam ein neuer und ebenfalls moderner Typus von Vergleichen auf, der sich dagegen wandte, Menschen kategorial zu vereinheitlichen und auf die

⁴ „Allgemeinmenschliches“ wurde allerdings nur den (männlichen) Bürgern des eigenen Landes attestiert, wie Léa Renard am Beispiel der deutschen Kolonialstatistik belegt (vgl. Abschn. 3).

Verschiedenheit oder sogar Unvergleichbarkeit des Einzelnen pochte: Jeder Mensch ist anders – eben *simply different*. Deshalb kann man ihn zwar von anderen unterscheiden, aber vergleichen kann man ihn nicht. Aber auch Behauptungen einer grundsätzlichen Andersartigkeit beruhen letztlich auf einem Vergleich, denn anders oder sogar ungleichbar kann man nur im Vergleich zu anderen sein – kein Genie ohne Mittelmaß.

Beide modernen Formen des Personenvergleichs – Besser/schlechter- und Einfach-anders-Vergleiche – haben ihre kulturelle Basis in den beiden Formen des Individualismus, die Georg Simmel unterschieden hat (Simmel 1995[1901], 1999[1917]). Der im 18. Jahrhundert entstehende „Individualismus der Gleichheit“, Simmel bezeichnet ihn auch als „quantitativen Individualismus“, hebt das allen Menschen Gemeinsame – „den Menschen als solchen“ (Simmel 1999[1917], S. 132) – hervor und macht sie darüber vergleichbar. Die Vorstellung einer in ihrem Menschsein begründeten Gleichartigkeit ist die Voraussetzung dafür, um sie hinsichtlich ihrer Unterschiede vergleichen und bewerten zu können. Der Individualitätsbegriff des späten 18. und 19. Jahrhunderts – Simmel spricht hier von „qualitativem Individualismus“ – stellt demgegenüber die Einzigartigkeit des Individuums in den Mittelpunkt. „Sobald das Ich im Gefühl der Gleichheit und Allgemeinheit hinreichend erstarrt war, suchte es wieder die Ungleichheit, aber nur die von innen heraus gesetzte. ... Nicht mehr darauf, dass man überhaupt ein freier Einzelner ist, kommt es an, sondern dass man dieser Bestimmte und Unverwechselbare ist“ (Simmel 1999[1917], S. 143). Bekannte Beispiele für diesen „Individualismus der Andersheit“ (Simmel) sind der Geniekult des ausgehenden 18. Jahrhunderts oder das Ideal der romantischen Liebe, die im Erleben der Unverwechselbarkeit des Anderen ihre Bestätigung findet.

Aber nicht nur Individuen, auch Kulturen, Staaten oder ganze Personenkategorien wurden nach dem Schema „besser/schlechter“ oder „einfach anders“ verglichen. Ein frühes Beispiel solcher Besser/schlechter-Vergleiche sind die europäischen Staatenvergleiche des 18. und 19. Jahrhunderts, denen *Willibald Steinmetz* in seinem Beitrag nachgeht. Auch „Einfach-anders“-Vergleiche waren nicht auf den Vergleich von Individuen beschränkt. Neben dem Kulturvergleich (s. u.) waren es vor allem Geschlecht und „Rasse“, an denen die Behauptung einer fundamentalen Verschiedenheit exemplifiziert wurde. Paradoxerweise wurde die Behauptung einer grundsätzlichen Unvergleichbarkeit aus einem akribischen Vergleich ihrer mentalen und körperlichen Merkmale abgeleitet und dadurch (pseudo-)wissenschaftlich legitimiert (Müller 2003, Kap. 1; Schiebinger 1993). In den Augen der damaligen Anatomen und Phrenologen, die die Gehirne, Skelette, Hautfarben und Haare der Geschlechter und „Rassen“ akribisch vermaßen und verglichen, äußerte sich ihre Andersheit nicht nur in einer Differenz kognitiver und psychischer Merkmale, sondern erstreckte sich bis in die kleinste Faser ihres Körpers.

Ähnlich kontrovers verlief auch die Geschichte des *wissenschaftlichen Vergleichs* (umfassend Eggers 2016). Die Kontroverse glich jener, die ich oben dargestellt habe. Vor allem in der Romantik und im Historismus formierten sich vergleichsskeptische Positionen, die dem Vergleich und der mit ihm einhergehenden kategorialen Vereinheitlichung die Singularität des Geschehens und die Notwendigkeit entgegenstellten, Phänomene in ihrer Ganzheit und Besonderheit zu erfassen: Jeder Mensch,

jedes Ding und jede Kultur ist anders als alle anderen. Aus dieser Sicht bot sich die Analogie als die geeignetere Methode an: Über das Denken in Analogien ließen sich Phänomene zueinander in Beziehung setzen, ohne ihre Eigenart zu zerstören.

Ein prominenter Vertreter dieser Auffassung war Johann Gottfried Herder. Seine Kritik am Kulturvergleich richtete sich nicht nur gegen den Vergleich an sich, sondern vor allem auch gegen das Vergleichen im Modus von Besser/schlechter-Vergleichen und die damit verbundene Unterwerfung des Verglichenen unter einen einzigen Bewertungsmaßstab (ausführlicher Eggers 2016, Kap. 6). Wer sollte der „unparteiische Richter“ sein, der darüber entscheiden kann, nach welchen Kriterium Kulturen verglichen werden können, und was wäre „der genauestimmte, der unverfälschbare Maasstab?“ (Johann Gottfried Herder, 1794, zit. in W. Steinmetz 2015, S. 40). Wenn Kulturen und Menschen verschieden sind, wie will man sie dann vergleichen und bewerten, ohne dem einen den Gesichtspunkt des anderen aufzudrängen? Während diese Frage im 18. Jahrhundert noch durchaus selbstkritisch diskutiert wurde und Anderssein nicht notwendig Überlegenheit meinte, setzte sich im 19. Jahrhundert das westliche Modell als alleiniger Vergleichs- und Bewertungsmaßstab durch. Darauf bezieht sich Jürgen Osterhammels schöne Formulierung: „Asien war Europa vergleichbar, solange sich Europa noch nicht für unvergleichlich hielt“ (Osterhammel 1998, S. 378). Von nun an ließen sich die „rückständigen Völker“ an Europa messen und im Rahmen von Stufenmodellen auf einer fiktiven Zeitachse anordnen, auf der sich auch die Vergangenheit Europas widerspiegelte. Die Entwicklungstypologien von Auguste Comte, Emile Durkheim und Ferdinand Tönnies sind ein Widerschein dieser Spiegelung im Anderen.

Herders Kritik am Kulturvergleich markiert den Beginn einer Auseinandersetzung, die den Gesellschaftsvergleich seitdem begleitete. Diese Auseinandersetzungen sind der Hintergrund des Aufsatzes von *George Steinmetz*, mit dem der Band eingeleitet wird. Der Beitrag schlägt eine Brücke zwischen dem außerwissenschaftlichen, dem sozialen Vergleichen und dem Vergleich als wissenschaftlicher Methode und argumentiert, dass die komparative Methodik des 20. Jahrhunderts den zentralen Prämissen des frühen Kulturvergleichs über weite Strecken verhaftet blieb. Wie sehr das wissenschaftliche und außerwissenschaftliche Vergleichen miteinander verwoben sind, zeigt George Steinmetz einerseits am Beispiel der interkulturell vergleichenden Soziologie, die nach dem Zweiten Weltkrieg im Kontext der damaligen Modernisierungstheorie entstanden ist, und andererseits am Beispiel der in den späten 1970er-Jahren aufkommenden qualitativen Methodik des Gesellschaftsvergleichs (Skocpol 1979).

Der wissenschaftliche Kulturvergleich des 19. Jahrhunderts und die ihn begleitenden Ideen von Fortschritt und Evolution waren geprägt durch den imperialen Kontext, in dem sie aufkamen. Auch wenn sich die damaligen Soziologen vom nationalistischen Chauvinismus distanzierten, hielten sie doch am Glauben an ein stetiges Fortschreiten und an der Annahme fest, dass Gesellschaften abgeschottete – „selbstgenügsame“ (Parsons 1966) – Gebilde sind, die unbeeinflusst von den Außenbeziehungen, in die sie eingebettet sind, analysiert und verglichen werden können. Diese Annahmen gingen als latente Prämissen in den modernisierungstheoretischen Gesellschaftsvergleich nach dem Zweiten Weltkrieg ein und sie erklären,

weshalb die international vergleichende Sozialforschung lange Zeit eine US-amerikanische Spezialität blieb und in Europa nur marginal rezipiert wurde.

Für die USA als Land ohne (nennenswerte) koloniale Tradition und als Vorreiterin einer Entwicklungspolitik, die die „Dritte Welt“ auf denselben Modernisierungspfad bringen wollte, den die westlichen Länder schon durchlaufen hatten, war es naheliegend, an die Fortschrittsideen des 19. Jahrhunderts und an die Vorstellung anzuschließen, dass Gesellschaften voneinander unabhängige Einheiten bilden, die sich im Wesentlichen aus sich selbst heraus entwickeln (aufschlussreich Tenbruck 1989). Für die imperialen Länder Europas, allen voran Großbritannien, Frankreich und Belgien, war es dagegen sehr viel weniger selbstverständlich, die Welt in eine Ansammlung autonomer – oder werdender – Nationalstaaten aufzugliedern. Dazu kam, dass sich die Entwicklungspolitik der imperialen Mächte in erster Linie auf die eigenen Kolonien bezogen und die europäischen Soziologen – Pierre Bourdieu ist dafür ein herausragendes Beispiel – dort ihre wissenschaftlichen Erfahrungen sammelten, und nicht wie ihre US-amerikanischen Kollegen in den sogenannten „Entwicklungsländern“ (ausführlicher G. Steinmetz 2020). Diese unterschiedlichen Erfahrungs- und Vergleichshorizonte führten dazu, dass die modernisierungstheoretische Vergleichsmethodik in Europa erst nach der endgültigen Entkolonialisierung breiter rezipiert wurde. Steinmetz' historische Rekonstruktion der Verstrickungen zwischen der wissenschaftlichen Komparatistik und ihren vorwissenschaftlichen Vorläufern führt überzeugend vor, dass man mit einer solchen Analyse nicht nur eine reflektiertere Sicht auf die Methodik der komparativen Forschung gewinnt, sondern auch zu einem genaueren Verständnis der blinden Flecken und regionalen Sonderwege der Soziologie gelangt.

Wie kann sich die komparative Methodologie aus diesen latenten Verstrickungen befreien, ohne zu verdrängen, dass sich der wissenschaftliche Vergleich von den Alltagspraktiken des Vergleichens nie vollständig lösen kann? Zur Beantwortung dieser Frage orientiert sich George Steinmetz an Bourdieus reflexiver Soziologie und verbindet dessen Forderung nach einem zweifachen „epistemologischen Bruch“ mit einer alternativen Vergleichsmethodologie, die sich an den kritischen Realismus von Roy Bhaskar anlehnt. Die Besonderheit seines Beitrags liegt darin, dass er es nicht bei einer methodologischen Kritik der Methodik des interkulturellen Vergleichs bewenden lässt (dazu instruktiv Knöbl 2007, Kap. 5). Vielmehr zeigt er einen methodologischen Weg auf, der – im Sinne von Bourdieus doppeltem epistemologischen Bruch – zu den Prämissen des vorwissenschaftlichen wie auch des konventionellen wissenschaftlichen Gesellschaftsvergleichs Distanz hält und gleichzeitig berücksichtigt, dass wissenschaftliche Vergleiche „Konstruktionen zweiten Grades“ (Schütz) sind, die sich immer auf eine bereits „vorverglichene“ Welt beziehen. Damit positioniert sich George Steinmetz zwischen den verschiedenen Richtungen, in die sich die Auseinandersetzung um die wissenschaftliche Legitimität von Gesellschaftsvergleichen aufgespalten hat. Er wendet sich einerseits gegen das positivistische Methodologieverständnis, das er sowohl dem quantitativen Gesellschaftsvergleich wie auch dem makrokausalen Fallanalysenansatz à la Skocpol vorwirft, aber auch gegen postkoloniale Theorien, die die Legitimität kausaler Analysen grundsätzlich in Abrede stellen (ausführlicher G. Steinmetz 2004). Gegen diese Positionen setzt er eine Methodologie, die es erlaubt, der Singularität und Kontingenz historischer

Ereignisse Rechnung zu tragen, ohne auf das Erklärungspotenzial vergleichender Analysen zu verzichten.

Auch *Willibald Steinmetz* befasst sich mit der Geschichte des Gesellschaftsvergleichs. Sein empirisches Beispiel ist die Entstehung und Veränderung des (quantifizierten) europäischen Staatenvergleichs zwischen dem späten 17. und dem frühen 20. Jahrhundert. Im Gegensatz zum Beitrag von George Steinmetz liegt der Fokus weniger auf der wissenschaftlichen Methodik des Vergleichs als auf dessen politischer Praxis. Heute stoßen wir nahezu täglich auf weltweite Staatenrankings, allein die Weltbank hat Hunderte von Indikatoren entwickelt, nach denen sie die Staaten vergleicht und in eine Rangordnung bringt. Der Beitrag von Willibald Steinmetz zeigt für eine frühe Periode auf, wie voraussetzungsvoll solche Rankings sind und welche historischen Konstellationen dazu führten, dass sich Staaten mittels Zahlen zu vergleichen und zu bewerten begannen.

Um Staaten miteinander vergleichen zu können, braucht es nicht nur Organisationen, die die Autorität haben, Kategorien und Messstandards verbindlich festzulegen. Ebenso entscheidend ist eine kulturelle Voraussetzung, nämlich die Vorstellung, dass „Geschichte“ ein zukunftsöffener Prozess des ständigen Fortschreitens ist. Erst unter dieser Bedingung konnte sich eine „culture of improvement“ (Slack 2015) herausbilden, die ihrerseits die Voraussetzung dafür war, sich selbst und die anderen hinsichtlich ihrer Fort- und Rückschritte zu beobachten. Neu war mit anderen Worten nicht unbedingt der Vergleich an sich, neu war ein Vergleich, der Staaten unter dem Gesichtspunkt eines ständigen Wettbewerbs des Nachholens und Überholens beobachtete. Und schließlich macht der Beitrag von Willibald Steinmetz auch deutlich, dass die Einheiten des Staatenvergleichs nicht Nationalstaaten im heutigen Verständnis waren, sondern ein heterogenes Konglomerat unterschiedlichster und aus heutiger Sicht unvergleichbarer politischer Organisationsformen bildeten: Imperien wurden mit souveränen Staaten, politisch abhängige Gebilde mit Staatenförderationen verglichen. Dies bedeutete, dass Staatlichkeit (im modernen Sinne) nicht gesetzt war, sondern selbst ein Vergleichs- und Bewertungskriterium bildete, wobei sich der souveräne (National-)Staat erst mit der Zeit als Referenzmodell etablierte. Insofern führten die frühen Staatenvergleiche auch zu einer neuen Vorstellung dessen, was ein „Staat“ überhaupt ist und was er zu leisten hat, und stellten damit einen wesentlichen Schritt in der Naturalisierung des Nationalstaates als Grundkategorie der internationalen Statistik dar (Speich Chassé 2020; am Beispiel der frühen UN-Statistik Heintz 2012).

Anhand einer materialreichen Rekonstruktion der Kontroversen, die den Staatenvergleich im 18. und 19. Jahrhundert begleiteten, belegt der Beitrag, dass sich die Vergleichsgrößen über die Zeit hinweg veränderten, und zeigt gleichzeitig auf, wie verwickelt die Geschichte des europäischen Staatenvergleichs verlief. In einer ersten Phase bezog sich der Vergleich primär auf die staatlichen Machtressourcen – auf Bevölkerungszahl, Fläche, Staatsfinanzen und militärische Stärke. Willibald Steinmetz interpretiert diese Zentrierung auf Machtindikatoren als eine Reaktion auf den nahezu permanenten Kriegszustand, in dem sich Europa bis 1815 befand. Mit der Friedensperiode nach dem Wiener Kongress verlagerte sich das Vergleichsinteresse auf Indikatoren, die die Leistungsfähigkeit und den Zivilisationsgrad, also das Prestige, eines Landes anzeigten und insofern auch legitimatorische Funktion

hatten – auf Wirtschaftsgrößen und kulturelle und soziale Errungenschaften (zur Prestigekonkurrenz zwischen Nationalstaaten Werron 2012).

Willibald Steinmetz zeigt aber auch, dass die Art – und die Akzeptanz – des Vergleichens und Bewertens keineswegs linear verliefen. Ähnlich wie im Falle des Personenvergleichs (s. o.) gerieten auch die Staatenvergleiche im Modus quantitativer Besser/schlechter-Vergleiche unter Kritik. Quantifizierenden Staatenvergleichen, die mit den Berechnungen der frühen britischen „Politischen Arithmetik“ in der zweiten Hälfte des 17. Jahrhunderts einsetzten, später von der deutschen Tabellenstatistik weitergeführt und im Zuge der Einrichtung amtlicher statistischer Büros zunehmend professionalisiert und thematisch erweitert wurden, standen schon fast von Beginn an Positionen gegenüber, die sich gegen den numerischen Reduktionismus von Zahlenvergleichen wandten und für einen bloß beschreibenden, qualitativen Staatenvergleich plädierten. Den sogenannten „Tabellenknechten“ warfen sie vor, Disparates und Singuläres unter inhaltsleere abstrakte Kategorien zu subsumieren und messend zu vergleichen, was nicht vergleichbar war. Mit den Weltausstellungen seit Mitte des 19. Jahrhunderts kam ein neues Medium hinzu, über das sich der Staatenvergleich besonders anschaulich darstellen ließ. Anhand der Exponate, die in kunstvoll arrangierten Vergleichsräumen ausgestellt wurden, ließ sich das Modernitäts- und Zivilisationsgefälle zwischen den Ländern in einem wörtlichen Sinne sichtbar machen. Die Weltausstellungen hatten zudem den Vorteil, beide Vergleichsformen gleichzeitig zur Anschauung zu bringen – sowohl die Besser/schlechter-Vergleiche, über die sich die Staaten direkt zueinander in Beziehung setzen und auf einer Skala der wirtschaftlichen und kulturellen Leistungsfähigkeit einordnen ließen, als auch die Einfach-anders-Vergleiche, die jeden Staat in seiner nationalen Eigenart darstellten.

Die beiden Vergleichsformen, deren historische Genese Willibald Steinmetz (2019b) rekonstruiert hat und in seinem Beitrag auf Staatenvergleiche anwendet, bildeten sich im europäischen Kontext des 18. und 19. Jahrhunderts heraus. Was ist aus ihnen geworden? Sind sie auch heute noch relevant und ist ihr Verhältnis nach wie vor ein antagonistisches? Mit diesen Fragen befassen sich *Alex V. Barnard* und *Marion Fourcade*. Ihr Beitrag greift auf Studien zurück, die Marion Fourcade alleine oder in Zusammenarbeit mit Kieran Healy verfasst hat (Fourcade 2016; Fourcade und Healy 2013, 2016), und präsentiert sie in einer kondensierten Form. Fourcade verwendet zwar eine andere Terminologie und setzt auch etwas andere Akzente, aber dennoch lassen sich ihre Arbeiten direkt an Willibald Steinmetz' Befunde zur Geschichte des Vergleichs anschließen. Zusammengelesen beschreiben sie, welchen Gestaltwandel Besser/schlechter- und Einfach-anders-Vergleiche in den letzten 200 Jahren durchgemacht haben.

Aus der Sicht von Barnard und Fourcade bildet sich gegenwärtig eine „ordinale Gesellschaft“ heraus – eine Gesellschaft, in der alles einer ständigen Bewertung unterzogen und jeder und jedes mithilfe digitaler Technologien verglichen, gerankt und kategorisiert wird. Im Mittelpunkt ihres Beitrags steht die Frage, wie sich diese zunehmende „Ordinalisierung“ zu einer Einteilung verhält, die Menschen nominal nach ihrer kategorialen Zugehörigkeit klassifiziert. Was verändert sich, wenn eine Ordnung, die Menschen in erster Linie nach ihren sozialen Merkmalen einstuft (ihrem Geschlecht, ihrer Nationalität, ihrer Hautfarbe usw.), durch eine Ordnung abgelöst wird, die sie unabhängig von ihren kategorialen Zugehörigkeiten in eine

individualisierte und zunehmend quantifizierte Besser/schlechter-Relation bringt? Lösen sich kategoriale Differenzen auf oder schleichen sie sich in anderer Form und auf anderem Wege wieder ein? Die Parallelen zur Fragestellung von Willibald Steinmetz sind unschwer zu erkennen. Es geht im Kern um das Verhältnis zwischen (nominalen) *Simply-different*- und (ordinalen) *Better/worse*-Vergleichen oder in der Terminologie von Marion Fourcade: um das Verhältnis zwischen „judgments of kind“ and „judgments of worth“ (Fourcade 2016, S. 179).

Barnard und Fourcade unterscheiden drei Varianten von Vergleichen, die sie im Anschluss an die Theorie der Messniveaus als *nominal*, *kardinal* und *ordinal* bezeichnen. Sie sprechen zwar von Klassifikationen und nicht von Vergleichen, der Sache nach handelt es sich aber um Vergleiche. *Nominale* Klassifikationen sortieren die Welt in „dieses“ und „anderes“ („judgments of kind“). Ähnlich wie Willibald Steinmetz' Einfach-anders-Vergleiche beruht auch die Feststellung kategorialer Verschiedenheit auf einem Vergleich: Frauen können nur dann als von Männern verschieden eingestuft werden, wenn sie vorher mit ihnen verglichen wurden. In ihrer Reichweite erstrecken sich nominale Vergleiche von der Behauptung eines bloßen Andersseins bis hin zum expliziten Anspruch auf Unvergleichbarkeit: „nominality veers toward uniqueness“ (Fourcade 2016, S. 188). *Kardinale* Klassifikationen vergleichen Sachverhalte ausschließlich nach ihren Größenverhältnissen. Facebook zählt die Anzahl der Likes und leitet daraus Popularitätsscores ab, Staaten zählen ihre Universitäten und Soldaten und vergleichen ihre Anzahl mit jener anderer Staaten. Die Besonderheit kardinaler Vergleiche liegt darin, dass sie von qualitativen Gesichtspunkten abstrahieren – nur die Zahl zählt. *Ordinale* Klassifikationen sind Vergleiche, die die verglichenen Einheiten in ein Besser/schlechter-Verhältnis bringen: Etwas ist besser als etwas anderes („judgments of worth“). Im Unterschied zu nominalen Vergleichen, die das Anderssein in den Vordergrund stellen, unterstellen ordinale Vergleiche Äquivalenz.

Was nominale Vergleiche ausmacht und worin sie sich von ordinalen Vergleichen unterscheiden, veranschaulichen Barnard und Fourcade am Beispiel des französischen *Terroir*-Prinzips (ausführlich Fourcade 2012). In Frankreich werden die einzelnen Weine nicht in eine gradualisierte Rangfolge gebracht, wie es bei Punktesystemen der Fall ist, sondern nach Anbaugebiet in Qualitätskategorien eingeteilt. Da jedes *Terroir* durch seine je besonderen Anbaubedingungen charakterisiert ist, gelten die Weine aus unterschiedlichen *Terroirs* als unvergleichbar – als „simply different“. Es gibt mit anderen Worten keine gemeinsame Metrik, um die Weine aus unterschiedlichen Anbaugebiete auf einer einheitlichen Skala miteinander zu vergleichen und keine Möglichkeit, mit einem Wein aus dem Languedoc in die Liga der Grand-Cru-Weine aus dem Médoc aufzusteigen. Das Punktesystem der Weinklassifikation folgt dagegen der Logik von Besser/schlechter-Vergleichen. Parker-Punkte kümmern sich nicht um das *Terroir* oder das soziale Standing der Weinproduzenten, sondern beziehen potenziell alle Weine mit ein und bringen sie in eine quasinumnerische Rangfolge, die beansprucht, die individuelle Leistung eines Pro-

duzenten abzubilden.⁵ Ein Kleinbauer aus der Toskana hat im Prinzip die gleichen Chancen wie eine Gutsbesitzerin aus dem Médoc, 100 Parker-Punkte zu erhalten, und ein Grand Cru aus dem Burgund ist ebenso abstiegsgefährdet wie ein Douro aus Portugal.

Ähnlich wie für Willibald Steinmetz steht auch für Barnard und Fourcade das konfliktive Verhältnis zwischen (nominalen) Einfach-anders- und (ordinalen) Besser/schlechter-Vergleichen im Mittelpunkt – oder in ihrer Terminologie: das Verhältnis zwischen nominalen und ordinalen Klassifikationen. Fourcade (2016) verortet diesen Konflikt in einem Dilemma, mit dem moderne Gesellschaften konfrontiert sind. Auf der einen Seite sehen sie sich dazu gedrängt, kulturelle Differenzen anzuerkennen, formalisiert im Ideal der „multikulturellen Gesellschaft“ und einer Politik, die benachteiligten Personengruppen besondere Rechte einräumt. Auf der anderen Seite erfordert das Selbstbild einer sich als meritokratisch verstehenden Gesellschaft, dass Individuen ausschließlich aufgrund ihrer Leistungen beobachtet und beurteilt werden, gewissermaßen „ohne Ansehen der Person“ (Weber). Eine „gute“ Gesellschaft ist eine Gesellschaft, die anerkennt, dass Menschen und Kollektive „simply different“ sind. Aber eine „gute“ Gesellschaft ist auch eine Gesellschaft, die bei der Verteilung von Anerkennung von personalen Merkmalen absieht und unterstellt, dass alle Menschen gleich sind und entsprechend auch nach gleichen Maßstäben beurteilt werden müssen – als besser oder als schlechter. Beide Ansprüche lassen sich nicht unter einen Hut bringen, sondern sabotieren sich gegenseitig. Die Forderung, im Namen der Gleichberechtigung bestimmten Personenkategorien besondere Rechte einzuräumen, unterläuft das individualistische Gleichheitsprinzip, und der Anspruch, bei der Leistungsbeurteilung von persönlichen Merkmalen abzusehen, ignoriert die faktisch bestehenden Differenzen.

Das antagonistische Verhältnis zwischen diesen beiden Grundprinzipien ist der Ausgangspunkt, von dem aus Barnard und Fourcade ihre Argumentation entwickeln. Ihre These ist die, dass die Ordinalisierung der Gesellschaft kategoriale Zuschreibungen unterminiert, ähnlich wie die Weinbewertung nach Punkten dem französischen Terroir-Prinzip den Rang abläuft, und an seine Stelle ein Ordnungssystem setzt, das jede und jeden denselben Beurteilungsmaßstäben unterwirft. Das ist aber nur die eine Seite. Die andere Seite ist die, dass kategoriale Unterschiede unter der Hand wiedereingeführt werden. Diese These illustrieren Barnard und Fourcade am Beispiel des Wandels der Kriterien für die Kreditvergabe (ausführlicher Fourcade und Healy 2013, 2016). Bis in die 1970er-Jahre orientierten sich Kreditgeber an kategorialen Merkmalen, um die Bonität möglicher Kreditnehmer zu beurteilen. Je nach Geschlecht, Ethnie, Wohnort und Familienstand wurden Kredite vergeben oder auch nicht. Seitdem hat sich ein grundlegender Wandel der Kreditvergabe vollzogen. Grundlage sind nun nicht mehr kategoriale Merkmale, sondern von Drittorganisationen (Fico in den USA, Schufa in Deutschland) gesammelte Daten über Kontostände oder ausstehende Zahlungen, aus denen Algorithmen die jeweilige Kreditwürdigkeit berechnen. Der für jeden Einzelnen erstellte Kreditscore wird ständig angepasst,

⁵ In der markanten Formulierung von Parker: „What I’ve brought is a democratic view. I don’t give a shit that your family goes back to pre-Revolution and you’ve got more wealth than I could imagine. If this wine’s no good, I’m gonna say so“ (zit. in Fourcade 2012, S. 538).

je nachdem, welche Käufe getätigt wurden oder welche Kredite noch offen sind. Das Ergebnis ist ein inklusives und hoch individualisiertes System, das alle Individuen auf einer Skala finanzieller Rechtschaffenheit abbildet und die Ursachen unterschiedlicher Kreditkonditionen dem Einzelnen zurechnet (zu wenig sparsam, zu wenig fleißig), und nicht mehr mit sozialen Merkmalen in Verbindung bringt.

Fourcade und Healy (2013) sprechen in diesem Zusammenhang und unter Bezugnahme auf Max Webers Begriff der Klassenlage von „classification situations“. Mithilfe von algorithmischen Scoringtechniken werden Individuen je einzeln in unterschiedliche Klassifikationslagen sortiert, die ähnlich wie konventionelle Klassenlagen mit ungleichen sozialen Chancen assoziiert sind. Beispielhaft dafür sind Kredit- und Popularitätsscores oder Scores, die das Gesundheitsverhalten oder die Delinquenzwahrscheinlichkeit erfassen. Das Neue liegt darin, dass sich diese Sortierung individualisiert vollzieht – die Kompaktheit von Klassen wird in individuelle Positionen auf einer Vielzahl von ordinalen Skalen aufgelöst. Dies hat zur Folge, dass die strukturellen Ursachen der Ungleichheit weitgehend unsichtbar bleiben und deshalb auch nur schlecht politisierbar sind. Barnard und Fourcade sprechen deshalb in Anlehnung an Erving Goffmans (1975) Begriff der „Scheinnormalität“ von einer „Schein-Entkategorisierung“. Die ordinale Gesellschaft tritt zwar mit dem Anspruch auf, von persönlichen Merkmalen abzusehen und alle Individuen „ohne Ansehen der Person“ nach den gleichen, „fairen“ Maßstäben zu bewerten. Damit werden kategoriale Unterschiede jedoch nicht zum Verschwinden gebracht, sondern nur anders erzeugt, nun aber eingebettet in einen individualisierenden Moraldiskurs, der eine strukturelle Interpretation erschwert.

Setzt man die Texte von Willibald Steinmetz und Barnard und Fourcade zueinander in Beziehung und verortet sie auf einer historischen Dimension, dann lässt sich daraus einiges über den Gestaltwandel des Verhältnisses von Bewertung, Vergleich, Kategorisierung und Quantifizierung lernen und über die Rolle, die digitale Technologien dabei spielen. Denn was Barnard und Fourcade als Charakteristikum der ordinalen Gesellschaft begreifen – die immer feinkörnigere Beobachtung von Individuen auf der Basis dessen, was sie *tun*, und nicht mehr anhand dessen, was sie sozial gesehen *sind* – wurde durch die Digitalisierung wesentlich vorangetrieben. Digitale Technologien radikalisieren, was in den Besser/schlechter-Vergleichen des 19. Jahrhunderts angelegt, wenn auch nicht durchgängig verwirklicht war: den Vergleich von Personen nicht mehr nach ihrem sozialen Stand, sondern anhand ihrer individuellen Qualitäten. Die heutigen Besser/schlechter-Vergleiche beziehen alle ein und verfügen mit den digitalen Technologien über eine technische Infrastruktur, die es erlaubt, auf der Basis von Verhaltensdaten die jeweiligen Rangpositionen ständig neu zu berechnen. Der kompetitive und temporalisierte Charakter von Besser/schlechter-Vergleichen, den Willibald Steinmetz in Ansätzen bereits für das 19. Jahrhundert konstatiert hat, wird in der „ordinalen Gesellschaft“ auf die Spitze getrieben mit der Folge, dass Unterschiede immer mehr der Leistung oder dem Versagen des Einzelnen zugeschrieben werden.

Bettina Heintz entwickelt Barnard und Fourcades These eines Zusammenhangs von Digitalisierung und Ordinalisierung weiter und führt dazu das Konzept des Beobachtungsformats ein. Der Beitrag vergleicht digitale Technologien mit der Statistik und untersucht, worin sich ihre Beobachtungstechniken unterscheiden. Der Vergleich

mit einem Beobachtungsformat, das 200 Jahre früher entstanden ist, erlaubt es, präziser zu bestimmen, worin genau die Affinität zwischen digitalen Technologien und der von Barnard und Fourcade konstatierten Ordinalisierung liegt. Das heißt anstatt digitale Technologien als etwas grundsätzlich Neues zu begreifen, verortet Bettina Heintz sie in der Geschichte sozialer Beobachtungsformate, die mit der Entstehung der modernen Statistik begann (vgl. Abschn. 1). Wie die amtliche Statistik zielen auch Suchmaschinen, soziale Netzwerke, Selbstvermessungsgeräte und Shoppingportale darauf ab, in der Masse der aufgezeichneten Daten Muster zu entdecken und aus diesen Mustern Erwartungen über zukünftige Entwicklungen abzuleiten. Die Art und Weise, wie sie das tun, unterscheidet sich jedoch. Für ihren Beitrag greift sie zwei Beispiele heraus – die Bevölkerungsstatistik als historisch erstes modernes Beobachtungsformat und personalisierte Recommendersysteme als prototypischer Fall digitaler Beobachtungsinstrumente – und vergleicht sie hinsichtlich ihrer Beobachtungsverfahren: Wie werden die Zahlen fabriziert, mit deren Hilfe die amtliche Statistik die Gesellschaft beobachtet, und welche Beobachtungsverfahren setzen Recommendersysteme ein, um personalisierte Empfehlungen zu erzeugen, und welche Rolle spielen dabei Vergleich, Bewertung, Kategorisierung und Quantifizierung?

Beobachtungsformate beruhen auf impliziten Theorien des Sozialen, d. h. auf Vorstellungen darüber, was sich zu zählen und zu beobachten lohnt, und wie das Soziale, das sie zu erfassen suchen, beschaffen und strukturiert ist. Sie entstehen nicht in einem luftleeren Raum, sondern sind Ausdruck der jeweiligen Gesellschaft, in der sie entwickelt wurden, und dem, was in dieser Gesellschaft denk- und sagbar ist. Verortet man die Bevölkerungsstatistik und Recommendersysteme auf einer historischen Zeitachse, macht der Vergleich nicht nur sichtbar, wie die Bevölkerungsstatistik und Recommendersysteme beobachten und worin sich ihre Beobachtungstechnik unterscheidet, sondern er gibt auch Aufschluss darüber, wie sich die Prämissen sozialer Beobachtung in den letzten 200 Jahren verändert haben. Die Pointe des Beitrags von Bettina Heintz liegt darin, dass sie im Detail rekonstruiert, wie digitale Technologien und die amtliche Statistik zur Aufdeckung von Regelmäßigkeiten gelangen. Dieses Herunterbuchstabieren auf die Ebene der konkreten Operationsweise macht Unterschiede sichtbar, die auf einen grundlegenden Wandel ihrer Sozialontologie verweisen. Digitale Beobachtungstechnologien ersetzen die amtliche Statistik nicht, aber sie machen deutlich, dass es heute noch andere Weisen der Sozialbeobachtung gibt, die der gegenwärtigen Gesellschaft offenbar angepasster sind als die Beobachtung durch die Linsen der Statistik.

3 Die vermessene Welt. Kategorisieren, Vergleichen und Bewerten in der inter/nationalen Statistik und in der Umweltdebatte

Die Entstehung der modernen Statistik wird üblicherweise auf die erste Hälfte des 19. Jahrhunderts datiert und mit dem Aufbau statistischer Ämter in Zusammenhang gebracht. Wie Lars Behrisch (2016) in seiner Studie zu den Anfängen der modernen Statistik im letzten Drittel des 18. Jahrhunderts zeigt, ist der entscheidende Wendepunkt aber nicht organisatorischer, sondern *epistemologischer* Natur: Die maßgebliche Zäsur war nicht die Einrichtung statistischer Ämter, sondern ein grundlegend

neues Verständnis der Aussagekraft und des Potenzials von Zahlen. Den modernen Statistikern ging es nicht mehr um bloßes Zählen, sondern um die Aufdeckung übergreifender sozialer Zusammenhänge, die sie aus der Masse der gesammelten Daten herauszulesen suchten. Im Unterschied zu den protostatistischen Erhebungen der Vormoderne waren sie nicht mehr an konkreten Einzelinformationen interessiert, sondern an systemischen Größen und ihren Interrelationen – an Altersverteilungen, Ernährungsgewohnheiten, Kriminalitätsraten und vielen anderen Größen mehr –, die sich vergleichen und aus denen sich politische Maßnahmen ableiten ließen.

Michel Foucault (2004) bringt die Entstehung der modernen Statistik in Zusammenhang mit einer neuen Technik des Regierens, die er als Biopolitik bezeichnet und von den Techniken der Disziplinierung abgrenzt. Im Gegensatz zur Disziplinarmacht, die auf die Abrichtung und Kontrolle des Individualkörpers ausgerichtet war, wirkt Biopolitik indirekt über die Vermessung, Erklärung und Regulierung der Prozesse, die auf der Ebene der Bevölkerung als kollektive Phänomene in Erscheinung treten. „Bevölkerung“ gab es als Begriff und Sachverhalt natürlich schon vorher. Neu war aber, dass Bevölkerung nicht mehr als Summe der Individuen verstanden wurde, die in einem Territorium leben, sondern als eigenständiges, emergentes Objekt, das nach eigenen, gewissermaßen „natürlichen“ Gesetzmäßigkeiten funktioniert (Foucault 2004, S. 107 ff.). Diese Gesetzmäßigkeiten aufzudecken, ihre Randbedingungen und Veränderungen systematisch zu beobachten, war die Aufgabe der modernen Statistik, die im letzten Drittel des 18. Jahrhunderts entstand und durch die Gründung statistischer Ämter in der ersten Hälfte des 19. Jahrhunderts verstetigt wurde.

Mit der Einrichtung nationaler statistischer Ämter erhielt das biopolitische Projekt eine organisatorische Infrastruktur. Die statistische Beobachtung wurde zentralisiert, national vereinheitlicht und auf Dauer gestellt. Die Institutionalisierung der amtlichen Statistik führte zu einem enormen Anstieg der verfügbaren Daten, die das Objekt – die „Bevölkerung“ –, das sie zu messen beanspruchten, in gewissem Sinne erst erzeugten und in Gestalt von Tabellen sichtbar machten (Ruppert 2011). Für Ian Hacking (1991) war die Verfügbarkeit von immer mehr Zahlen – „the avalanche of printed numbers“ (Hacking 1982) – ein wesentlicher Anstoß dafür, statistische Aggregate nicht mehr als bloße Aufsummierung individueller Geschehnisse zu interpretieren, sondern als Ausdruck sozialer Gesetzmäßigkeiten. Statistik als Beobachtungsformat meint also nicht die Anhäufung numerischer Fakten, sondern die Entdeckung von Ordnungsmustern in diesen Fakten (ausführlicher Heintz).

Die Menschen, die die Bevölkerung ausmachen und mit den Instrumenten der Statistik beobachtet werden, sind nicht mehr das „Volk“ und auch keine Untertanen, die dem Souverän in einem Verhältnis des Gehorsams gegenüberstehen, sondern Individuen, die ihr gemeinsames Menschsein teilen (Foucault 2004, S. 107 ff.). Nur unter dieser Bedingung, also Georg Simmels „Individualismus der Gleichheit“ (vgl. Abschn. 2), konnte eine Beobachtungsapparatur entwickelt werden, die alle Bürger und Bürgerinnen eines Landes unabhängig von ihren sozialen, regionalen und konfessionellen Unterschieden erfasste und miteinander verglich. „It makes no sense to count people if their common personhood is not seen as somehow more significant

than their differences“ (Porter 1986, S. 25).⁶ Damit verbunden ist eine zweite Voraussetzung, nämlich die Unterscheidung zwischen der eigenen, nationalen Bevölkerung und allen anderen Völkern. Am Ursprung der nationalen Statistik liegt mit anderen Worten ein primärer Klassifikationsakt, der jene, die als Teil der Bevölkerung zählen (und folglich auch gezählt werden), abgrenzt von jenen, die nicht dazu gehören. Mit dieser Unterscheidung erzeugt der Staat einen nationalen „Äquivalenzraum“ (Desrosières 2005, S. 20f.), der entlang politisch-territorialer Grenzen verläuft und der, wie die Beiträge von Léa Renard, Theresa Wobbe sowie Hannah Bennani und Marion Müller zeigen, ausgeweitet werden musste, um Menschen weltweit zählen, einteilen und vergleichen zu können. Denn als statistisch vergleichbar galten damals nur jene Personen, die als Teil der Bevölkerung und damit der nationalen Gesellschaft angesehen wurden. Mit diesem Zusammenhang zwischen Bevölkerung, Vergleichbarkeit und Statistik befasst sich am Beispiel der deutschen Kolonialstatistik der Aufsatz von Léa Renard.

Der Beitrag untersucht anhand einer detaillierten Analyse der sogenannten „Denkschriften über die deutschen Schutzgebiete“, die vom Deutschen Reichskolonialamt zwischen 1885 und 1914 herausgegeben wurden, wie das neue Instrument der statistischen Beobachtung in den Kolonien eingesetzt wurde, und kommt zum Resultat, dass dies nur sehr punktuell geschah. Dieser Befund widerspricht der gängigen Auffassung, dass die Statistik auch in den Kolonien als Herrschaftsinstrument eingesetzt wurde (etwa Christopher 2008; Leonhard und von Hirschhausen 2011, Kap. 3; Touchelay 2019). Da die Imperien in besonderem Maße mit dem Problem des „governing at a distance“ konfrontiert waren, ist es an sich naheliegend anzunehmen, dass sie über ihre kolonialisierten Bevölkerungen möglichst viel in Erfahrung bringen wollten. Die Herrschaft über kulturell fremde und geografisch weit entfernte Gebiete erforderte ein exaktes Wissen darüber, wie viele Menschen in den Kolonien leben, wie sie sozial gegliedert sind und inwieweit sie sich als Arbeitskräfte und Steuerzahler einsetzen lassen. Dass diese Informationen im Deutschen Kaiserreich kaum und auch von anderen Imperien nur vereinzelt erhoben wurden,⁷ hatte, wie der Aufsatz von Léa Renard belegt, nicht nur organisatorische und infrastrukturelle Gründe. Das Haupthindernis war das imperiale Differenzmodell, d. h. die Annahme einer grundsätzlichen Unvergleichbarkeit zwischen Mutterland und Kolonien und zwischen der eigenen und der kolonialen Bevölkerung.

Diese Separierung setzte sich auch in den statistischen Erhebungen fort. Die Prämisse, dass die Kolonisierten und die Angehörigen des Zentrums prinzipiell unvergleichbar sind, führte dazu, dass die beiden Gruppen anders gezählt und kategorisiert wurden. Léa Renard führt dafür den treffenden Begriff der „methodischen Alterität“ ein. Während die deutschen Kolonisatoren und die in den Kolonien lebenden „Europäer“ als Einzelpersonen gezählt wurden, wurden die Einheimischen nur über eine Auszählung ihrer „Hütten“ erfasst, so wie es vor der Einführung der

⁶ Wie voraussetzungsvoll diese Gleichheitsannahme war und wie widersprüchlich sie gehandhabt wurde, zeigt die Geschichte der Berufsstatistik im 19. Jahrhundert, bei der es mehrere Jahrzehnte dauerte, bis ständische Klassifikationskriterien vollständig verschwunden waren (Vanderstraeten 2006).

⁷ Zu Frankreich Renard (2021); zum englischen Imperium mit besonderen Blick auf Indien Christopher (2008) und von Hirschhausen (2011).

Individualzählung auch in Europa vielerorts praktiziert wurde (Rusnock 1995). Die sogenannten „Eingeborenen“ traten mit anderen Worten nicht als Individuen in Erscheinung, sondern als eine amorphe Masse, deren Größe bloß geschätzt wurde. Die Ergebnisse dieser beiden Zählungen wurden nicht miteinander verrechnet, sondern in separaten Tabellen dargestellt. Es wurde also nie eine Zahl publiziert, die sich auf die gesamte unter deutscher Herrschaft lebende Bevölkerung bezog. Eine Gesamtzahl gab es nur für die sogenannten „Weißen“, die in den Kolonien lebten, und zwar unabhängig von ihrer Staatsbürgerschaft. Die primäre Trennlinie verlief damit nicht zwischen den Deutschen und den in den Kolonien lebenden „Ausländern“, sondern zwischen den Weißen und den „Anderen“ – den Kolonialiserten. Es ist offensichtlich, dass dieses Klassifikationsschema nicht nominal und symmetrisch war, sondern als eine „hierarchische Opposition“ (Dumont 1980) auftrat und insofern eine stark ordinale Dimension mit sich führte. Die „Weißen“ repräsentierten das Allgemeine – die „Menschen“ – und bildeten gleichzeitig den Gegenpol zu den „Eingeborenen“. Die „Eingeborenen“ verkörperten gegenüber dem Allgemeinen das Andere, und das Andere war das Gegenteil der „Weißen“.

Zusammengenommen belegt der Aufsatz von Léa Renard, wie tief die Annahme einer Inkommensurabilität von Zentrum und Peripherie, Weißen und Einheimischen im kolonialen Denken verankert war. Kolonisierte und Kolonisatoren wurden als unterschiedliche Menschensorten angesehen, für die es keine übergeordnete Kategorie gab. Deshalb ließen sie sich statistisch auch nicht miteinander vergleichen und deshalb fiel es auch schwer, die auf die deutsche Bevölkerung bezogenen statistischen Verfahren auf die kolonisierte Bevölkerung anzuwenden. Der Beitrag ist deshalb auch ein wichtiges Korrektiv zum methodologischen Nationalismus (George Steinmetz), der auch viele Arbeiten zur Geschichte der Statistik durchzieht. Er verweist darauf, dass die Entstehung der amtlichen Statistik in eine imperial strukturierte Welt eingebettet ist und der Foucaultsche Begriff der „Bevölkerung“ angesichts der imperialen Verfasstheit der damaligen europäischen Gesellschaften zu spezifizieren ist (Stoler 1995; Cordell et al. 2010). Denn zur „Bevölkerung“ gehörte nur ein kleiner Teil der Menschen im Herrschaftsbereich der Imperien.

Schon Adolphe Quetelet träumte von einer Statistik, die „das gesamte Menschengeschlecht“ vermisst (Quetelet 1921[1869], S. 151 f.). Aber eine solche „Weltstatistik“ lag im Zeitalter der imperialen „Politik der Differenz“ (Burbank und Cooper 2010) noch in weiter Ferne. Sie konnte erst realisiert werden, als sich die menschenrechtliche Idee der Gleichheit aller Menschen und aller Völker gegen das imperiale Differenzmodell durchgesetzt hatte (Heintz 2012). Die lange Geschichte der modernen Statistik von ihren Anfängen Ende des 18. Jahrhunderts hin zur internationalen Statistik belegt, welche Schwellen überwunden werden mussten, bis weltumspannende quantitative Vergleiche auf Dauer gestellt werden konnten. Eine erste und entscheidende Schwelle ist *kultureller Art*. Ähnlich wie eine nationale Statistik erst entstehen konnte, als die Gemeinsamkeiten zwischen den Bewohnern eines Landes als relevanter erachtet wurden als ihr sozialer Stand, konnte sich auch die internationale Statistik erst mit der Institutionalisierung der Menschenrechte nach 1945 konsolidieren. Erst als alle Menschen, und nicht nur die Bürger eines Landes, als prinzipiell gleich erachtet wurden, wurden sie weltweit vergleichbar. Eine zweite Voraussetzung sind *Zentralisierung und Organisationsbildung*. Flächendeckende

Bevölkerungserhebungen erfordern einen zentralisierten Staat mit einem funktionierenden Behördenapparat und eine zentrale Agentur, die verbindlich Kategorien und Messstandards festsetzen kann (W. Steinmetz). Obschon es bereits im Völkerbund und in der International Labor Organization (ILO) erste Ansätze zum Aufbau einer internationalen Statistik gegeben hatte (Nichols 1942; Wobbe und Renard 2017), konnte eine weltweite statistische Beobachtung erst nach dem Zweiten Weltkrieg auf Dauer gestellt werden (Ward 2004).

Der Aufbau einer Statistik erfordert drittens *Standardisierung und Vereinheitlichung*. Um die Bevölkerung quantitativ zu erfassen, muss die Gesellschaft nach Maßgabe ihrer Quantifizierbarkeit umgestaltet werden. „Society must be remade before it can be the object of quantification“ (Porter 1992, S. 201). Menschen und Dinge müssen unter Vernachlässigung ihrer individuellen Qualitäten kategorial vereinheitlicht werden, Maße müssen ineinander übersetzbar sein, Arbeit und Produkte müssen aus dem lokalen Kontext herausgelöst und über ein gemeinsames Maß vergleichbar gemacht werden. Die Geschichte der nationalen Statistik zeigt, welche immensen sozialen Eingriffe dazu erforderlich waren. Im Falle weltweiter Statistiken, die mit einer sehr viel größeren Vielfalt konfrontiert sind und wo jedes Land seine eigenen Erhebungsverfahren und Klassifikationsschemata hat, stellt sich dieses Problem noch sehr viel schärfer. Viertens setzen internationale Statistiken universelle Vergleichskriterien voraus, die überall und auf alle anwendbar sind. Die Schwierigkeit, der Heterogenität der Verhältnisse gerecht zu werden, zeigt sich bereits bei so augenscheinlich einfachen Vergleichskriterien wie der Erwerbstätigkeit oder der Familiengröße. Menschen gehen zwar überall einer Beschäftigung nach, aber nicht überall ist diese Beschäftigung marktvermittelt oder als Beschäftigung registriert. Und überall werden Familien gegründet, aber nicht überall werden die Grenzen kulturell gleich gezogen. Wie geht die internationale Statistik mit diesem Problem um?

Damit befasst sich *Theresa Wobbe* am Beispiel der Erwerbstätigkeit. Ihr Beitrag rekonstruiert die Geschichte der statistischen Unterscheidung von Erwerbs- und Hausarbeit und die ihr zugrundeliegende Trennung von Markt und Haushalt in verschiedenen historischen Phasen: bei der Einführung der Kategorie der Erwerbstätigkeit in die deutsche Berufsstatistik Ende des 19. Jahrhunderts und anhand ihrer Übernahme in die internationale Beschäftigungsstatistik der ILO in der Zwischenkriegszeit sowie nach 1945. Dieser lange Zeitraum und der Wechsel der Betrachtungsebene von der nationalen zur internationalen Statistik erlauben es ihr, zu untersuchen, wie eine in der nationalen Statistik eingeführte Unterscheidung in der internationalen Statistik nachwirkt.

Die Trennung zwischen Hausarbeit und Erwerbsarbeit ist für uns heute selbstverständlich. Der Beitrag geht der Frage nach, wie es dazu kam und welche Rolle die nationale und internationale Statistik bei ihrer Objektivierung und Institutionalisierung gespielt hat. Dazu führt Theresa Wobbe den Begriff des „inter/nationalen statistischen Feldes“ ein und kritisiert an den bestehenden Studien, dass die nationale und die internationale Statistik in der Regel separiert voneinander untersucht werden. Demgegenüber begreift sie nationale und internationale Statistiken als eng miteinander verwobene Beobachtungsebenen, die in ihren Zusammenhängen untersucht werden müssen. In ihrem Beispielsfall äußert sich diese Verwobenheit darin,

dass die ILO-Statistik direkt an die Klassifikation der nationalen Statistik anknüpfte und die dort formalisierte Trennung von Erwerbs- und Hausarbeit übernahm und weiter verfestigte. Die Unterscheidung von Markt und Haushalt wurde damit als latentes Beobachtungsschema in die internationale Statistik transferiert und nicht mehr weiter reflektiert.

Um diese Transferprozesse zu erhellen, geht sie zunächst auf den deutschen Berufszensus ein, der die Kategorie der Erwerbstätigkeit 1882 erstmals einführt und sie über die Kriterien „marktbezogen“, „regulär“ und „bezahlt“ von Nichterwerbstätigkeit abgrenzte. Arbeit ist von diesem Zeitpunkt an bezahlte und marktvermittelte Arbeit. Über die Kategorie der Erwerbstätigkeit wird die Bevölkerung in einen „aktiven“ und einen „nichtaktiven“ Teil aufgeteilt mit der Folge, dass unbezahlte Tätigkeiten, allen voran die Hausarbeit, mit Abhängigkeit und Unselbstständigkeit assoziiert und abgewertet werden. Wie sich die Einführung dieser Kategorie vollzog und welche Prämissen ihr zugrunde lagen, beschreibt Theresa Wobbe im ersten empirischen Kapitel. In einem zweiten Schritt zeigt sie am Beispiel der ILO, dass die in der nationalen Statistik verwendete Kategorie der Erwerbstätigkeit und die sie begleitende Invisibilisierung von Hausarbeit als Arbeit nahezu nahtlos übernommen und dadurch weiter stabilisiert wurde. Die Unterscheidung Markt vs. Haushalt schrieb sich mit anderen Worten schon von Beginn an in die internationale Statistik ein und wurde zu einer zentralen Beobachtungskategorie des internationalen Staatenvergleichs. Erst mit der Zeit entwickelte die ILO eine eigene Definition von Erwerbstätigkeit, die den verschiedenen nationalen Klassifikationspraktiken Rechnung trug, aber auch den Anforderungen eines internationalen Vergleichs genügen sollte. Mit dem Aufkommen des Konzepts der „Labour Force“ nach dem Zweiten Weltkrieg veränderte sich das Verständnis von Erwerbstätigkeit insofern, als sich der Fokus nun auf die verfügbaren Arbeitskraftressourcen verschob. Damit änderte sich auch der Blick auf die Hausarbeit, indem Hausfrauen neu auch unter dem Aspekt ihrer möglichen zukünftigen Erwerbstätigkeit beobachtet wurden. Bezeichnend ist aber, dass Hausarbeit nach wie vor als nichtproduktive Arbeit betrachtet wurde und sich die Anpassungen nur auf die Kategorie der Erwerbstätigkeit bezogen.

Der Beitrag macht auf drei Dinge aufmerksam. Erstens auf die Pfadabhängigkeit und Eigendynamik statistischer Beobachtungskategorien. Um zeitliche Vergleichbarkeit zu ermöglichen, gehen frühere Klassifikationsentscheidungen als nicht weiter reflektierte Vorentscheidungen in spätere Statistiken ein, auch wenn sich der Beobachtungsraum gewandelt hat. Zweitens zeigt Theresa Wobbe anhand einer dichten und materialreichen Analyse auf, wie und über welche Überlegungen und Entscheidungen die Trennung zwischen Markt und Haushalt und die damit verbundene Abwertung der Hausarbeit in die nationale Statistik eingeführt und später von der internationalen Statistik übernommen wurde. Soziale Wandlungsprozesse und die Ausweitung des Beobachtungshorizonts haben zwar zu einer Differenzierung der Kategorie der Erwerbstätigkeit geführt, die Grundunterscheidung zwischen Erwerbstätigkeit und Hausarbeit blieb aber intakt. Und schließlich macht der Beitrag auf die Verflochtenheit der nationalen und internationalen Statistik und auf die Notwendigkeit aufmerksam, sie in ihren Zusammenhängen, als inter/nationales statistisches Feld, zu untersuchen.

Angesichts der Tatsache, dass der Beobachtungsraum der ILO auch Länder einschloss, in denen viele Beschäftigte nicht in einem formellen Arbeitsverhältnis standen und Marktwirtschaft nicht die dominante Wirtschaftsform war, ist die Persistenz eines Beobachtungsschemas, das auf monetarisierte und marktwirtschaftlich organisierte Industriegesellschaften zugeschnitten ist, symptomatisch für den kulturellen Partikularismus angeblich universeller Vergleichsmaßstäbe. Man könnte die Ergebnisse des Beitrags von Theresa Wobbe deshalb auch mit der Frage in Zusammenhang bringen, die Herder so präzise gestellt hatte (vgl. Abschn. 2): Wer wäre der „unparteiische Richter“ und was wäre „der genauestimmte, der unverfälschbare Maasstab“ eines internationalen Vergleichs? Herder wollte mit dieser Frage darauf aufmerksam machen, dass Kulturen grundsätzlich verschieden sind und es deshalb nicht möglich ist, sie demselben Vergleichsregime zu unterziehen. Wovon Herder warnte, ist in der Zwischenzeit geschehen: Beobachtungskategorien, die an westliche Gesellschaften angepasst sind, sind zu quasiuniversellen Vergleichsgrößen aufgestiegen, die alles und alle demselben Standard unterwerfen und damit die Diversität der Lebensformen – die „Ungleichheit des Vergleichenen“ (Luhmann 2002, S. 65) – weitgehend unsichtbar machen (am Beispiel globaler Indikatoren Merry 2016; anhand des Bruttosozialprodukts Speich Chassé 2013). Joachim Matthes (1992) spricht in diesem Zusammenhang von „Nostrifizierung“ im Sinne „einer Aneignung des Anderen nach eigenem Maß“: Die „andere“ Wirklichkeit wird konzeptuell an die vertraute ‚angeglichen‘ ... und was sich nicht fügt, wird in den Status einer Randdifferenz versetzt“ (Matthes 1992, S. 84). Am Ende wird nur noch das verglichen, was sich in die bestehenden Kategorien einfügt, und unterstellt wird dabei, dass sich überall dieselben Phänomene auffinden lassen, wenn auch vielleicht erst im Schattenriss eines „not yet“ (Chakrabarty 2007).

Auch der Beitrag von *Hannah Bennani* und *Marion Müller* befasst sich mit der internationalen Statistik, aber aus einem anderen Blickwinkel. Ausgangspunkt ist die Beobachtung, dass sich das Personal der Weltgesellschaft in den letzten 30 Jahren zunehmend diversifiziert hat (Bennani und Müller 2018). Immer neue Personengruppen werden auf globaler Ebene als weltweit relevante Personenkategorien identifiziert und zum Teil mit distinkten Rechten ausgestattet – Indigene (Bennani 2017), kleine Kinder (Tag 2020), Menschen mit Behinderung (Bennani und Müller 2020). Es gab natürlich schon vorher kleine Kinder, Menschen mit Behinderungen und sogenannte „erste Bewohner“. Neu ist aber, dass sie trotz ihrer kulturellen und sozialen Unterschiede als Angehörige einer weltweit identischen Kategorie beobachtet werden und sich teilweise auch selbst so wahrnehmen.

Ein entscheidender Schritt im Zuge der Institutionalisierung einer globalen Personenkategorie ist ihre Zählung: Wie viele von jeder „Sorte“ gibt es weltweit? Was auf den ersten Blick als eine einfache Auszählung erscheinen mag, ist jedoch, wie Hannah Bennani und Marion Müller in einer detaillierten Analyse der Herstellung globaler statistischer Zahlen belegen, ausgesprochen voraussetzungsvoll. Dass Zahlen Wirklichkeit nicht abbilden, sondern sie (mit-)erzeugen, ist nachgerade ein soziologischer Gemeinplatz. So verbreitet diese Behauptung ist, so wenig weiß man aber darüber, wie dies genau geschieht. Die Besonderheit des Beitrags von Hannah Bennani und Marion Müller liegt darin, dass sie die Blackbox der statistischen Fabrikation von Zahlen öffnen und anhand verschiedener Beispiele – Alter, Geschlecht,

Ethnizität und Behinderung – die vielfältigen Entscheidungen und Verfahrensweisen rekonstruieren, über die Zahlen mit globalem Geltungsanspruch hergestellt werden. Dazu zerlegen sie den Prozess des Zählens in einzelne Teilschritte und rekonstruieren für jedes Beispiel und für jeden Schritt – Definition der Kategorie, Operationalisierung, Kodierung und Aggregation –, wie konkret vorgegangen wurde.

Mit ihrer Fokussierung auf die Praktiken der Herstellung statistischer Zahlen gelangen sie zu einer Reihe interessanter Erkenntnisse. Die erste Einsicht betrifft die Frage, wie es der Statistik gelingt, die Kontingenz der Zahlenproduktion zu invisibilisieren und Zahlen als objektive Beschreibungen der Wirklichkeit erscheinen zu lassen. Wie ihre Analyse deutlich macht, liegt der entscheidende Schritt in der Aggregation der über Fragebögen erhobenen individuellen Informationen zu einer Zahl, die sich auf die von den Einzelindividuen abgelöste Gesamtheit aller Fälle bezieht. Während die Datenerhebung bestrebt ist, die je individuellen Konstellationen zu erfassen und die Vielfalt der menschlichen Körper und Identitäten möglichst detailgenau abzubilden, wird diese Vielfalt durch den Akt des Aggregierens eingedampft und auf eine oft dichotome Variable reduziert: behindert/nichtbehindert, indigen/nichtindigen, kleine Kinder/ältere Kinder. Mit diesem letzten Schritt wird eine neue Entität in die Welt gesetzt und in Gestalt einer Zahl objektiviert, die keine Spuren ihrer Herstellung mehr erkennen lässt – die globale Kategorie der „Indigenen“, der „kleinen Kinder“ oder der „Behinderten“.

Damit ist ein zweites Paradox verbunden – das Paradox, dass soziale Bewegungen, die sich für die Rechte einer bestimmten Personenkategorie einsetzen, einerseits die Starrheit kategorialer Grenzen kritisieren, mit ihrer Forderung nach Quantifizierung aber gleichzeitig zu ihrer Verfestigung beitragen. Denn um zu wissen, wie viele Behinderte es gibt, müssen Behinderte von Nichtbehinderten kategorial unterschieden werden. Diese Unterscheidung steht jedoch quer zur Auffassung, dass Behinderung ein graduelles Phänomen ist und alle – mehr oder weniger – behindert oder nicht-behindert sind. Und schließlich ist der Beitrag auch deshalb instruktiv, weil er zeigt, dass sich *Barnards* und *Fourcades* These einer Verschränkung von ordinalen Skalen und nominalen Klassifikationen sogar auf der Mikroebene der Zahlenproduktion nachweisen lässt: Auf der einen Seite werden kategoriale Merkmale immer häufiger über ordinal skalierte Variablen erhoben (unterschiedliche Grade von Behinderung), um dann in einem zweiten Schritt wieder in eine nominale Variable transformiert zu werden (behindert/nichtbehindert).

Um einen ganz anderen Typus von globalen Quantifizierungs- und Kategorisierungsprozessen geht es im Aufsatz von *Daniel Speich Chassé*, der den zweiten Teil des Bandes abschließt. Sein Gegenstand sind nicht Statistiken wie in den ersten drei Beiträgen, sondern die wissenschaftliche Vermessung von Umweltschäden sowie deren buchhalterische Verrechnung, z.B. im Rahmen des Emissionshandels. Ohne Zahlen, so seine leitende These, hätte die Umwelt nicht zu einem global kommunizierbaren und verhandelbaren Gegenstand werden können. Dieser These geht Daniel Speich Chassé aus einer historischen Perspektive nach. Ausgangspunkt ist die Beobachtung, dass Umweltfragen erst in den 1970er-Jahren auf die weltpolitische Agenda gesetzt wurden und erst in dieser Zeit die Vorstellung einer „Welt-Umwelt“ entstand. Angesichts der Tatsache, dass sich weder das Klima noch die Bodenerosion, weder Giftstoffe noch Mikroorganismen an nationale Grenzen halten

und es seit dem 19. Jahrhundert eine Vielzahl von privaten Organisationen gibt, die sich für den Schutz der Umwelt einsetzen, ist diese späte weltpolitische Reaktion erklärungsbedürftig. In seinem Beitrag skizziert Daniel Speich Chassé zunächst die Vorgeschichte des umweltpolitischen Epochenbruchs der 1970er-Jahre und setzt diesen Bruch in Beziehung zu den lange Zeit unscharfen Konturen des Umweltbegriffs. Natur wird lokal erfahren und die Art und Weise, wie sich diese Erfahrung semantisch niederschlägt, ist kulturell variabel. Um eine weltweit anschlussfähige Umweltpolitik formulieren zu können, musste von diesen lokalen Deutungen abstrahiert werden und ein verallgemeinerbarer und genauerer Begriff von Umwelt entwickelt werden.

Ein abstrakter Begriff von Umwelt im Sinne eines weltweiten ökologischen Zusammenhangs, der alle betrifft und alle einschließt, war, so Daniel Speich Chassé, in einem doppelten Sinne an Quantifizierung gebunden. Eine eminente Rolle spielte zum einen die wissenschaftliche Vermessung der Natur, die mit der Entstehung der modernen Naturwissenschaften einsetzte und im 20. Jahrhundert zu jenen Modellierungen führte, die es erlaubten, die Umwelt als einen weltumspannenden Zusammenhang zu begreifen. Ebenso wichtig waren Zahlen aber auch für die weltpolitische Kommunikation. Zahlen entzogen den lokalen Naturerfahrungen ihren sinnlichen Gehalt und verwandelten sie in abstrakte Größen, die sich untereinander vergleichen ließen. Es ist eine Sache, Hitze am Körper zu erfahren oder zu beobachten, dass die Ernte verdorrt, aber eine ganz andere, sie auf einer weltweit standardisierten Skala zu messen und mit anderen Temperaturen in Beziehung zu setzen.

Die Erfassung der Umwelt in Termini globaler Kennziffern war die Voraussetzung für eine weitere Quantifizierungswelle, die primär ökonomischer – präziser: buchhalterischer – Natur war. Das Wissen darum, dass Naturressourcen endlich sind und Umweltschäden Kosten nach sich ziehen, führte zur Idee, sie zu inventarisieren und preislich abzubilden. Umweltschäden oder der Verbrauch natürlicher Ressourcen ließen sich als Kosten bilanzieren und den Verursachern, seien das Unternehmen oder Staaten, als „Schulden“ zurechnen. Dahinter stand die Vorstellung, dass sich Eingriffe in die Natur nicht nur quantifizieren, sondern auch monetär beziffern und extern zuschreiben ließen. Diese Buchhaltungspraxis wurde mit der Zentrierung des umweltpolitischen Diskurses auf die Klimafrage perfektioniert. Der CO₂-Ausstoß eines Landes lässt sich nicht nur messen und vergleichen, sondern auch monetär beziffern und in Form von Emissionszertifikaten handeln. Wie Daniel Speich Chassé in seinem Beitrag ausführt, ist eine solche ökonomische Quantifizierung nicht neu. Sie begann mit der Vergleichbarmachung und Kommodifizierung von Landwirtschaftsflächen im 17. Jahrhundert und der Neudeutung von Landbesitz als Kapital, setzte sich mit der Auszählung und Monetarisierung von Naturressourcen fort und nahm in der zweiten Hälfte des 20. Jahrhundert die Form einer globalen Umweltbuchhaltung an. Mit diesem vorläufig letzten Schritt wurde alles, was es in der „Natur“ gibt – Tiere, Schadstoffe und Gase, Pflanzen, Steine und Erden – vergleichbar und gegenseitig verrechenbar gemacht. Es ist diese gewaltige Abstraktionsleistung, die der wissenschaftlichen und buchhalterischen Quantifizierung der Umwelt zugrunde liegt, und sie zu einem weltweit kommunizierbaren Phänomen macht.

4 Wertende Vergleiche: Monitoringinstrumente und Rankings

Die amtliche Statistik war das erste moderne Beobachtungsformat, das Daten in regelmäßigen Abständen erhob und in diesen Daten nach übergreifenden Zusammenhängen suchte. Im Verlaufe des 20. Jahrhunderts entstanden weitere Beobachtungsformate. Ein für die Gesellschaftsbeobachtung besonders relevantes und für digitale Technologien folgenreiches Format war die Umfrageforschung, die sich in den 1930er-Jahren verbreitete und im Gegensatz zur amtlichen Statistik mehrheitlich von privaten Unternehmen betrieben wurde (Igo 2007; Keller 2001; in Bezug auf digitale Technologien Turow 2011). Während die amtliche Statistik vor allem die äußeren Merkmale des *homme moyen* (Quetelet) vermaß, versuchte die Meinungs-, Markt- und Publikumsforschung das Innere des *averaged man* (Igo 2007) zu ermitteln, seine Einstellungen, Werte und Präferenzen. In den 1960er- und 1970er-Jahren kamen neue Beobachtungsinstrumente hinzu – das Sozialmonitoring und das Format der Rankings und Ratings –, die im Unterschied zur Statistik das Vergleichene nicht nur auf Ähnlichkeit und Differenz hin beobachten, sondern es gleichzeitig in eine Rangfolge bringen. Die Beiträge von *Martin Petzke* sowie von *Leopold Ringel* und *Tobias Werron* befassen sich mit diesen beiden neuen Beobachtungsformaten und rücken zwei Aspekte in den Vordergrund, die in den vorangehenden Beiträgen eher implizit blieben: zum einen die zeitliche Dimension von Vergleichen und Bewertungen (Ringel und Werron) und zum andern die Werturteile, die auch in augenscheinlich neutrale Beobachtungsinstrumente eingehen (Petzke).

Obschon die Herstellung von Zahlen eine Vielzahl von Entscheidungen erfordert, werden sie oft für die Sache selbst gehalten: „What is counted usually counts“ (Miller 2001, S. 386). Entscheidungen, die sich auf quantitative Daten stützen, führen eine Aura des Sachnotwendigen mit sich: Sie präsentieren sich als Entscheidungen, die keine sind. Oder wie es Theodore Porter formulierte: „Quantification is a way of making decisions without seeming to decide“ (Porter 1995, S. 8). Demgegenüber belegen Studien zu den Praktiken der Quantifizierung, dass Zahlen Wirklichkeit nicht einfach abbilden, sondern sie nach ihren eigenen Gesichtspunkten mitkonstruieren und dadurch implizit auch Bewertungen mit sich führen. Wenig thematisiert wurde jedoch, *wie* sich Bewertungen in angeblich objektive Wirklichkeitsbeschreibungen einschreiben.

Hier setzt der Aufsatz von *Martin Petzke* ein. Er befasst sich mit den deutschen Integrationsmonitoren und zeigt an diesem Beispiel, dass die im Monitoring involvierten Klassifikations-, Vergleichs-, und Vermessungsprozesse nahezu zwangsläufig mit Bewertungen einhergehen. Der Integrationsmonitor tritt mit dem Anspruch auf, die Integration der Immigrantenbevölkerung anhand von statistischen Daten zu vermessen und damit eine objektive Grundlage für wissenschaftlich abgestütztes politisches Handeln zu liefern. Der Beitrag belegt jedoch, dass die sozialstatistische Beobachtung regelmäßig „Bewertungsüberschüsse“ produziert – moralische Urteile, die systematisch auf Kosten der Immigranten gehen. Wie solche Bewertungsüberschüsse erzeugt werden, beschreibt er anhand einer detaillierten Analyse von drei statistischen Basisoperationen: die Klassifikation der Bevölkerung in Personen mit und ohne Migrationshintergrund, der Vergleich der beiden Bevölkerungsgruppen entlang verschiedener Parameter und die Auswahl der Integrationsindikatoren.

So führt bereits die bloße Unterscheidung zwischen Personen mit und ohne Migrationshintergrund eine Bewertung mit sich – das Nominale wird sozusagen automatisch ordinalisiert. Wie dieses Umkippen ins Ordinale geschieht, erläutert Martin Petzke unter Bezugnahme auf die Theorie der Markiertheit (Brekhus 1998; Zerubavel 2018) und illustriert diesen Prozess an vielen empirischen Beispielen. Die Theorie der Markiertheit argumentiert, dass nominale Unterscheidungen immer einen ordinalen Überschuss mit sich führen. Der Grund dafür liegt darin, dass die beiden Seiten einer Unterscheidung in der Regel nicht in einem symmetrischen Verhältnis stehen, sondern die eine Seite das Normale, Selbstverständliche und normativ Geltende repräsentiert und die andere Seite für das Außergewöhnliche und Abweichende steht. Beispiele dafür, dass nur das Ungewöhnliche sprachlich markiert wird, während das für selbstverständlich Gehaltene nicht spezifiziert wird, lassen sich mühelos finden. Man spricht von „offen homosexuell“, aber nicht von „offen heterosexuell“, von „Familienvätern“, aber nicht von „Familienmüttern“, von „African Americans“, aber nicht von „European Americans“.

Auch im Falle der Klassifikation der Bevölkerung in Personen mit und ohne Migrationshintergrund ist die asymmetrische Markierung offensichtlich. Die Aufmerksamkeit wird auf Personen mit Migrationshintergrund gelenkt, während die andere Seite nicht speziell benannt wird. An Cem Özdemir haftete das Label „Politiker mit Migrationshintergrund“ an, bei Claudia Roth wäre niemand auf die Idee gekommen, sie als Politikerin *ohne* Migrationshintergrund zu kategorisieren. Der Umstand, dass die markierte Seite über ihre Abweichung vom „Normalen“ definiert ist, führt dazu, dass sie mit Defizit und Devianz konnotiert ist und der Migrationsstatus zu einem Masterstatus wird, der alle anderen Rollen einfärbt. Aus der Kategorie „Person mit Migrationshintergrund“ kann man kaum aussteigen, auch dann nicht, wenn die kategoriale Zuordnung funktional irrelevant ist. Gleichzeitig werden die Unterschiede zwischen den beiden Personengruppen überzeichnet und Differenzen, die quer zu dieser Kategorisierung liegen, invisibilisiert oder unter die Primärdifferenz „mit oder ohne Migrationshintergrund“ subsumiert. Klassenunterschiede werden so unter der Hand zu ethnischen Unterschieden, die öffentliche Aufmerksamkeit verlagert sich von „wir da unten“ versus „ihr da oben“ zu Unterschieden zwischen „uns“ und den ethnisch „Anderen“.

Ähnliches geschieht auch beim Vergleichen und Vermessen. Auch diese beiden Operationen sind nicht neutral, sondern erzeugen Bewertungsüberschüsse, die am Ende das anscheinend objektive Monitoring mit Werturteilen durchziehen. Die Pointe des Beitrags von Martin Petzke liegt im Nachweis, dass die Tendenz des Integrationsmonitors, Personen mit Migrationshintergrund aus einer Defizitperspektive zu beobachten, nicht den Vorurteilen der Statistiker und Integrationsbeamten angelastet werden kann, sondern durch die statistische Beobachtung selbst erzeugt wird, und zwar oft gegen den erklärten Willen der Zuständigen (ausführlicher Petzke 2020). Es ist die Eigenlogik der Klassifikation, des Vergleichs und der Vermessung, die unintendierte Bewertungen hervorbringt. Damit weist der Beitrag auch über seinen Anwendungsfall hinaus. Was Martin Petzke am Beispiel des Integrationsmonitors demonstriert, ließe sich auch auf die Statistik übertragen, die ebenfalls mit dem Anspruch auf bewertungsfreie Sachrationalität auftritt. Für die in der Literatur zur Statistik breit diskutierte Frage, inwieweit Statistiken eben doch nicht so neutral sind

und woran man das zeigen kann, liefert er eine theoretisch elegante und empirisch überzeugende Antwort. Zudem ist sein Beitrag ein anschauliches Beispiel dafür, welchen Gewinn es bringt, wenn man alle vier Operationen berücksichtigt.

Leopold Ringel und *Tobias Werron* befassen sich mit der zeitlichen Dimension des Vergleichens und Bewertens und unterscheiden zwischen verschiedenen Varianten von Temporalität. Im einfachsten Fall werden Sachverhalte nur hinsichtlich ihrer Differenz in der Zeit verglichen: Was war früher, was ist jetzt (temporale Vergleiche)? Komplexer sind Vergleiche, die den Vergleich wiederholen und damit in der Lage sind, nicht nur die *Gegenstände* des Vergleichs, sondern auch seine *Ergebnisse* miteinander zu vergleichen. Ringel und Werron sprechen hier von temporalisierten Vergleichen. Serielle Vergleiche sind eine Unterkategorie temporalisierter Vergleiche. Ihre Besonderheit besteht darin, dass der Vergleich in regelmäßigen Abständen und anhand identischer Kriterien erfolgt und es dadurch möglich wird, die zu unterschiedlichen Zeitpunkten durchgeführten Vergleiche ihrerseits zu vergleichen. Amtliche Statistiken, Rankings oder die Jahresbilanzen von Unternehmen sind typische Beispiele für serielle Vergleiche.

Der Beitrag geht der Entstehungsgeschichte serieller Vergleiche am Beispiel von Kunst- und Hochschulrankings nach und untersucht, unter welchen Bedingungen sie eine serielle Form annahmen. Der Beobachtungszeitraum reicht von den frühen, quasinumerischen Künstlerbewertungen im 18. Jahrhundert, den sogenannten „balances des peintres“ (Spoerhase 2014), über das Aufkommen der ersten Universitätsvergleiche Ende des 19. Jahrhunderts bis hin zur Gegenwart und belegt anhand vieler historischer Beispiele, welche Hürden zu überwinden waren, bis Praktiken des vergleichenden Bewertens zu ihrem heutigen seriellen Format fanden. In ihrer detaillierten Analyse der Vorgeschichte von Kunst- und Universitätsrankings rekonstruieren Ringel und Werron die komplexe Genese jener Merkmale, die das gegenwärtige Verständnis von Rankings ausmachen, in ihrer Historizität aber nur selten reflektiert werden: Erstens die Serialität von Rankings, d.h. die Tatsache, dass sie in regelmäßiger Wiederholung erstellt werden und es dadurch erlauben, Vergleiche, die zu verschiedenen Zeitpunkten stattfanden, ihrerseits zu vergleichen und daraus auf Aufstiegs- und Abstiegsbewegungen zu schließen; zweitens die Orientierung auf Leistungsvergleiche: Alles, was in den Bannkreis von Rankings gelangt, wird als Leistung ausgeflaggt; drittens diese vergleichende Leistungsbuchhaltung kombiniert mit dem Umstand, dass auf jedem Platz nur einer Platz hat, erzeugt eine kompetitive Dynamik und ein Streben nach permanenter Verbesserung. Diese Dynamik wird viertens durch den quantifizierenden Charakter von Rankings weiter gesteigert. Im Unterschied etwa zu Restaurantratings sind die Abstände zwischen den einzelnen Positionen bezifferbar und verstärken damit den Eindruck von Durchlässigkeit und ständiger Bewegung.

Auch wenn es heute so erscheinen mag, entstanden serielle Rankings nicht auf einen Schlag, vielmehr sind sie der Endpunkt zahlreicher und oft auch gescheiterter Experimente, wie Ringel und Werron in ihrer historischen Rekonstruktion zeigen. Im Falle von Kunstrankings musste der Überzeugung des Feldes Rechnung getragen werden, dass sich das künstlerische Schaffen nicht wie eine profane sportliche Leistung vermessen lässt. Dies gelang erst in den 1970er-Jahren mit der Publikation des *Kunstkompass*, der von einer Quantifizierung der künstlerischen Qualität absah und

Künstler und Künstlerinnen nur nach der ihnen zugeschriebenen Reputation in eine Rangfolge brachte (vgl. auch Buckermann 2020). Im Wissenschaftsfeld waren die Widerstände anderer Art. Hier lag das Hindernis vor allem in der Auffassung, dass das Streben nach Erkenntnis rein sachlich motiviert ist und nur die Wissenschaftler und Wissenschaftlerinnen selbst in der Lage sind, wissenschaftliche Leistung zu beurteilen. Es gab zwar seit Ende des 19. Jahrhunderts immer wieder Initiativen, die Qualität von Hochschulen zu vergleichen, aber es handelte sich entweder um Klassifikationssysteme, die zwar wiederholt wurden, aber die verglichenen Universitäten nicht in eine Rangfolge brachten, oder um bewertende Vergleiche, die aber nur einmalig veröffentlicht wurden. Es dauerte noch mehrere Jahrzehnte, bis sich mit dem *U.S. News* Hochschulranking das erste serielle Universitätsranking etablieren konnte, das Hierarchisierung und Wiederholung miteinander verband.

Der Aufsatz von Leopold Ringel und Tobias Werron belegt, welchen Gewinn es bringt, die Praktiken des Vergleichens, Bewertens und Quantifizierens auf ihre zeitliche Ordnung hin zu beobachten. Der Beitrag eröffnet der Vergleichs- und Bewertungssoziologie nicht nur eine neue Dimension, sondern er ist auch die bisher umfassendste Darstellung der Entstehungsgeschichte von Rankings. Die historische Prozessperspektive hat zudem den Vorteil, aufzeigen zu können, dass die für die heutigen Rankings typische Kombination von Vergleich, Bewertung und Quantifizierung nicht auf einen Schlag entstand, sondern schrittweise erfolgte. Es bedurfte spezifischer Bedingungskonstellationen, damit Bewertungen auf Dauer gestellt und quantifiziert werden konnten.

5 Digitale Beobachtungsformate: Vergleichen, Bewerten und Kategorisieren im Medium algorithmischer Berechnungen

Digitalisierung wird häufig als eine fundamentale Zäsur beschrieben. Demgegenüber plädieren die Beiträge des letzten Teils dafür, den Kontinuitäten mehr Aufmerksamkeit zu schenken. Digitale Technologien sind in vielerlei Hinsicht neu, sie führen aber auch Vorhandenes weiter, wenn auch in anderer Gestalt. Indem der Band das „Beobachten“ in den Mittelpunkt stellt, hebt er eine Kontinuitätslinie besonders hervor, nämlich dass es sich auch bei digitalen Technologien um Beobachtungsformate handelt, die ähnlich wie ihre Vorläufer Daten generieren und sie auf ihre Ordnungsmuster hin analysieren.

Die Art und Weise, wie sie das tun und zu welchem Zweck, unterscheidet sich allerdings. Während die früheren Beobachtungsformate an der Aufdeckung übergeordneter Zusammenhänge interessiert sind, wollen digitale Beobachtungsinstrumente etwas über den Einzelnen in Erfahrung bringen: Mit welcher Wahrscheinlichkeit wird A eine Straftat begehen? Wie groß ist die Gefahr, dass B eine Sepsis entwickelt? Mit welchen Empfehlungen kann C dazu gebracht werden, auch in Zukunft bei Netflix zu streamen? Ein weiterer Unterschied bezieht sich auf die Datenform. Heute verwenden zwar alle Beobachtungsformate *digitalisierte* Daten, aber nicht *digitale* Daten im strengen Sinne: Ihre Grundlage sind Umfragen (Statistik, Meinungsforschung) oder sekundäranalytische Auswertungen bereits erhobener Daten (Monitoringsysteme, partiell Rankings). Demgegenüber stützen sich digitale Tech-

nologien im Wesentlichen auf Online-Daten, also auf (digitale) Daten, die sie aus der Dauerbeobachtung der Netzaktivitäten ihrer Nutzer und Nutzerinnen gewinnen. Dass sie diese Daten mithilfe von Algorithmen analysieren, gilt zwar als Sondermerkmal digitaler Technologien, bei näherem Hinsehen sind die Grenzen jedoch porös. Denn auch statistische Verfahren verwenden Algorithmen (sonst könnte keine Statistiksoftware entwickelt werden), und umgekehrt greifen auch algorithmische Berechnungen auf statistische Methoden zurück (Boelaert und Ollion 2018). Ein weiterer Unterschied liegt darin, dass sich Vergleichen, Bewerten und Kategorisieren bei ihnen praktisch zeitgleich vollzieht, während dies bei den anderen Beobachtungsformaten tendenziell sequenziell geschieht.

Die im letzten Teil des Bandes versammelten Beiträge geben Auskunft darüber, wie digitale Technologien beobachten, zu welchem Zweck und mit welchen Folgen. Sie decken ein breites Spektrum an Applikationen ab: Dating-Apps (Peetz), Vermittlungsplattformen wie Uber oder Airbnb (Mennicken und Kornberger), Recommendersysteme (Unternährer) und Anwendungen, die prädiktive Algorithmen im Bereich der Arbeitsvermittlung (Büchner und Dosdall) und bei der Polizeiarbeit (Brayne) einsetzen. In teilweise sehr detaillierten Analysen der algorithmischen Vorgehensweise zeigen sie auf, dass diese mit Begriffen wie „Big Data“ und „Algorithmen“ nicht ausreichend beschrieben ist. Entscheidend ist, wie die Daten mithilfe von Algorithmen analysiert werden. Obschon sie sich mit unterschiedlichen Anwendungen befassen, gelangen sie zu einem ähnlichen Befund: Die Operationsweise von Algorithmen lässt sich im Wesentlichen darüber erfassen, dass sie Vergleichen, Bewerten und Kategorisieren in einem ständigen rekursiven Prozess miteinander verknüpfen und sich diese Verknüpfung im Medium algorithmischer Berechnungen vollzieht. Insofern bringen digitale Technologien das Anliegen des Bandes besonders augenfällig zum Ausdruck.

Den Auftakt macht ein Beitrag von Stefanie Büchner und Henrik Dosdall, der aus einer organisationssoziologischen Perspektive den Einsatz prädiktiver Algorithmen in der Arbeitsvermittlung untersucht. Ihr Untersuchungsfall ist ein Algorithmus zur Berechnung der Integrationschancen arbeitssuchender Personen, der seit 2019 vom österreichischen Arbeitsmarktservice (AMS) eingesetzt wird. Im Gegensatz zu vielen anderen Studien, die die „social power of algorithms“ (Beer 2017) den Algorithmen selbst zuschreiben, legen Büchner und Dosdall den Fokus auf den organisatorischen Kontext, in den sie eingebettet sind: Es ist ihre Einbindung in organisatorische Entscheidungsprozesse, die Algorithmen „actionable“ machen, d. h. mit Handlungsfähigkeit und sozialer Relevanz ausstatten. Dies gilt besonders für Entscheidungsalgorithmen, wie sie etwa bei der polizeilichen Ermittlung oder bei der Verteilung knapper Ressourcen eingesetzt werden.

Um diese These zu fundieren, erläutern sie zunächst, wie der AMS-Algorithmus konkret funktioniert, und zeigen, dass die Berechnung der Integrationschancen auf einer Verknüpfung von vergleichenden, bewertenden und kategorisierenden Operationen beruht. Im Unterschied zu vielen anderen digitalen Technologien greift der AMS-Algorithmus nicht auf Online-Daten, sondern auf bereits vorhandene soziodemografische und berufsbiografische Daten zurück. Die Arbeitssuchenden werden in einem ersten Schritt anhand solcher Merkmale je einzeln miteinander verglichen. Dieser Vergleich ist die Grundlage dafür, sie auf einer ordinalen Skala einzustufen,

die den Grad ihrer jeweiligen Integrationschancen wiedergibt. Um die Einstufung organisatorisch handhabbar zu machen, werden sie am Ende in drei Kategorien, sogenannte „Chancengruppen“, eingeteilt, aus denen sich die jeweiligen Unterstützungsmaßnahmen ableiten.

Die Klassifikation, die der Algorithmus vornimmt, ist für sich genommen noch nicht relevant. Soziale Relevanz erhält sie erst dadurch, dass Organisationsentscheidungen auf sie Bezug nehmen. Erst dadurch wird der Algorithmus mit „social power“ (Beer) ausgestattet. Wie Büchner und Dosdall zeigen, kann die Art und Weise, wie Organisationen Algorithmen in ihre Entscheidungsprozesse einbauen, variieren. Im Falle des AMS-Algorithmus, der in einer Organisation eingesetzt wird, in der die Fallbearbeitung selbst schon in gewissem Maße „algorithmisiert“ ist und den Vermittlern nur wenig persönlicher Spielraum bleibt, werden die Fallentscheidungen durch den Algorithmus weitgehend vorstrukturiert. Das muss nicht zwangsläufig so sein. Die Berechnungen von Algorithmen könnten auch zu explorativen Zwecken oder als zusätzliche Entscheidungshilfe verwendet werden, von der nur fallweise und nach Ermessen der Vermittlerinnen Gebrauch gemacht wird. Es ist mit anderen Worten die Organisation und ihre Entscheidungsarchitektur – und nicht der Algorithmus selbst –, die darüber bestimmen, wie er eingesetzt wird. Indem Büchner und Dosdall den organisatorischen Kontext in den Mittelpunkt rücken, stellt ihr Aufsatz eine wichtige Ergänzung zur bisherigen Literatur dar. Es gibt zwar viele Studien, die sich mit den Auswirkungen der Digitalisierung auf die Arbeitsverhältnisse beschäftigen (exemplarisch Rosenblat 2018) oder untersuchen, wie Algorithmen in organisationale Prozesse eingreifen und menschliches Entscheiden unter Umständen auch ersetzen (als Überblick Kellog et al. 2020). Sie gehen aber nicht so weit wie Büchner und Dosdall, die den Einsatz von Algorithmen konsequent als ein organisationssoziologisches Problem rahmen und damit nicht mehr die Mensch/Maschine-Beziehung, sondern das Verhältnis von Algorithmus und Organisation in den Mittelpunkt stellen (aus einer ähnlichen Perspektive Kette 2020).

Auch der Beitrag von Sarah Brayne zur polizeilichen Ermittlung befasst sich mit prädiktiven Algorithmen. Der Fokus liegt aber nicht auf organisationssoziologischen Fragen, sondern auf den sozialen Folgen des Einsatzes von Algorithmen. Anhand einer mehrjährigen ethnografischen Fallstudie im Polizeidepartment von Los Angeles zeigt sie materialreich auf, wie sich die Überwachungspraktiken der Polizei durch die Verwendung von Big Data und prädiktiven Methoden verändert haben. Es ist eine der wenigen Studien, die im Detail untersuchen, wie Big-Data-Überwachung konkret funktioniert und welche Folgen sich daraus für die soziale Ungleichheit ergeben. Im Einzelnen identifiziert sie fünf Entwicklungen, die die Polizeiarbeit nachhaltig verändert haben. Der entscheidende Trend ist die Umstellung von reaktiver zu prädiktiver Ermittlung. Während die konventionelle polizeiliche Ermittlungsarbeit in der Regel erst nach der Tat anläuft, werden heute vermehrt Algorithmen eingesetzt, die potenzielle Täter prospektiv zu erfassen suchen.

Im Zuge der Umstellung auf prädiktive Ermittlungsverfahren wird das Erfahrungswissen der Polizisten zunehmend durch automatisierte Systeme ersetzt, die für sie die Entscheidungen fällen oder sie zumindest vorbereiten. Der Polizist wird in gewissem Maße zu einem Ausführungsorgan dessen, was Softwaresysteme wie

Palantir oder Predpol ihm vorschlagen.⁸ Ein Beispiel dafür ist die Verlagerung von abfragegestützten zu alarmgestützten Systemen. Im ersten Fall geht die Suchanfrage von den Polizistinnen und Polizisten aus, etwa wenn sie die Zentrale um Überprüfung eines Autokennzeichens bitten. Im zweiten Fall erhalten sie automatisch eine Meldung, sobald eine bestimmte Variablenkonfiguration in den Daten vorliegt, z. B. wenn in einem bestimmten Gebiet ein Haftbefehl ausgestellt wurde und sich gleichzeitig ein Auto mit einem bestimmten Kennzeichen dort befindet. Das ermöglicht Überwachung „at a distance“ – eine Rundumbeschattung ist nicht mehr notwendig. Gleichzeitig erweitert sich der Kreis jener, die ins Visier der Überwachung geraten. In den polizeilichen Datenbanken sind nicht mehr nur Personen erfasst, über die bereits Akten vorliegen, sondern zunehmend auch Menschen, die bisher noch keinen Kontakt mit der Polizei hatten, z. B. indem über netzwerkanalytische Methoden auch Personen identifiziert werden, die mit dem Verdächtigen in Verbindung stehen oder irgendwann einmal standen. Die in Datenbanken registrierten Personen können mithilfe von Algorithmen entlang verschiedener Parameter verglichen und hinsichtlich ihrer mutmaßlichen Delinquenzbereitschaft bewertet und klassifiziert werden. Je nachdem in welche Risikokategorie sie fallen, werden andere polizeiliche Überwachungsmaßnahmen ausgelöst.

Die Möglichkeit, Unmengen an Daten in Sekundenschnelle zu analysieren, motiviert dazu, ehemals separierte Datenbanken zu integrieren und immer mehr Daten einzuspeisen – neben polizeilichen Daten auch Daten, die von anderen Behörden oder von Privatunternehmen, wie Facebook oder Amazon, gesammelt wurden. Der Zugriff auf immer umfassendere Datenbanken führt dazu, dass Personen nicht nur nach ihrer mutmaßlichen Delinquenzbereitschaft eingestuft werden, sondern diese Einstufung mit Sortierungen verbunden wird, die aus anderen institutionellen Kontexten stammen. Dies ist ein typischer Fall dessen, was Fourcade und Healy (2013) als „classifications situations“ bezeichnet haben (vgl. Abschn. 2). Die Studie von Sarah Brayne ist insofern ein instruktives Beispiel für Fourcades These, dass der Einsatz von algorithmischen Techniken neue Muster sozialer Ungleichheit hervorbringt. An Personen mit einem hohen polizeilichen Risikoscore haftet ein „kategorialer Verdacht“ (Lyon 2002), der zusammen mit anderen individuellen Markern, etwa Popularitätsscores, Kreditscores oder Punkten, die Versicherungen für das Gesundheitsverhalten vergeben, ein digitales Stigma erzeugt – und bei anderen ein digitales Privileg.

Digitale Beobachtungstechnologien fungieren mit anderen Worten als eine Art „Sortiermaschinen“, die jede und jeden anhand einer Vielzahl von Kriterien einstufen und mit einem individuellen digitalen Profil versehen, das erhebliche ungleichheitsrelevante Effekte hat. Wer als wenig kreditfähig eingestuft wird, erhält Kredite zu schlechteren Konditionen und wird bei einer Bewerbung möglicherweise nicht berücksichtigt, wessen Fahr- oder Gesundheitsverhalten als riskant beurteilt wird, zahlt höhere Versicherungsprämien, und wer einen erhöhten Kriminalitätsscore hat, wird möglicherweise Schwierigkeiten haben, Sozialleistungen zu erhalten. Fourcade und Healy (2016) sehen in dieser Verschränkung von Ordinalisierung und (Re-)Ka-

⁸ Ähnliche Systeme werden auch in Deutschland und der Schweiz unter dem bezeichnenden Namen *Pre Crime Observation System* (Precobs) eingesetzt.

tegorisierung eine neue Form von Kapital, das sie „Überkapital“ nennen: „We can think of übercapital as made up of all the digital information available about an individual, encapsulating the totality of her relations as expressed through digital traces, ordered and made tractable through scoring and ranking methods“ (Fourcade und Healy 2016, S. 18). Das Sozialkredit-System in China, das die soziale Bonität der Bürger über eine Vielzahl von Skalen erfasst, bringt diese Entwicklung zur Perfektion (Wong und Dobson 2019).

Prädiktive Algorithmen sind eine Variante digitaler Beobachtungsinstrumente. Verhaltensdaten werden dabei mit dem Ziel analysiert, aus ihnen auf die Wahrscheinlichkeit eines bestimmten Verhaltens in der Zukunft zu schließen. Recommendersysteme treffen zwar auch Voraussagen, aber es geht nicht bloß um die Prognose zukünftigen Verhaltens, sondern darum, die Nutzer zu einem bestimmten Verhalten zu motivieren, etwa ein Produkt zu kaufen oder weiterhin bei Netflix zu streamen. Deshalb bezeichnet Nick Seaver (2019) Recommendersysteme auch als „traps“. Mit dieser vor allem bei Onlineshops, Suchportalen und Dating-Apps verbreiteten Technologie befassen sich die Beiträge von Markus Unternährer und Thorsten Peetz.

Recommendersysteme sprechen uns als gute Bekannte an: „Songs extra für Dich“, „Dieses Buch könnte Dich auch interessieren“. Den Umstand, dass uns Empfehlungssysteme persönlich adressieren, nehmen viele Autoren zum Nennwert und interpretieren ihre Verbreitung als Indiz einer um sich greifenden Singularisierung (exemplarisch Reckwitz 2017; Mau 2017). Nimmt man jedoch die Prozesse in den Blick, die sich hinter der Bildschirmoberfläche abspielen, erhält man ein ganz anderes Bild: dort wird nicht singularisiert, sondern relationiert. Wie dies genau geschieht, zeigt Markus Unternährer in einer detaillierten Analyse der Operationsweise von Recommendersystemen. Er räumt dabei auch mit dem Vorurteil auf, dass alle Empfehlungen personalisiert sind. Es gibt auch nicht personalisierte Empfehlungssysteme, etwa Netflix’ „Top 10 in der Schweiz heute“, sowie schwach personalisierte Recommender, die unterschiedlichen kategorialen Gruppen unterschiedliche Empfehlungen anbieten: Frauen werden andere Bücher empfohlen als Männern, jüngeren Menschen andere Sportartikel als Älteren.

Von solchen nicht oder nur schwach personalisierten Recommendersystemen unterscheidet Markus Unternährer zwei personalisierte Empfehlungssysteme. Diese schneiden ihre Empfehlungen zwar auf den einzelnen Nutzer zu, sie beobachten ihn aber gerade nicht in seiner Individualität. *Inhaltsbasierte* Recommender, wie sie etwa bei wissenschaftlichen Suchmaschinen zum Einsatz kommen, leiten ihre Empfehlungen aus Ähnlichkeitsrelationen zwischen Objekten, etwa angeklickten Artikeln, ab. Recommendersysteme, die auf *Collaborative Filtering* beruhen, berechnen sie auf der Basis der Bewertungen aller anderen Nutzer. Deshalb auch der Begriff „*kollaboratives* Filtern“. Wie Markus Unternährer in einem instruktiven Abschnitt darlegt, wurde dieser Begriff ursprünglich für Online-Gruppen verwendet, in denen die Beteiligten untereinander Empfehlungen austauschten. Recommendersysteme, die das Verfahren des Collaborative Filtering nutzen, ersetzen diesen sozialen Tausch durch algorithmische Relationen: An die Stelle sozialer Beziehungen tritt die „algorithmische Sozialität“ (Unternährer) jener, die Dinge ähnlich bewerten, sich aber weder

untereinander kennen noch davon wissen, dass der Algorithmus sie miteinander in Verbindung bringt.

Um die Funktionsweise von personalisierten Recommendersystemen zu erfassen, führt Markus Unternährer den Begriff des *Matching* ein. Matchings sind für ihn Vergleiche, die nicht die Nutzer selbst vergleichen, sondern die *Relation* – die Passung – zwischen spezifischen Nutzern und spezifischen Dingen indiziert, z. B. durch einen Kauf. Es interessiert nicht, wer die Nutzer kategorial gesehen sind oder welcher Art die Produkte sind, die sie gekauft haben. Es interessiert allein, welche Dinge sie angeklickt, gekauft oder gestreamt haben. Um zu ermitteln, welche Produkte zu welchen Nutzern passen, also „good matches“ sind, wird jeder Nutzer mit allen anderen Nutzern danach verglichen, ob sie ähnliche Dinge ähnlich bewertet haben. Den Begriff „good matches“ übernimmt er von Viviana Zelizer (2006) und adaptiert ihn an seine Fragestellung. Über passende personalisierte Empfehlungen werden nicht nur „good matches“ zwischen Nutzern und Produkten hergestellt, sondern auch Bindungen zwischen Nutzern und Unternehmen. Sie sind der „Kitt“, der Nutzer an ein Unternehmen bindet, die *Nudges*, die sie dazu bringen, weiterhin dort zu kaufen oder zu streamen – und dem Unternehmen folglich auch künftig Daten zur Verfügung zu stellen, die für weitere Empfehlungen genutzt oder auch an Dritte verkauft werden können.

Der Beitrag von Markus Unternährer dokumentiert, welche Einsichten sich gewinnen lassen, wenn man sich nicht mehr mit dem begnügt, was auf der Bildschirmoberfläche sichtbar ist, sondern die dahinterliegenden algorithmischen Prozesse in den Blick nimmt. Sein Vergleich der vier Arten von Recommendern belegt nicht nur, dass die personalisierende Adressierung von Empfehlungssystemen eine kommunikative Fiktion ist, die man besser nicht zum Ausgangspunkt aufgeregter Gesellschaftsdiagnosen machen sollte (vgl. auch Heintz). Er zeigt darüber hinaus, dass sich Recommendersysteme zwar als das grundlegend Neue präsentieren, sie aber in vielerlei Hinsicht in Kontinuität zu früheren Verfahren stehen – zum einen zu Rankings und Ratings, die in gewisser Weise als Recommendersysteme *avant la lettre* betrachtet werden können, zum anderen zu den Segmentations-techniken der klassischen Marktforschung. Während sich die Marktforschung ursprünglich am „Durchschnittskonsumenten“ orientierte, spalten Segmentations-techniken die Masse der Konsumentinnen und Konsumenten in immer kleinere soziodemografische Segmente auf (Turow 2011; Pridmore und Hämäläinen 2017). Schwach personalisierte Recommender schließen direkt an die Technik des segmentierten Targeting an, indem sie ihre Empfehlungen nach soziodemografischen Merkmalen differenzieren. Personalisierte Recommender treiben diese Technik auf die Spitze und stellen sie gleichzeitig auf den Kopf, indem sie ihre Empfehlungen nur noch an „segments of one“ (Keith Weed 2017) adressieren, aber dazu die Gesamtheit aller Nutzer einbeziehen.

Recommendersysteme sind praktisch überall im Gebrauch – Suchmaschinen verwenden sie, Onlineshops, Streamingdienste oder Nachrichtenportale. Eine besonders vielschichtige Variante von Recommendersystemen sind Dating-Apps. Mit ihnen befasst sich die Fallstudie von Thorsten Peetz am Beispiel von Tinder. Paarbildung vollzieht sich immer häufiger über die Mittlerdienste von Dating-Apps, die unterschiedlichste soziale oder sexuelle Vorlieben bedienen: Veggly führt Veganer

zusammen, Joyce Menschen, die eher an flüchtigen Begegnungen interessiert sind, und Lumen – 50+ richtet sich an Personen, deren Paarbildungschancen alterswegen prekär sind. Auf den ersten Blick stehen Dating-Apps in eklatantem Widerspruch zum modernen Liebesideal, dessen Kommunikationscode es erfordert, den andern in seiner Einzigartigkeit bedingungslos zu bestätigen: Wer liebt, vergleicht nicht. Aber genau das tun die Nutzer und Nutzerinnen von Dating-Apps, zumindest bis zu jenem Moment, wo die Richtige, der Richtige gefunden ist. Dazu kommt, dass nicht nur die Suchenden vergleichen, sondern die Vorselektion von einem Algorithmus getroffen wird, der die Nutzer systematisch miteinander vergleicht und ihnen auf dieser Basis ausgewählte Profile vorlegt. Das heißt bevor sich die Nutzer und Nutzerinnen durch das Angebot wischen und die einen mit einem „Like“, die anderen mit einem „Nope“ versehen, werden sie von Algorithmen zueinander in Beziehung gesetzt, um herauszufinden, wer sich hinsichtlich Vorlieben, Wünschen oder Eigenschaften am ähnlichsten ist.

Der Beitrag von Thorsten Peetz befasst sich nicht so sehr mit der Frage, wie die Nutzer und Nutzerinnen selbst zu einer Wahl gelangen, sondern er rückt die infrastrukturelle Grundlage ihrer Bewertungspraktiken in den Vordergrund – die im Hintergrund ablaufenden algorithmischen Vergleichs- und Kategorisierungsprozesse. Ausgehend von den bewertungssoziologischen Konzepten, die er im ersten Teil einführt, analysiert er die Vorgehensweise von Dating-Apps und zeigt, dass sie nicht anders funktionieren als Recommender, die Nutzern Kaffeemaschinen, Publikationen, Filme oder Zeitungsartikel empfehlen. Sie vergleichen, bewerten und kategorisieren und leiten daraus „good matches“ ab (*Unternährer*). Die Differenz besteht einzig darin, dass im einen Fall Dinge empfohlen werden, im anderen Fall potenzielle (Liebes-)Partner. Im Unterschied zu anderen digitalen Beobachtungsinstrumenten, die Bewertungen in Form von Leistungs- oder Popularitätsscores öffentlich machen, wird die Bewertung, die Tinder selbst vornimmt, unsichtbar gemacht. Die Tatsache, dass Tinder kein Begehrtheitsranking veröffentlicht (aber davon vermutlich Gebrauch macht), ist möglicherweise, so Thorsten Peetz, eine „implizite Verneigung“ vor der Idee romantischer Liebe. Wo es um Gefühle geht, haben Zahlen und Skalen nichts zu suchen, ähnlich wie Popularitätsmetriken nicht mit der Selbstdarstellung einer Plattform verträglich sind, die sich wie Instagram das „sharing“ und „caring“ auf die Fahnen geschrieben hat (Mosseri 2019). Im Falle von Dating-Apps reichen „good matches“ aber offenbar nicht aus, um die Nutzer an das Unternehmen zu binden, es braucht zusätzliche Vorrichtungen, die verbergen, dass der Herstellung von persönlichen Beziehungen unpersönliche Vergleichs- und Rechenoperationen zugrunde liegen.

Damit wird gleichzeitig Reaktivität verhindert. Während Universitäten darum wissen, dass sie beobachtet und vermessen werden und im Wissen um diese Beobachtung ihr Verhalten darauf einstellen (Espeland und Sauders 2007), wird der Umstand, dass auch Tinder selbst beobachtet und berechnet, unsichtbar gemacht. Diese Invisibilisierung ist aber selbst hoch reflexiv. Da die Softwareentwickler antizipieren, dass die Sichtbarmachung des „desirability ranking“ bei den Nutzern Reaktionen auslösen könnte, die die Logik der App unterlaufen, richten sie ihr Design prospektiv darauf aus (zu einer solchen „second-order reactivity“ Petzke 2020). Das Wissen darum, dass Tinder die Likes und die Nopes zu einem Begehrtheitscore

verrechnet und ihn in ihre Vorschläge einfließen lässt, würde nicht nur die Fiktion von Wahlfreiheit erschüttern, sondern auch Strategien evozieren, den eigenen Score zu verbessern.

Der Band wird mit einem Beitrag von *Andrea Mennicken* und *Martin Kornberger* abgeschlossen, der einen analytischen Rahmen für die Analyse digitaler Bewertungstechnologien entwirft. Theoretischer Ausgangspunkt sind die *Valuation Studies* und das dort entwickelte Argument, dass der Wert eines Gutes ihm nicht inhärent ist, sondern durch Bewertungspraktiken und -technologien hervorgebracht wird. Wie solche Bewertungspraktiken in der Welt des Digitalen funktionieren, was sie ermöglichen und auf welche Weise sie bestehende Praktiken der Bewertung verändern, untersuchen Mennicken und Kornberger am Beispiel von Uber und Airbnb. Die Besonderheit solcher digitalen Plattformen besteht darin, dass sie eine Infrastruktur bereitstellen, über die Käufer und Verkäufer in einen direkten Austausch treten können.

Plattformorganisationen unterscheiden sich in vielerlei Hinsicht von traditionellen Unternehmen. Sie besitzen die Produkte nicht, die sie vermitteln, die Beschäftigten sind nicht in ein formales Arbeitsverhältnis eingebunden und die Leistungsbewertung erfolgt nicht von oben, sondern über die Bewertungen der Käuferinnen und Käufer. Aus Sicht der Marktsoziologie ist der Erfolg digitaler Marktplätze an sich hochgradig unwahrscheinlich, da sich das Problem der Koordination bei ihnen noch um einiges radikaler stellt als in konventionellen Märkten (Beckert 2007). Die Abwicklung von Transaktionen erfolgt im Netz weitgehend anonym und die Transaktionen vollziehen sich unter größter Unsicherheit. Es fehlen die üblichen Indizien, um den Wert eines Produktes oder einer Dienstleistung einzuschätzen, und die Teilnehmer haben kaum Anhaltspunkte zur Hand, um die Vertrauenswürdigkeit ihres potenziellen Partners zu beurteilen. Wie soll man wissen, ob die angebotene Unterkunft tatsächlich ihrer Beschreibung entspricht, wenn man sie sich nicht persönlich ansehen kann, und wie kann man einschätzen, ob die Vermieterin (oder der Gast) die Verpflichtungen einhält? Das jedem sozialen Handeln eigene Problem der doppelten Kontingenz stellt sich mit anderen Worten in digitalen Märkten in besonderer Schärfe, ohne dass für seine Lösung auf etablierte Mechanismen zurückgegriffen werden könnte. Wie lässt sich angesichts dieser Problemlage der Erfolg digitaler Plattformen erklären?

Zur Beantwortung dieser Frage rücken Andrea Mennicken und Martin Kornberger die digitalen *evaluativen Infrastrukturen* in den Vordergrund, die das Rückgrat digitaler Plattformen bilden, und argumentieren, dass diese nicht nur eine Lösung für das Koordinationsproblem digitaler Märkte anbieten, sondern darüber hinaus neue Strukturen erschaffen. Den Begriff der digitalen evaluativen Infrastruktur bestimmen sie über drei Elemente: Erstens verstehen sie unter evaluativen Infrastrukturen das gesamte System miteinander verknüpfter quantifizierter Bewertungstechniken und nicht bloß einzelne Bewertungsinstrumente. Zweitens zeichnen sich evaluative Infrastrukturen aus ihrer Sicht durch eine spezifische Machttechnik aus, die sie in Anlehnung an den informationstechnologischen Begriff des Protokolls als „protokollarische Macht“ bezeichnen. Protokollarische Macht ist eine Form der Machtausübung, die auf direkte Kontrolle verzichtet und sich darauf beschränkt, die Regeln zu setzen, und diese Regeln in der Architektur der Plattform implementiert. Auf

digitalen Plattformen werden Anbieter, seien das Taxifahrer, Gastgeberinnen oder Verkäufer, durch die Kunden kontrolliert, aber es sind die Plattformbesitzer, die darüber entscheiden, welche Daten gesammelt und wie sie ausgewertet werden und zu welchem Zweck. Und drittens zeichnen sich digitale Plattformen durch Generativität aus, indem sie ständig neue Verbindungen, neue Kategorisierungen und neue Formen der Wertschöpfung generieren. Mit ihrer Betonung der Generativität von Plattformen grenzen sich Mennicken und Kornberger von Ansätzen ab, die die Reaktivität (exemplarisch Espeland und Sauder 2007) oder Performativität (exemplarisch MacKenzie und Millo 2003) von Modellen oder Bewertungstechniken in den Vordergrund stellen.

Wie digitale Plattformen im Einzelnen funktionieren, beschreiben Andrea Mennicken und Martin Kornberger anhand einer detaillierten Analyse der Operationsweise von Uber und Airbnb. Ihre Analyse der Operationsweise digitaler evaluativer Infrastrukturen macht deutlich, dass digitale Plattformen ähnlich vorgehen wie die digitalen Technologien, mit denen sich die vorangehenden Beiträge beschäftigen. Uber unterwirft die Fahrer einer ständigen Beobachtung. Alles, was sie tun (oder unterlassen, zu tun) wird in Echtzeit aufgezeichnet – die gefahrenen Strecken, die Anzahl abgelehnter Passagiere, die Einhaltung der vorgeschriebenen Routen und vieles andere mehr – und unterzieht diese Daten einer algorithmischen Analyse, die Vergleichs-, Kategorisierungs- und Bewertungsoperationen in einem ständigen Wechselspiel miteinander verknüpft. Die Ergebnisse lösen ihrerseits Entscheidungen aus, die oft automatisiert erfolgen, z. B. Entscheidungen über Leistungsvorgaben oder die Entscheidung, eine Fahreraccount zu deaktivieren oder neue Fahrpreise festzulegen. Die angeblich „selbstständigen“ Fahrer werden zu einem Rädchen in einem weitgehend automatisierten System degradiert, in das sie nicht eingreifen können und dessen Regeln für sie nicht transparent sind.

Der Beitrag von Andrea Mennicken und Martin Kornberger erklärt, wie die Online-Plattformen der Sharing-Ökonomie das oben angesprochene Koordinationsproblem lösen und welche Rolle evaluative Infrastrukturen dabei spielen. Die auf der Basis der aufgezeichneten Daten und Kundenbewertungen erstellten Ratings sind für die Nutzer ein wichtiger Orientierungspunkt, der dazu verhilft, die Unsicherheit über die Qualität der Dienstleistung zu reduzieren. Gleichzeitig verhilft das Ratingsystem dazu, Vertrauen zu generieren, ohne dass digitale Transaktionen nicht zustande kämen. Darüber hinaus zeigt der Beitrag, dass evaluative digitale Infrastrukturen nicht bloß vorhandene Beziehungen stabilisieren, sondern ständig neue Konnektivitäten generieren und dadurch jene Strukturen erzeugen, aus denen digitale Plattformen letztlich bestehen.

Die Forschung zu den Praktiken des Vergleichens, Bewertens, Kategorisierens und Quantifizierens findet weitgehend arbeitsteilig statt: Jeder Begriff hat sein eigenes Revier. Der vorliegende Sammelband will dieser Fragmentierung entgegenwirken und dazu anregen, systematischer über ihre empirischen und begrifflichen Verbindungen nachzudenken. Die Beiträge veranschaulichen, welche neuen Einsichten man gewinnt, wenn man die vier Verfahren nicht künstlich separiert, sondern sie in ihrem Zusammenwirken analysiert. Die Autorinnen und Autoren waren vor die Aufgabe gestellt, ihren Gegenstand nicht nur unter einem Aspekt zu betrachten, sondern zu er-

proben, was sich entdecken lässt, wenn man alle vier Praktiken in den Blick nimmt, auch jene, an die man zunächst vielleicht nicht denken würde. Die Fallbeispiele stammen aus ganz unterschiedlichen empirischen Kontexten. Sie befassen sich mit Kolonialstatistiken und Kunstrankings, mit dem frühen Staatenvergleich und Dating-Apps, mit der Methodik des interkulturellen Vergleichs und mit prädiktiven Algorithmen. Die Disparatheit der Beispiele macht die Anlage des Bandes zwar komplex, sie hat aber den Vorteil, seine Grundannahmen an einem breiten Spektrum zeitlich und sachlich auseinanderliegender Fälle testen zu können.

Die Verbindung zwischen den verschiedenen Beiträgen wird einerseits über die Fragestellung des Bandes hergestellt: Welche Bedeutung haben Vergleich, Kategorisierung, Quantifizierung und Bewertung in den einzelnen Untersuchungsfällen und wie greifen sie ineinander? Da die Fallbeispiele aus historisch unterschiedlichen Epochen stammen und vom späten 17. Jahrhundert bis in die Gegenwart reichen, geben sie gleichzeitig Auskunft darüber, ob und wie sich der Stellenwert der einzelnen Verfahren und deren wechselseitiges Verhältnis verändert haben. Eine zweite Verbindung ergibt sich aus dem Vorschlag, Statistiken, Monitoringsysteme, Rankings und digitale Technologien als soziale Beobachtungsformate zu interpretieren, also als Formate, die regelmäßig Daten generieren und in diesen Daten nach Ordnungsmustern suchen. Über das Konzept des Beobachtungsformats wird ein gemeinsames Bezugsproblem formuliert, das die Beiträge zusammenhält, nämlich die Frage, auf welche Weise die einzelnen Formate beobachten und wie die vier Verfahren in den einzelnen Fällen miteinander verknüpft sind. Aus den Gemeinsamkeiten und Unterschieden lässt sich nicht nur etwas über ihre Verbindung lernen, sondern auch über den Wandel der gesellschaftlichen Beobachtungstechniken und ihrer kulturellen und sozialen Voraussetzungen.

Danksagung Ich danke Hannah Bennani, Marion Müller, Martin Petzke und Tobias Werron für ihre präzisen und kritischen Kommentare.

Funding Open access funding provided by University of Luzern

Open Access Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

Literatur

- Ajunwa, Ifeoma, und Daniel Greene. 2019. Platforms at Work: Automated Hiring Platforms and Other New Intermediaries in the Organization of Work. *Work and Labor in the Digital Age. Research in the Sociology of Work* 33:61–91.
- Arlinghaus, Franz, und Peter Schuster. Hrsg. 2020. *Rank oder Ranking? Dynamiken und Grenzen des Vergleichs in der Vormoderne*. Konstanz: UVK Verlagsgesellschaft.
- Aspers, Patrick. 2018. Forms of uncertainty reduction. Decision, valuation, and contest. *Theoretical Sociology* 47:33–47.
- Beckert, Jens. 2007. Die soziale Ordnung von Märkten. In *Märkte als soziale Strukturen*, Hrsg. Jens Beckert, Rainer Diaz-Bone und Heiner Ganßmann, 43–62. Frankfurt a. M.: Campus.
- Beckert, Jens, und Christine Musselin (Hrsg.). 2013. *Constructing Quality: The Classification of Goods in the Market*. Oxford: Oxford University Press.
- Beer, David. Hrsg. 2017. Introduction: The Social Power of Algorithms. In *The Social Power of Algorithms*, Hrsg. David Beer. Information, Communication & Society 20(1):1–13. London: Routledge.
- Behrlich, Lars. 2016. *Die Berechnung der Glückseligkeit. Statistik und Politik in Deutschland und Frankreich im späten Ancien Régime*. Ostfildern: Thorbecke.
- Bennani, Hannah. 2017. *Die Einheit der Vielfalt. Zur Institutionalisierung der globalen Kategorie „indigene Völker“*. Frankfurt a. M.: Campus.
- Bennani, Hannah, und Marion Müller. 2018. „Making up people“ globally. Die Institutionalisierung globaler Personenkategorien am Beispiel Indigener Völker und Menschen mit Behinderungen. *Zeitschrift für Soziologie* 47:306–331.
- Bennani, Hannah, und Marion Müller. 2020. Globale Personenkategorien und die Pluralisierung der Menschenrechte. Zur „Vermenschenrechtlichung“ indigener Völker und Menschen mit Behinderungen. In *Global beobachten und vergleichen. Soziologische Analysen zur Weltgesellschaft*, Hrsg. Hannah Bennani, Martin Bühler, Sophia Cramer und Andrea Glauser, 243–282. Frankfurt a. M.: Campus.
- Bennani, Hannah, Martin Bühler, Sophia Cramer und Andrea Glauser. Hrsg. 2020. *Global beobachten und vergleichen. Soziologische Analysen zur Weltgesellschaft*. Frankfurt a. M.: Campus.
- Boelaert, Julien, und Etienne Ollion. 2018. The Great Regression. Machine Learning, Econometrics, and the Future of Quantitative Social Sciences. *Revue française de sociologie* 59:475–506.
- Bowker, Geoffrey C., und Susan Leigh Star. 2000. *Sorting Things Out: Classification and its Consequences*. Cambridge: MIT Press.
- Brekhus, Wayne. 1998. A Sociology of the Unmarked: Redirecting Our Focus. *Sociological Theory* 16:34–51.
- Buckermann, Paul. 2020. Warum eine Rangliste? Die Vorstellungen von Kunst hinter dem Künstler_innenranking *Kunstkompass*, 1970–2018. In *Global beobachten und vergleichen. Soziologische Analysen zur Weltgesellschaft*, Hrsg. Hannah Bennani, Martin Bühler, Sophia Cramer und Andrea Glauser, 349–380. Frankfurt a. M.: Campus.
- Bühler, Martin. 2019. *Von Netzwerken zu Märkten. Die Entstehung eines globalen Getreidemarktes*. Frankfurt a. M.: Campus.
- Burbank, Jane, und Frederick Cooper. 2010. *Empires in World History. Power and the Politics of Difference*. Princeton: Princeton University Press.
- Chakrabarty, Dipesh. 2007. *Provincializing Europe. Postcolonial Thought and Historical Difference*. Princeton: Princeton University Press.
- Christopher, Anthony J. 2008. The quest for a census of the British Empire c.1840–1940. *Journal of Historical Geography* 34:268–285.
- Cordell, Dennis D., Karl Ittmann und Gregory H. Maddox. 2010. Counting Subjects. Demography and Empire. In *The Demographics of Empire: The Colonial Order and the Creation of Knowledge*, Hrsg. Karl Ittmann, Dennis D. Cordell und Gregory H. Maddox, 1–21. Athens: Ohio University Press.
- Desrosières, Alain. 2005. *Die Politik der großen Zahlen. Eine Geschichte der statistischen Denkweise*. Heidelberg: Springer.
- Dewey, John. 1939. *Theory of Valuation*. International Encyclopedia of Unified Science, Vol. II, Nr. 4.
- Diaz-Bone, Rainer, und Emmanuel Didier. Hrsg. 2016. *Conventions and Quantification*. Historical Social Research 41.
- Dumont, Louis M. 1980. *Homo Hierarchicus. The Caste System and Its Implications*. Chicago: Chicago University Press.

- Durkheim, Emile, und Marcel Mauss. 1987/1902. Über einige primitive Formen der Klassifikation. Ein Beitrag zur Erforschung der kollektiven Vorstellungen. In *Emile Durkheim. Schriften zur Soziologie der Erkenntnis*, Hrsg. Hans Joas, 169–256. Frankfurt a. M.: Suhrkamp.
- Eggers, Michael. Hrsg. 2011. *Von Ähnlichkeiten und Unterschieden. Vergleich, Analogie und Klassifikation in Wissenschaft und Literatur (18./19. Jahrhundert)*. Heidelberg: Universitätsverlag.
- Eggers, Michael. 2016. Vergleichendes Erkennen. Zur Wissenschaftsgeschichte und Epistemologie des Vergleichs und zur Genealogie der Komparatistik. Heidelberg: Universitätsverlag Winter.
- Epple, Angelica, Walter Erhard und Johannes Grave. Hrsg. 2020. *Practices of Comparing. Towards a New Understanding of a Fundamental Human Practice*. Bielefeld: Transcript.
- 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 Mitchell L. Stevens. 2008. A Sociology of Quantification. *European Journal of Sociology* 49:401–436.
- Foucault, Michel. 1974. *Die Ordnung der Dinge*. Frankfurt a. M.: Suhrkamp.
- Foucault, Michel. 2004. *Sicherheit, Territorium, Bevölkerung. Geschichte der Gouvernementalität I*. Frankfurt a. M.: Suhrkamp.
- Fourcade, Marion. 2012. The Vile and the Noble: On the Relation between Natural and Social Classifications in the French Wine World. *The Sociological Quarterly* 53:524–545.
- Fourcade, Marion. 2016. Ordinalization. Lewis A. Coser Memorial Award for Theoretical Agenda Setting 2014. *Sociological Theory* 34:175–195.
- Fourcade, Marion, und Kieran Healy. 2013. Classification Situations: Life-chances in the Neoliberal Era. *Accounting, Organizations, and Society* 38:559–572.
- Fourcade, Marion, und Kieran Healy. 2016. Seeing like a market. *Socioeconomic Review* 15:9–29.
- Goffman, Erving. 1975. *Stigma. Über Techniken der Bewältigung beschädigter Identität*. Frankfurt a. M.: Suhrkamp.
- Hacking, Ian. 1982. Biopower and the Avalanche of Printed Numbers. *Humanities and Society* 5:279–295.
- Hacking, Ian. 1991. How Should We Do the History of Statistics? In *The Foucault Effect. Studies in Governmentality*, Hrsg. Graham Burchell, Colin Gordon und Peter Miller, 181–195. Chicago: The University of Chicago Press.
- Heintz, Bettina. 2010. Numerische Differenz. Überlegungen zu einer Soziologie des (quantitativen) Vergleichs. *Zeitschrift für Soziologie* 39:162–181.
- Heintz, Bettina. 2012. Welterzeugung durch Zahlen Modelle politischer Differenzierung in internationalen Statistiken, 1948–2010. In *Welterzeugung durch Bilder*, Hrsg. Cornelia Bohn, Arno Schubbach und Leo Wansleben, 7–39. Stuttgart: Lucius & Lucius.
- Heintz, Bettina. 2016. „Wir leben im Zeitalter der Vergleichung.“ Perspektiven einer Soziologie des Vergleichs. *Zeitschrift für Soziologie* 39:162–181.
- Hirschauer, Stefan. 2014. Un/doing Differences. Die Kontingenz sozialer Zugehörigkeiten. *Zeitschrift für Soziologie* 43:170–191.
- Hirschauer, Stefan. 2017. *Un/doing differences. Praktiken der Humandifferenzierung*. Weilerswist: Velbrück.
- von Hirschhausen, Ursula. 2011. People that Count. The Imperial Census in Nineteenth- and Early Twentieth-Century Europe and India. In *Comparing Empires. Encounters and Transfers in the Long Nineteenth Century*, Hrsg. Jörn Leonhard und Ursula von Hirschhausen, 145–170. Göttingen: Vandenhoeck & Ruprecht.
- Igo, Sarah E. 2007. *The Averaged American: Surveys, Citizens, and the Making of a Mass Public*. Cambridge: Harvard University Press.
- Karpik, Lucien. 2010. *Valuing the Unique: The Economics of Singularities*. Princeton: Princeton University Press.
- Keller, Felix. 2001. *Archäologie der Meinungsforschung*. Konstanz: UVK Verlagsgesellschaft.
- Kellog, 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.
- Kette, Sven. 2020. „Computer says no.“ Konsequenzen der Algorithmisierung von Entscheidungsprozessen. Manuskript. Universität Luzern.
- Knöbl, Wolfgang. 2007. *Die Kontingenz der Moderne. Wege in Europa, Asien und Amerika*. Frankfurt a. M.: Campus.
- Lamont, Michelle. 2012. Toward a Comparative Sociology of Valuation and Evaluation. *Annual Review of Sociology* 38:201–221.
- Leonhard, Jörn, und Ursula von Hirschhausen (Hrsg.). 2011. *Comparing Empires. Encounters and Transfers in the Long Nineteenth Century*. Göttingen: Vandenhoeck & Ruprecht.

- Luhmann, Niklas. 1995a. Kultur als historischer Begriff. In *Gesellschaftsstruktur und Semantik, Bd. 4*, 31–54. Frankfurt a. M.: Suhrkamp.
- Luhmann, Niklas. 1995b. Jenseits von Barbarei. In *Gesellschaftsstruktur und Semantik, Bd. 4*, 138–150. Frankfurt a. M.: Suhrkamp.
- Luhmann, Niklas. 2002. *Das Erziehungssystem der Gesellschaft*. Frankfurt a. M.: Suhrkamp.
- Lyon, David. 2002. Everyday Surveillance: Personal data and social classifications. *Information, Communication & Society* 5:242–257.
- Lyon, David. 2007. *Surveillance Studies: An Overview*. Cambridge, UK: Polity Press.
- MacKenzie, Donald, und Yuval Millo. 2003. Constructing a Market, Performing Theory: The Historical Sociology of a Financial Derivatives Exchange. *American Journal of Sociology* 109:104–145.
- Marx, Gary T. 2015. *Surveillance Studies*. *International Encyclopedia of the Social and Behavioral Sciences* 43:733–741.
- Matthes, Joachim. 1992: The Operation Called „Vergleichen“. In *Zwischen den Kulturen? Die Sozialwissenschaften vor dem Problem des Kulturvergleichs*. Hrsg. Joachim Matthes, 75–99. Göttingen: Otto Schwartz & Co.
- Mau, Steffen. 2017. *Das metrische Wir. Über die Quantifizierung des Sozialen*. Frankfurt a. M.: Suhrkamp.
- 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.
- Merry, Sally Engle. 2016. *The Seductions of Quantification. Measuring Human Rights, Gender Violence, and Sex Trafficking*. Chicago: Chicago University Press.
- Miller, Peter. 2001. Governing by Numbers. Why Calculative Practices Matter. *Social Research* 28:379–390.
- Mosseri, Adam. 2019. „Wir wollen nicht, dass Instagram ein Wettbewerb ist.“ *FAZ*, 1.7.19, <https://www.faz.net/aktuell/wirtschaft/digitec/instagram-chef-adam-mosseri-im-interview-ueber-sozialen-druck-16261737.html?premium=0x7c74b85e989f3a07fc975cd6672c9c71>.
- Müller, Marion. 2003. *Geschlecht und Ethnie. Historischer Bedeutungswandel und interaktive Konstruktion und Interferenzen*. Wiesbaden: Westdeutscher Verlag.
- Nassehi, Armin. 2019. *Muster. Theorie der digitalen Gesellschaft*. München: Beck.
- Navis, Chad, und Mary Ann Glynn. 2010. How New Market Categories Emerge. Temporal Dynamics of Legitimacy, Identity and Entrepreneurship in Satellite Radio, 1990–2005. *Administrative Science Quarterly* 55:439–471.
- Nichols, Charles K. 1942. The Statistical Work of the League of Nations in Economic, Financial and Related Fields. *Journal of the American Statistical Association* 37:336–342.
- Ormod, Susan. 1994. „Let's Nuke the Dinner.“ Discursive Practice of Gender in the Creation of a New Cooking Process. In *Bringing Technology Home. Gender and Technology in a Changing Europe*. Hrsg. Cynthia Cockburn und Ruza Fürst-Dilic, 42–58. Buckingham: Open University Press.
- Osterhammel, Jürgen. 1998. *Die Entzauberung Asiens. Europa und die asiatischen Reiche im 18. Jahrhundert*. München: Beck.
- Parsons, Talcott. 1966. *Societies. Evolutionary and Comparative Perspectives*. Englewood Cliffs, N.J.: Prentice-Hall.
- Peetz, Thorsten, Enno Aljets, Frank Meier und Désirée Weibel (Hrsg.). 2016. *Soziologie der Bewertung*. Berliner Journal für Soziologie 26, issue 3–4.
- Petzke, Martin. 2020. Measures and Their Countermeasures: Reflexivity and Second-Order Reactivity in Quantifying Immigrant Integration. *Sociological Forum* <https://onlinelibrary.wiley.com/doi/full/10.1111/socf.12668>.
- Porter, Theodore W. 1986. *The Rise of Statistical Thinking. 1820–1900*. Princeton: Princeton University Press.
- Porter, Theodore W. 1992. Objectivity as Standardization: The Rhetoric of Impersonality in Measurements, Statistics, and Cost-Benefit-Analysis. In *Rethinking Objectivity*, Hrsg. Allan Megill, 197–237. Durham: Duke University Press.
- Porter, Theodore W. 1995. *Trust in Numbers. The Pursuit of Objectivity in Science and Public Life*. Princeton: Princeton University Press.
- Pridmore, Jason, und Lalu Elias Hämäläinen. 2017. Market Segmentation in (In)Action: Marketing and “Yet to Be Installed” Role of Big and Social Media Data. *Historical Social Research* 42:103–122.
- Quetelet, Adolphe. 1921[1869]. *Soziale Physik oder Abhandlungen über die Entwicklung der Fähigkeiten des Menschen*. Jena: Verlag von Gustav Fischer.
- Reckwitz, Andreas. 2017. *Die Gesellschaft der Singularitäten. Zum Strukturwandel der Moderne*. Frankfurt a. M.: Suhrkamp.
- Reichardt, Sven. 2016. Einführung: Überwachungsgeschichte(n). *Geschichte und Gesellschaft* 42:5–33.

- Renard, Léa. 2021. *The Statistical Construction of Otherness: Colonial, National and Migratory Classification Principles in France and Germany (1880–2010)*. Frankfurt a. M.: Campus (im Erscheinen).
- Rosenblat, Alex. 2018. *Überland. How Algorithms are Rewriting the Rules of Work*. Oakland: University of California Press.
- Ruppert, Evelyn S. 2011. Making Populations: From Censuses to Metrics. In *Sichtbarkeitsregime*, Hrsg. Leon Hempel, Susanne Krasmann und Ulrich Bröckling, 157–173. Wiesbaden: VS Verlag.
- Rusnock, Andrea. 1995. Quantification, Precision, and Accuracy: Determination of Population in the Ancien Régime. In *The Values of Precision*. Hrsg. M. Norton Wise, 17–38. Princeton: Princeton University Press.
- Sauder, Michael, und Wendy N. Espeland. 2006. Strength in Numbers. The Advantage of Multiple Rankings. *Indiana Law Journal* 81:205–227.
- Schiebinger, Londa. 1993. Anatomie der Differenz. „Rasse“ und Geschlecht in der Naturwissenschaft des 18. Jahrhunderts. *Feministische Studien* 11:48–64.
- Seaver, Nick. 2019. Captivating Algorithms: Recommender Systems as Traps. *Journal of Material Culture* 24:421–436.
- Simmel, Georg. 1995[1901]. Die beiden Formen des Individualismus. In *Georg Simmel Aufsätze und Abhandlungen 1901–1908, Bd. 1*, Hrsg. Rüdiger Kramme, Angela Rammstedt und Otthein Rammstedt, 49–56. Frankfurt a. M.: Suhrkamp.
- Simmel, Georg. 1999[1917]. Grundfragen der Soziologie. In *Georg Simmel. Gesamtausgabe Bd. 16*. Hrsg. Gregor Fitzi und Otthein Rammstedt, 59–160. Frankfurt a. M.: Suhrkamp.
- Skocpol, Theda. 1979. *States and Social Revolutions: A Comparative Analysis of France, Russia, and China*. Cambridge: Cambridge University Press.
- Slack, Paul. 2015. *The Invention of Improvement: Information and Material Progress in Seventeenth-Century England*. Oxford: Oxford University Press.
- Speich Chassé, Daniel. 2013. *Die Erfindung des Bruttosozialprodukts. Globale Ungleichheit in der Wissensgeschichte der Ökonomie*. Göttingen: Vandenhoeck & Ruprecht.
- Speich Chassé, Daniel. 2020. Der Staatsvergleich in historischer Perspektive. Warum, seit wann und wie werden politische Mächte miteinander verglichen? In *Global beobachten und vergleichen. Soziologische Analysen zur Weltgesellschaft*, Hrsg. Hannah Bennani, Martin Bühler, Sophia Cramer und Andrea Glauser, 79–112. Frankfurt a. M.: Campus.
- Spoerhase, Carlos. 2014. Das Maß der Potsdamer Garde. Die ästhetische Vorgeschichte des Rankings in der europäischen Literatur- und Kunstkritik des 18. Jahrhunderts. *Jahrbuch der Deutschen Schillergesellschaft* 58:90–126.
- Steinmetz, George. 2004. Odious Comparisons: Incommensurability, the Case Study, and “Small N’s”. *Sociological Theory* 22:371–400.
- Steinmetz, George. 2020. Soziologie und Kolonialismus. Über die Geburt der Soziologie aus der kolonialen Erfahrung. *Mittelweg* 36:57–78.
- Steinmetz, Willibald. 2015. „Vergleich“ – eine begriffsgeschichtliche Skizze. In *Die Welt beobachten. Praktiken des Vergleichens*, Hrsg. Angelika Epple und Walter Erhart, 85–134. Frankfurt a. M.: Campus.
- Steinmetz, Willibald. Hrsg. 2019a. *The Force of Comparison. A New Perspective on Modern European History and the Contemporary World*. Oxford: Berghahn Books.
- Steinmetz, Willibald. 2019b. Above/Below, Better/Worse or Simply Different? Metamorphoses of Social Comparison 1600–1900. In *The Force of Comparison. A New Perspective on Modern European History and the Contemporary World*, Hrsg. Willibald Steinmetz, 80–112. Oxford: Berghahn Books.
- Stoler, Ann Laura. 1995. *Race and the Education of Desire: Foucault’s History of Sexuality and the Colonial Order of Things*. Durham: Duke University Press.
- Stollberg-Rilinger, Barbara. 2014. Logik und Semantik des Ranges in der Frühen Neuzeit. In *Konkurrenz in der Geschichte*, Hrsg. Ralph Jessen, 197–227. Frankfurt a.M.: Campus.
- Tag, Miriam. 2020. *The Making of a Global Issue: Early Childhood at the United Nations*. Dissertation. Universität Luzern.
- Tenbruck, Friedrich H. 1989. Gesellschaftsgeschichte oder Weltgeschichte? *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 41:417–439.
- Touchelay, Béatrice. 2019. British and French Colonial Statistics: Development by Hybridization from the Nineteenth to the Mid-Twentieth Centuries. In *British and French Colonialism in Africa, Asia and the Middle East. Connected Empires across the Eighteenth to the Twentieth Centuries*, Hrsg. James R. Fichter, 249–274. Cham: Palgrave Macmillan.
- Turow, Joseph. 2011. *The Dayly You. How the New Advertising Industry is Defining Your Identity and Worth*. New Haven: Yale University Press.

- Vanderstraeten, Raf. 2006. Soziale Beobachtungsraster. Eine wissenssoziologische Analyse statistischer Klassifikationsschemata. *Zeitschrift für Soziologie* 35:193–211.
- Ward, Michael. 2004. *Quantifying the World. UN Ideas and Statistics*. Bloomington: Indiana University Press.
- Weed, Keith. 2017. The Future of Consumer Marketing? <https://www.thinkwithgoogle.com/consumer-insights/consumer-trends/unilever-consumer-marketing-segmentation/>
- Werron, Tobias. 2012. Worum konkurrieren Nationalstaaten? Zu Begriff und Geschichte der Konkurrenz um „weiche“ globale Güter. *Zeitschrift für Soziologie* 41:338–355.
- Wobbe, Theresa, und Léa Renard. 2017. The category of “family workers” in International Labour Organization statistics (1930s–1980s): a contribution to the study of globalized gendered boundaries between household and market. *Journal of Global History* 12:340–360.
- Wong, Karen Li Xan, und Amy Shields Dobson. 2019. We’re just data: Exploring China’s social credit system in relation to digital platform ratings cultures in Westernised democracies. *Global Media and China* 4:220–232.
- Zelizer, Viviana. 2006. Money, Power, and Sex. *Yale Journal of Law and Feminism* 18:303–315.
- Zerubavel, Eviatar. 1996. Lumping and splitting: Notes on social classification. *Sociological Forum* 11:421–433.
- Zerubavel, Eviatar. 2018. *Taken for Granted. The Remarkable Power of the Unremarkable*. Princeton: Princeton University Press.
- Zuckerman, Ezra W. 1999. The Categorical Imperative: Securities Analysts and the Illegitimacy Discount. *American Journal of Sociology* 104:1398–1438.

Bettina Heintz 1949, Dr. phil., Sen. Prof. für Soziologie, Universität Luzern. Forschungsgebiete: Soziologie des Vergleichs und der Quantifizierung, Soziologie der Weltgesellschaft, Menschenrechte, soziologische Theorie. Veröffentlichungen: „Wir leben im Zeitalter der Vergleichung.“ Perspektiven einer Soziologie des Vergleichs, *Zeitschrift für Soziologie* 45, 2016; Good – Better – Best: Comparisons and the Power of Ranking Orders, in: W. Steinmetz, (Hrsg.), *The Force of Comparison: A New Perspective on Modern European History and the Contemporary World*. New York 2019; Menschenrechte in der Weltgesellschaft. Deutungswandel und Wirkungsweise eines globalen Leitwertes. Frankfurt a. M. (als Hrsg., mit B. Leisering).