

Zwischen Konsens und Kontroverse

Grenzarbeit in der Geschichte der Stichprobe

Lukas Griessl

1. Einleitung

Einen Tag vor Beginn der Fußball-Weltmeisterschaft 2018 tauchten Bilder der deutschen Nationalspieler Mesut Özil und İlkay Gündoğan mit dem türkischen Präsidenten Recep Tayyip Erdoğan auf. In Deutschland führten diese zu viel Kritik von Seiten der Medien, Fans und Politik und sie waren der Auftakt für eine Kontroverse über die Loyalität der beiden Spieler zu Deutschland und deren Platz in der Nationalmannschaft.¹ Wenn der Fototermin und die dadurch ausgelöste Kontroverse sicherlich Politik- und Sportgeschichte geschrieben haben, so nimmt diese Episode auch einen Platz in der Geschichte der Meinungsumfragen ein: Einen Tag nachdem das Foto aufgenommen wurde, veröffentlichte Focus Online einen Artikel mit der Überschrift *Repräsentative Civey-Umfrage für FOCUS Online – Nach Erdogan-Fotos: Mehrheit für Ausschluss von Özil und Gündoğan aus Nationalelf*². Laut dieser als repräsentativ deklarierten Umfrage wollten 80 % der über 18-Jährigen in Deutschland Mesut Özil und İlkay Gündoğan nicht mehr in der deutschen Nationalmannschaft sehen. Einer anderen, von Forsa durchgeführten Umfrage zufolge sah die Lage etwas anders aus: Demnach sahen 71 % der Deutschen in dem Foto keinen Grund, die beiden Spieler nicht für die Weltmeisterschaft zu nominieren.³ Im Anschluss an die Veröffentli-

-
- 1 Dieser Aufsatz ist Teil meines Dissertationsprojektes an der University of Essex. Für hilfreiche Anmerkungen und Diskussionen bedanke ich mich bei Max Gropper, Sebastian Klimasch, Carsten Ohlrogge und Emilie Köhler sowie besonders dem*der anonymen Gutachter*in.
 - 2 Focus Online. 2018. *Repräsentative Civey-Umfrage für FOCUS Online. Nach Erdogan-Fotos: Mehrheit für Ausschluss von Özil und Gündoğan aus Nationalelf*. Online verfügbar: https://www.focus.de/politik/deutschland/repraesentative-civey-umfrage-fuer-focus-online-nach-erdogan-fotos-klare-mehrheit-will-oezil-und-guendogan-aus-nationalelf-ausschliessen_id_8934133.html [25.01.2022]
 - 3 welt.de. 2018. *RTL-Forsa-Umfrage: 61 Prozent gegen Treffen von Özil und Gündogan mit Erdogan*. Online verfügbar unter: <https://www.welt.de/newsticker/sport-news/article176424191/Fussball-RTL-Forsa-Umfrage-61-Prozent-gegen-Treffen-von-Oezil-und-Guendogan-mit-Erdogan.html> [25.01.2022]

chung der ersten Umfrage reichten die drei deutschen Meinungsforschungsinstitute Forsa, Infas und die Forschungsgruppe Wahlen eine Beschwerde gegen Focus Online beim deutschen Presserat ein. In der Begründung hieß es, dass »[b]ei der zugrundeliegenden Umfrage [...] ein Befragungsdesign verwendet [wurde], das nach den allgemein anerkannten wissenschaftlichen Kriterien der empirischen Sozialforschung *ganz grundsätzlich* nicht geeignet ist, »repräsentative« Ergebnisse zu liefern«⁴ (Hervorhebung durch den Verf.). Laut dieser Beschwerde überschritt Civey mit seiner Methodik also eine Grenze, indem sie das verletzte, was gemäß den Beschwerdeführern den allgemein anerkannten wissenschaftlichen Kriterien der empirischen Sozialforschung entspricht. Die Beschwerde, so lässt sie sich interpretieren, richtete sich dabei jedoch nicht nur gegen das Unternehmen Civey, sondern grundsätzlich gegen das verwendete Befragungsdesign. Dies wirft die Frage auf, wie der Konsens darüber, was als allgemein anerkannt gilt, historisch gewachsen ist. Die Aufarbeitung dieser Genealogie eröffnet die Möglichkeit, die Frage nach ebensolchen methodischen Grenzziehungen und deren rhetorischen Manifestationen als soziale Konstruktion zu erörtern.

Im vorliegenden Aufsatz beschäftige ich mich in diesem Sinne mit der Entwicklung der Stichprobenziehung und den rhetorischen Manifestationen methodischer Kontroversen und Konsensen. Zuerst stelle ich das *Konzept der »Grenzarbeit«* vor, um dieses für die anschließende historische Analyse nutzbar zu machen. Daran anknüpfend zeichne ich die *Geschichte der Stichprobenziehung* nach, mit einem besonderen Augenmerk auf rhetorische Figuren der Grenzarbeit. Hierbei erläutere ich zunächst den *epistemischen Kontext des 18. und 19. Jahrhunderts*, um in einem nächsten Schritt die *Realisierung einer innerdisziplinären Übereinkunft* über die grundsätzliche Akzeptanz von Stichproben im ersten Viertel des 20. Jahrhunderts zu beschreiben. In einem nächsten Schritt wende ich mich der daran anschließenden *Diskussion um die richtige Methode der Stichprobenziehung* zu, welche, wie im nächsten Schritt dargestellt, etwa Mitte des 20. Jahrhunderts entschieden wurde. Dabei argumentiere ich, dass es sich bei der in der Einleitung erwähnten Kontroverse strukturell um eine ähnliche Auseinandersetzung handelt wie bei jenen, die bereits in der ersten Hälfte des 20. Jahrhunderts geführt wurden. Abschließend gehe ich kurz auf die *kommerzielle Nutzung empirischer Forschung* ein, und skizziere ein mögliches Feld zukünftiger Grenzarbeit im Bereich der Umfrageforschung.

4 marktforschung.de. 2018. *Beschwerde beim Presserat wegen Civey-Umfrage*. Online verfügbar unter: [https://www.marktforschung.de/aktuelles/marktforschung/beschwerde-beim-presserat-wegen-civey-umfrage/\[25.01.2022\]](https://www.marktforschung.de/aktuelles/marktforschung/beschwerde-beim-presserat-wegen-civey-umfrage/[25.01.2022])

2. Methodenstreit als Grenzarbeit

Der Begriff der »Grenzarbeit«, den ich im Folgenden verwende, ist eine Übersetzung von »boundary-work«, einem Konzept in den englischsprachigen Science Studies. Grob zusammenfassen lässt sich dies als ein Aufziehen von Grenzen um die (und innerhalb der) Wissenschaften durch die in ihnen Arbeitenden, welches der Herstellung eigener Autorität dient. Boundary-work beschreibt nach Gieryn (1983) einen »ideologischen Stil«, der darin besteht, dass Wissenschaftler*innen ein öffentliches Bild von Wissenschaft zeichnen, indem sie es positiv von anderen Aktivitäten abheben. Gieryn ging es in seiner Arbeit um die Frage, wie Wissenschaft epistemische Autorität gegenüber anderen wissensproduzierenden Bereichen wie etwa der Religion erlangen konnte. Das Konzept der Grenzarbeit stellt dabei ein »dediziert konstruktivistisches« (Prietl und Ziegler 2016, S. 9) Analyseinstrument dar, welches in den Blick nimmt, »wie und welche Eigenschaften Wissenschaft selektiv zugeschrieben werden, um sie als Produzentin wahren Wissens auszuweisen« (ebd.). Der Fokus wissenschaftssoziologischer Betrachtung liegt dementsprechend auf der Darstellung und Wahrnehmung von Wissenschaft und besonders auf den »rhetorischen Prozesse[n], in denen wissenschaftlichen Praktiken und Akteur_innen bestimmte Eigenschaften zugeschrieben werden und in denen sie von anderen intellektuellen Aktivitäten abgegrenzt werden, um sie als wissenschaftlich auszuweisen« (ebd. S. 10). Wissenschaftliche Autorität und Autonomie sind demnach ein Ergebnis ebendieser Praktiken der Grenzziehung.

Während Gieryn mit seinem Konzept von Grenzarbeit auf die soziale Konstruktion einer Grenze zwischen Wissenschaft und Nicht-Wissenschaft abzielte, wurde dieses von Jasanoff in einem weiteren Sinne als »a communally approved drawing of lines between ›good‹ and ›bad‹ work (and, not trivially, between good and bad workers) within a single discipline, between different disciplines, and between ›science‹ and other forms of authoritative knowledge« (Jasanoff 1995, S. 53) weiterentwickelt. Nach Jasanoff beschreibt Grenzarbeit also nicht nur die Konstruktion und Aufrechterhaltung der Unterscheidung zwischen Wissenschaft und Nicht-Wissenschaft, sondern auch Auseinandersetzungen um Autorität und Anerkennung innerhalb einer wissenschaftlichen Gemeinschaft. Als Vorgehen bietet sich hierfür Latours (1987) Charakterisierung der Forschungsweisen *Science in the Making* und *Ready Made Science* an. Während Wissenschaft im Modus des *Science in the Making* als ergebnisoffen und kontrovers beschrieben werden kann, wird sie im Modus der *Ready Made Science* als eine auf »fertige Tatsachen und Forschungen gerichtete Perspektive« (Gertenbach und Laux 2019, S. 40) beschrieben. Anstatt eine bestimmte Realität einer fertigen Wissenschaft zu beschreiben, ermutigt Latour dazu, stattdessen die im Entstehen begriffene Wissenschaft als Ausgangspunkt zu nutzen. Diese Perspektive ermöglicht einen Blick darauf, wie wissenschaftliches Wissen konstituiert wird und durch welche Prozesse Gewissheit und Konsens innerhalb der Wissenschaft entste-

hen. Grenzarbeit stellt dabei einen Teil der Praktiken dar, wie Konsens hergestellt und aufgelöst wird.

3. Eine kurze Geschichte der Stichprobenziehung

3.1 Gefährliche Spekulation

Bevölkerungszählungen und Befragungen gibt es schon erstaunlich lange. Im Gegensatz zur wohl bekanntesten Volkszählung der Geschichte, bei der sich Josef aus Galiläa im Neuen Testament noch in Person nach Bethlehem aufmachen musste, »auf dass er sich schätzen ließe«, werden in der Gegenwart sehr unterschiedliche Methoden angewendet. Die Idee, Aussagen über eine Bevölkerung anhand einer Stichprobe, also eines kleinen Teils der Bevölkerung zu treffen ist dabei vergleichsweise jung: Zwar wurden Stichproben schon im 17. Jahrhundert vorgeschlagen, ernsthaft diskutiert und anerkannt wurden sie allerdings erst im 20. Jahrhundert. Um den epistemischen Kontext nachzuvollziehen, also die disziplinären Überzeugungen, innerhalb derer sich Stichproben als anerkannte wissenschaftliche Methode entwickelten, ist es notwendig, sich zunächst mit dem Beginn der Stichprobenziehung zu beschäftigen. Nennenswert sind im 17. und 18. Jahrhundert die Arbeiten von John Graunt (1620–1674) und William Petty (1620–1683), sowie die von Pierre Simon Laplace (1749–1827). Laplace stellte 1785 der Akademie der Wissenschaften in Frankreich eine Methode zur Schätzung der Einwohnerzahl vor (vgl. Laplace 1786). Sein Vorschlag war, die Einwohnerzahl Frankreichs auf Basis der Geburtenregister einer Stichprobe von 30 Departements aus dem gesamten Gebiet Frankreichs zu extrapolieren. Laplace wollte das Verhältnis zwischen der Zahl der Einwohner*innen und der Zahl der Geburten in diesen Stichprobenregionen berechnen, um es dann mit der Zahl der Geburten im ganzen Land (welche bekannt waren) zu multiplizieren. Im Laufe des 19. Jahrhunderts geriet das Werk von Laplace und damit der ernsthafte Vorschlag, Stichproben zu ziehen, jedoch fast vollständig in Vergessenheit.

Als entscheidende Figur in der Geschichte der Stichprobe für den Übergang vom 18. in das 19. Jahrhundert kann der Belgier Adolphe Quételet (1796 – 1874) gesehen werden. Quételet kannte Laplace und seine Arbeiten und war zunächst von dessen Vorschlag der Stichprobenziehung beeindruckt. Er schlug sogar vor, den Zensus für das Königreich der Vereinigten Niederlande im Sinne Laplaces zu erstellen. Wie Stigler (1986, S. 164) betonte, machte er jedoch eine abrupte Kehrtwende, offenbar aufgrund eines Arguments des Baron de Keverberg, eines hohen Beamten, der ein fundamentales Problem in Laplaces Vorschlag sah. Dieser lehnte Laplaces Ansatz als »unvollständig und spekulativ« (ebd. S. 165) ab, da solche Stichproben seiner Ansicht nach niemals repräsentativ für die Bevölkerung sein könnten, welche hinsichtlich

der Geburten- und Sterberaten zu heterogen sei, um Aussagen aus einzelnen Stichproben auf das ganze Land übertragen zu können. Seine Sorge war, dass das Land in fast so viele Teile geteilt werden müsste, wie es Menschen gibt, was keinen Vorteil mehr hinsichtlich der eingesparten Arbeit mit sich bringen würde. In einem Brief an den Herzog von Sachsen-Coburg und Gotha schrieb Quételet schließlich, dass Laplaces indirekte Methode so weit wie möglich zu vermeiden sei. Der Hauptgrund für seine Ablehnung partieller Daten war, dass er ihre mangelnde Genauigkeit befürchtete: »Die Statistik ist nur durch ihre Genauigkeit wertvoll; ohne diese wesentliche Eigenschaft wird sie nichtig, ja sogar gefährlich, da sie zu Fehlern führt⁵« (Quételet 1846, S. 293 [Übers. L.G.]). Teilerhebungen galten als dubios und passten nicht in die Vorstellung von Wissenschaft im 19. Jahrhundert. Es lässt sich also bereits hier eine Form der Grenzarbeit beobachten, wodurch Teilerhebungen epistemische Autorität abgesprochen und nur Vollerhebungen wissenschaftliche Autorität zugesprochen wurde.

Diese epistemische Verfassung, welche dieser Grenzarbeit zugrunde lag, lässt sich paradigmatisch etwa in der Rhetorik von Frederic J. Mouat (1816 – 1897) nachvollziehen, dem Gründer des International Statistical Institute (ISI). Dieser schrieb in seinem Aufsatz über die Geschichte der Statistical Society of London: »[S]tatistical inquiry has introduced order, method, and precision, in the place of speculation, conjecture, and uncertainty« (Mouat 1885). Mouat kontrastiert hier die im 17. und 18. Jahrhundert aufkeimenden Stichprobenerhebungen mit dem im 19. Jahrhundert vorherrschenden Primat der Vollerhebung. Solche Charakterisierungen können ebenfalls als Form der Grenzarbeit gelesen werden: Vollerhebungen, charakterisiert durch Ordnung, Methode und Präzision werden von Stichproben, charakterisiert durch Spekulation, Vermutung und Ungewissheit getrennt. Hier wird klar aufgezeigt, dass Stichprobenerhebungen nicht in das wissenschaftliche Terrain gehören und damit nicht als autorisierte Quelle von Wissen gelten können. Innerhalb des von Mouat gegründeten ISI vollzog sich im ausgehenden 19. und beginnenden 20. Jahrhunderts jedoch ein Wandel, welcher zum einen durch das Aufbrechen einer Grenze, wie auch durch das Errichten einer neuen Grenze beschrieben werden kann.

3.2 Grenzverschiebungen zur Jahrhundertwende

Das ISI wurde im Jahr 1885 als Nachfolgeorganisation des International Statistical Congress (ISC) gegründet und diente der Vernetzung und dem Austausch unter Statistikern (das ISI bestand zunächst nur aus Männern). Das ISI wurde unter anderem aufgrund seiner strikten Aufnahmebedingungen zu einer exklusiven Organisa-

5 Übersetzt aus dem Französischen: »La statistique n'a de valeur que par son exactitude; sans cette qualité essentielle, elle devient nulle, dangereuse même puisqu'elle conduit à l'erreur.«

tion mit einem hohen Maß an Kohäsion. Lusinchi (2021) macht am Beispiel des ISI deutlich, dass solche kohäsiven Gemeinschaften oft abweichende Meinungen ablehnen. Die Debatten am ISI zwischen 1895 und 1903 widmeten sich oft der Frage, ob »man das Ganze (also Vollerhebungen) in legitimer Weise durch einen Teil (das heißt durch Teilerhebungen) ersetzen kann« (Desrosières 2005, S. 236).

Als zentrale Figur dabei gilt der Norweger Anders Nicolai Kiær (1838–1919), Gründer und erster Leiter der Statistisk sentralbyrå, der zentralen Statistikbehörde Norwegens. Dieser nannte den von ihm entwickelten Ansatz *repräsentative Methode*, stellte diese zunächst im Jahr 1895 bei einem ISI-Treffen in Bern vor und entwickelte sie in den darauffolgenden Jahren bei weiteren Treffen weiter. In Anlehnung an Thomas Kuhn (2012) beschreibt Bethlehem (2009) diesen Zeitpunkt als »intellectually violent revolution«, als Einführung eines neuen Paradigmas im statistischen Denken. Zu dieser Zeit führte Kiær bereits seit mehr als 15 Jahren Stichprobenstudien in Norwegen durch und war davon überzeugt, dass eine Vollerhebung, wie es zu diesem Zeitpunkt Standard war, nicht immer notwendig sei. Kiær beschrieb seine Methode als »a partial exploration with observations on a large number of scattered localities, distributed over the whole territory so that they form a miniature of that whole« (Kiær in: Kruskal und Mosteller 1980, S. 176). Die Befragten sollten gemäß eines rationalen Verfahrens basierend auf Zensusdaten ausgewählt und überprüft werden. Die Stichprobe sollte also bewusst so gewählt werden, dass sie die Bevölkerung als ganze repräsentiere, daher auch der Name *repräsentative Methode*.

Als Kiær seine Methode im Rahmen des ISI-Treffens in Bern vorschlug, wurde sie von den Diskutanten mehrheitlich abgewiesen. Wie bereits dargestellt, war der epistemische Kontext zu dieser Zeit derart, dass Vollerhebungen die Praxis der Normalwissenschaft darstellten und Teilerhebungen von den meisten Statistiker*innen grundsätzlich abgelehnt wurden. Einer der Hauptkritiker Kiærs war der Statistiker Georg von Mayr (1841–1925), welcher die Methode als gefährlich und als etwas, wogegen man sich wehren müsse, beschrieb:

»[I]ch halte den Standpunkt, den man in seinem Werk findet[,] für sehr gefährlich. [...] Man kann die wahre Beobachtung von Tatsachen nicht durch Kalkulation ersetzen. [...] Es ist besonders gefährlich, repräsentative Stichproben inmitten einer Versammlung von Statistikern vorzuschlagen. [...] Wir müssen standhaft bleiben und sagen: keine Kalkulationen, wenn Beobachtungen gemacht werden können«⁶ (von Mayr in: IIS 1895, S. xciv [Übers. L.G.]).

6 Übersetzt aus dem Französischen : »[J]e regarde très dangereux le point de vue d'appréciation qui se trouve dans son travail. [...] On ne peut pas remplacer par le calcul l'observation réelle des faits. [...] C'est surtout dangereux de se déclarer pour ce système des investigations représentatives au sein d'une assemblée de statisticiens. [...] Mais il faut rester fermes et dire: pas de calcul là où l'observation peut être faite.«

Während von Mayr zumindest auf Kiærs Argumente einging, erwiderte etwa Dr. Heinrich Rauchberg (1860–1938), dass dieses Thema keinerlei Diskussion verdiene. In ähnlichem Duktus wehrte sich auch Edmund Wilhelm Milliet (1857–1931), welcher Kiærs repräsentative Methode von »seriöser« Statistik abgrenzte: »Ich glaube, es ist nicht richtig, der repräsentativen Methode (die letztlich nur ein Notbehelf sein kann) durch eine Stimme des Kongresses eine Bedeutung beizumessen, die ihr die seriöse Statistik niemals zugestehen würde⁷« (Milliet in: IIS 1895, S. xcvi [Übers. L.G.]). In diesen Äußerungen lassen sich erneut rhetorische Grenzziehungen finden, welche den epistemischen Kontext des 19. Jahrhunderts widerspiegeln. So wird Kiærs Vorschlag als »gefährlich« beschrieben, als etwas, wogegen man sich wehren müsse und etwas, das die »seriöse« Statistik niemals anerkennen würde. Unter den anderen Diskutanten während des Kongresses gab es allerdings auch andere Stimmen. So bezeichnete etwa Dr. Gustav Schmoller (1838–1917) Kiærs Gegner als »Statistikfanatiker« (Schmoller in: IIS 1895, S. xcvi [Übers. L.G.]), die seiner Meinung nach nicht einsehen wollen, dass, je komplexer statische Fragen werden, umso mehr die Notwendigkeit bestehe, sich nur einen Teil anzuschauen. Unter den Diskutanten war auch Louis Guillaume (1833–1924), der, obwohl er davon ausging, dass »das System der repräsentativen Erhebungen niemals an die Stelle der wissenschaftlichen Methode treten und die derzeit praktizierten Erhebungen vollständig ersetzen könnte« dafür warb, »keine kompromisslosen Doktrinäre«⁸ (Guillaume in: IIS 1895, S. ivxc [Übers. L.G.]) zu sein und Kiærs Vorschlag vorurteilsfrei zu begutachten. Obwohl Kiær auf heftigen Widerstand stieß, hatte er also auch Unterstützer und konnte in den Folgejahren seine Methode weiterentwickeln.

In der dargestellten Historie lassen sich zunächst rhetorische Grenzziehungen feststellen, wodurch das Terrain epistemischer Autorität innerhalb der Statistik abgesteckt wurde. Dabei zeigt sich eine besondere Form von Grenzarbeit: Im Sinne Gieryns stellt Grenzarbeit die soziale Konstruktion einer Grenze zwischen Wissenschaft und Nicht-Wissenschaft dar. Das trifft hier nur teilweise zu, vielmehr wird einer bestimmten wissenschaftlichen Praxis der Charakter der Wissenschaftlichkeit ab- und einer anderen Praxis zugesprochen. Es handelt sich also im Sinne Jasanoffs mehr um »a communally approved drawing of lines between ›good‹ and ›bad‹ work [...] within a single discipline« (Jasanoff 1995, S. 53), womit letztlich ebenso eine

7 Übersetzt aus dem Französischen : »Je crois qu'il n'est pas juste de donner par une voix du Congrès à la méthode représentative (qui enfin ne peut être qu'un expédient) une importance que la statistique sérieuse ne lui reconnaître jamais.«

8 Übersetzt aus dem Französischen : »Bien que le système des dénombrements représentatifs ne puisse jamais être substitué à la méthode scientifique et remplacer entièrement les dénombrements tels qu'ils sont pratiqués actuellement, il convient de montrer que nous ne sommes pas de doctrinaires intransigeants, mais que nous examinons sans préjugé les propositions [...].«

Grenze zwischen Wissenschaft und Nicht-Wissenschaft konstruiert wird. Die rhetorische Grenze zwischen guter und schlechter Arbeit innerhalb der Statistik hing maßgeblich davon ab, wie die jeweiligen Statistiker*innen zum Verhältnis des Ganzen zum Teil standen. Das Ganze durch einen Teil beschreiben zu wollen galt vielen als gefährlich und unseriös; diese Haltung sollte sich jedoch in den im folgenden Jahrhundert ändern. In diesem Schritt wird die historische Bedingtheit dieser Grenzen nochmal deutlich. Sie sind ein »historical phenomena, with a local and episodic (rather than transcendent) existence« (Gieryn 1999, S. x).

3.3 Zufällig oder bewusst?

Einen Unterstützer fand Kiær 1903 beim ISI-Kongress in Berlin in dem Franzosen Lucien March (1859 – 1933), der postulierte, dass Kiærs Stichprobenverfahren durch zufällige Auswahl der Stichprobe auf eine wissenschaftliche Grundlage gestellt werden könnte. March stellte fest, dass bereits im 18. Jahrhundert Stichproben erhoben wurden und erinnerte die Anwesenden an Laplaces Studien zur Schätzung der französischen Bevölkerung. Seine Idee wurde im Anschluss von Arthur Lyon Bowley (1869–1957) weiterentwickelt, der ebenfalls die Bedeutung einer zufälligen Auswahl der Teilnehmer*innen hervorhob und eine zentrale Rolle dabei spielte, das ISI von Kiærs Methode zu überzeugen.

In den 1890er Jahren war Bowley bereits von Kiærs Arbeiten zur »repräsentativen Methode« fasziniert. Im Folgenden sollte er die Person werden, die die statistische Theorie der Stichprobenziehung mit der Sozialforschung verband; ihm gelang der »decisive methodological breakthrough for the social survey as we know it today« (Marsh 1982, S. 25). In einer Rede vor der British Association for the Advancement of Science im Jahr 1906 (Bowley 1906) stellte Bowley eine theoretische Rechtfertigung der Stichprobenerhebung vor und demonstrierte zudem deren praktische Durchführbarkeit mittels einer 1912 in der britischen Stadt Reading durchgeführten Erhebung (Bowley 1913), welcher ähnliche Studien unter seiner Leitung in Northampton, Warrington und Stanley folgten. Im Jahr 1924 wurde Bowley Mitglied eines ISI-Ausschusses (zusammen mit Corrado Gini, Adolphe Jensen, Lucien March, Verriijn Stuart und Frantz Zizek), welcher mit der Aufgabe betraut wurde, die repräsentative Methode zu evaluieren. Dieser erarbeitete daraufhin einen Bericht mit dem Titel *Measurement of the precision attained in sampling* (Bowley 1925), der 1925 auf dem ISI-Kongress in Rom vorgestellt wurde und in dessen Anschluss der Ausschuss letztlich postulierte, dass die Anwendung der repräsentativen Methode in einer Reihe von Fällen genaue Ergebnisse liefern kann und sie somit unter bestimmten Voraussetzungen auch empfahl (vgl. IIS 1926, S. 62–69). 30 Jahre nach Kiærs ursprünglichem Vorschlag und sechs Jahre nach seinem Tod war damit der Grundstein für die moderne Theorie der Stichprobenerhebung gelegt. In diesem Jahr wurden sowohl die »bewusste Auswahl« Kiærs als auch die »zufällige Auswahl« Bowleys vom ISI als zwei

akzeptierte Ansätze der Stichprobenerhebungen anerkannt. Anders als bei Kiær erfordert Bowleys Methode keine Kenntnisse oder Annahmen darüber, wie sich die Grundgesamtheit von der Stichprobe unterscheidet, worin ihr großer Vorteil gegenüber ersterer besteht, wie sich im nächsten Teil dieses Textes zeigen wird.

Adolphe Jensen, ein Mitglied des ISI-Ausschuss, von welchem die repräsentative Methode evaluiert worden war, fasste diesen Einstellungswandel gut zusammen. Er befand 1926, dass es bei der frühen Debatte des Jahres 1903 hauptsächlich um die »recognition of the method in principle« (Jensen et al. 1926, S. 59) gegangen sei, während es nun kaum noch Statistiker*innen gebe, »who in principle will contest the legitimacy of the representative method« (ebd.). Das Stichprobenverfahren war zwischen 1895 und 1925 vom ISI anerkannt worden und hatte damit an epistemischer Autorität gewonnen. Dies war allerdings nur der Auftakt für neue Konflikte – mit neuen Grenzziehungen – welche bis heute eine Rolle in statistischen Fachdiskursen spielen.

3.4 Der Zufall setzt sich durch

Konfliktpotenzial lieferte im Jahrzehnt, das auf die Anerkennung von Stichprobenerhebungen folgte, die Frage danach, ob Stichproben bewusst oder zufällig ausgewählt werden sollten. Zwischen 1925 und 1934 fanden sowohl die »bewusste« und die »zufällige« Auswahl von Stichproben Anwendung; ab 1934 setzte sich die zufällige Auswahl durch. Eine maßgebliche Rolle dabei spielten die Arbeiten von Jerzy Neyman (1894–1981), welcher in seinem bahnbrechenden Artikel *On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection* (Neyman 1934) nicht nur die Theorie der Konfidenzintervalle entwickelte, sondern auch zeigte, dass die repräsentative Methode Kiærs unter bestimmten Bedingungen zu unvollständigen Ergebnissen führte. Konfidenzintervalle beschreiben einen Bereich, welcher den tatsächlichen Wert mit einer bestimmten Wahrscheinlichkeit enthält. Derlei Aussagen konnten im Falle von Kiærs Methode nicht getroffen werden.

Dass sich randomisierte Verfahren in Folge durchsetzten, war allerdings nicht nur das Ergebnis theoretischer Arbeiten wie denen Neymans, sondern auch konkreter Ereignisse, die die Mängel der bewussten Auswahl bei der Anwendung offenlegten. Großen Einfluss hatten die US-Präsidentenwahlen von 1936 und 1948, deren Ausgang von der Zeitschrift *Literary Digest* bzw. von dem Meinungsforschungsunternehmen Gallup falsch prognostiziert wurde.

Der 1890 gegründete *Literary Digest* führte schon lange Leser*innenbefragungen durch und baute dadurch eine mit der Zeit immer größer werdende Adressdatei auf. Im Jahr 1895 umfasste dieses Verzeichnis bereits mehr als 350 000 Adressen; 1932 waren es 32 Millionen. Zur Vorhersage des Wahlergebnisses von 1928 verschickte der *Literary Digest* 18 Millionen Fragebögen an seine Leser*innen und konnte da-

mit den Sieg Herbert Hoovers erstaunlich gut vorhersagen. Im Jahr 1936 scheiterte der bisher als sehr vertrauenswürdig eingestufte *Literary Digest* jedoch kläglich, als er fälschlicherweise den Sieg des Republikaners Alf Landon über den amtierenden Präsidenten Roosevelt prognostizierte.⁹ Als Grund für die falsche Voraussage des Digests gilt ein Stichprobenfehler: Befragte wurden vor allem aus dem Telefonbuch und einem Register von Autobesitzer*innen rekrutiert, wodurch mehr wohlhabende Personen in die Stichprobe gelangten als weniger wohlhabende. Zudem gab es eine Schweigeverzerrung und eine geringe Ausschöpfungsquote, was bedeutet, dass relativ wenige der Befragten tatsächlich auch antworteten (vgl. etwa Squire 1988); die vom *Literary Digest* gezogene Stichprobe war also nicht repräsentativ für die Bevölkerung. George Gallup (1901 – 1984), der gerade sein eigenes Umfrageunternehmen aufbaute, zog mit seinen eigenen Methoden eine Stichprobe von 3.000 Personen und sagte lange vor deren Veröffentlichung mit einem Fehler von nur einem Prozentpunkt voraus, wie die Digest-Vorhersagen ausfallen würden. Mit einer weiteren Stichprobe von etwa 50.000 Personen sagte Gallup korrekt den Sieg von Roosevelt voraus, auch wenn seine Vorhersage bei Roosevelts Stimmenanteil um einiges daneben lag. Der enorme Vertrauensverlust, den der *Literary Digest* 1936 aufgrund der Fehlprognose erlitt, gilt als Hauptursache für den Untergang des Magazins zwei Jahre später. Wie Keller (2001, S. 34) schreibt, war das Orakel »aber nur verstummt, um der wissenschaftlich rationalen Analyse das Wort zu übergeben«.

Zwölf Jahre nach seiner fulminanten Wahlprognose, zur Präsidentschaftswahl 1948, unterlief Gallup jedoch selbst eine phänomenale Fehlprognose. Im Duell zwischen dem amtierenden Präsidenten Harry Truman und dem Republikaner Thomas Dewey sagte das Unternehmen fälschlich einen Sieg von Dewey voraus; Gallups Wahlprognose genoss zu dieser Zeit so viel Vertrauen, dass die *Chicago Tribune* am Tag nach der Wahl fälschlicherweise »*Dewey Defeats Truman*« titelte.¹⁰ Im Anschluss widmeten sich Gremien wie der Social Science Research Council der Frage, was schiefgelaufen war. Der Grund für die Fehlprognose war, dass es den Interviewer*innen erlaubt war, frei zu entscheiden, wen sie befragten. In jeder der festgelegten Kategorien (u. a. Geschlecht, Alter, ökonomischer Status) waren Republikaner*innen wohl einfacher zu erreichen. Die Untersuchungen führten zu einem Ansehensverlust der Quotenstichprobe (vgl. Likert 1948), welche auch von Gallup selbst

9 Die Leser*innenbefragung des *Literary Digest* von 1936 hatte eine Stichprobengröße von etwa 10 Millionen Wähler*innen und eine Rücklaufquote von etwa 25 Prozent, worauf sich in der Ausgabe vom 31. Oktober 1936 die Voraussage stützte, dass Landon 370 von 531 Wahlmännerstimmen gewinnen würde. Nach der Wahl zeigte sich jedoch ein gänzlich anderes Bild: Landon gewann lediglich 8 Wahlmännerstimmen, Roosevelt hingegen 523.

10 Weniger bekannt als die Schlagzeile der *Chicago Tribune* ist der Fall des Münchner Merkur, welcher am Tag nach der Wahl mit dem Titel »Thomas E. Dewey Amerikas neuer Präsident« aufwartete.

durch Zufallsstichproben als Umfragebasis ersetzt wurde. Dieses Ereignis markierte den Beginn eines über mehrere Jahrzehnte langanhaltenden Konsens über die Überlegenheit probabilistischer Stichprobenverfahren, welche zu einem charakteristischen Element der Umfrageforschung und Branche wurde.

Dieser epistemische Kontext lässt sich in den Worten der deutschen Demoskopin und Allensbach-Gründerin Elisabeth Noelle-Neumann (1916–2010) und des Meinungsforschers Thomas Petersen wie folgt zusammenfassen:

»Der Übergang von schriftlichen Umfragen auf mündliche mit sorgfältig kontrollierter repräsentativer Stichprobe ist ja – soweit es die Verallgemeinerungsfähigkeit der Ergebnisse betrifft – der entscheidende Fortschritt der Demoskopie seit dem Durchbruch bei den amerikanischen Präsidentschaftswahlen von 1936. Damals wurde vor der ganzen Welt demonstriert, daß 10 Millionen verschickte Fragebogen zu einem falschen Ergebnis und wenige tausend Interviews mit repräsentativem Querschnitt zu der richtigen Prognose führten. Das Prinzip heißt: Es darf nicht von der Entscheidung des einzelnen abhängen, ob er in die Stichprobe gelangt oder nicht (›self selection‹), sondern alle Personen, die zum ›Universum‹ gehören, für das die Ergebnisse verallgemeinert werden sollen, müssen die gleiche oder berechenbare Chance haben, in die Stichprobe zu gelangen.« (Noelle-Neumann und Petersen 2005, S. 254)

Osborne und Rose (1999) argumentieren, dass die repräsentative Stichprobe heute den Charakter dessen trägt, was von Latour *Ready Made Science* genannt wurde, also eine Forschungsweise, die auf fertige Tatsachen und Forschungen abzielt. Während, wie in diesem Aufsatz gezeigt, zwischen 1925 und 1934 sowohl den Methoden der bewussten wie auch der zufälligen Auswahl epistemische Autorität zugesprochen wurde, änderte sich dies ab 1934, besonders im Anschluss an die US-Präsidentschaftswahl 1948, wodurch sich ein Primat zufallsbasierter Verfahren durchsetzte und diese zum Goldstandard erhoben wurde. In gewisser Weise wurde eine Kontroverse abgeschlossen, mit dem Ergebnis, dass die Umfrageforschung den Charakter der *Ready Made Science* bekam. Wie sich allerdings zeigen wird, wurde diese Kontroverse nur scheinbar geschlossen und es scheint, als hätte die Umfrageforschung den Modus der *Ready Made Science* – zumindest zum Teil – wieder hinter sich gelassen. Noelle-Neumann und Petersen führen weiter aus, indem sie aufzeigen, dass dieser Goldstandard heute nicht mehr zwingend grundsätzlicher Bestandteil jeglicher Meinungsumfragen ist.

»Heute, siebzig Jahre später, sind Meinungsumfragen in der ganzen Welt populär geworden, und damit hat sich die Neigung, Umfragen zu veranstalten, außerordentlich gesteigert, oft ohne daß die Vorbedingung, daß nämlich eine repräsentative Stichprobe ermittelt werden muß, noch beachtet wird. Das Ergebnis solcher nichtrepräsentativen Umfragen – dazu zählen heute auch die meisten Internet-

Umfragen – ist zwischen Zeitverschwendung und Betrug einzustufen.« (Noelle-Neumann und Petersen 2005, S. 254)

Auch hier lohnt es sich, die rhetorischen Figuren unter die Lupe zu nehmen, mit denen solche nicht-repräsentativen Umfragen beschrieben werden. Wenn Noelle-Neumann und Petersen schreiben, diese seien zwischen Zeitverschwendung und Betrug einzustufen, sprechen sie diesen die epistemische Autorität ab. Es lässt sich also erneut eine Form der Grenzarbeit feststellen, mit der das Territorium epistemischer Autorität abgesteckt wird. Entsprechende Grenzarbeit spielt eine große Rolle in gegenwärtigen Diskursen der Demoskopie; ein weiteres Beispiel stellt die in der Einleitung erwähnte Debatte um die Umfragen zu Mesut Özil und İlkay Gündoğan dar. Ein knappes Jahrhundert nach den Methodenstreits der 1920er und 1930er Jahre wird also erneut darüber diskutiert, ob Stichproben bewusst oder zufällig erhoben werden müssen, um den Anspruch von Wissenschaftlichkeit zu erfüllen.

Betrachtet man die jüngere Geschichte der Stichprobenerhebung, so stellt man fest, dass diese Wiederkehr der bewussten Auswahl in erster Linie auf zunehmende Probleme bei der Durchführung von Zufallsstichproben zurückzuführen ist. Diese Probleme haben in erster Linie mit sinkenden Responsequoten und damit steigenden Kosten zu tun. Zentral für die neuen Entwicklungen ist ein Wandel hin zu online durchgeführten Erhebungen. Die Pionierarbeit für diese Entwicklung nahm in etwa Mitte der 1980er Jahre ihren Ausgang. Von großer Bedeutung waren dabei Wissenschaftler*innen um Willem Saris und das niederländische Telepanel (Saris und De Pijper 1986). Man begann dabei zunächst mit 1.000 offline und zufällig ausgewählten Haushalten, welche mit entsprechender Technik (Personal Computer, Modems und falls erforderlich auch mit Telefonanschluss) ausgestattet wurden. Dadurch wurde ermöglicht, dass jede Woche ein neuer Fragebogen heruntergeladen werden konnte. Nach der Teilnahme wurde ein Zentralcomputer angewählt und die Antworten hochgeladen (Saris 1998). Auf diese Weise war es möglich, eine Zufallsstichprobe aufrechtzuerhalten und für eine neue Erhebungsmethode fruchtbar zu machen. Auf die Frage angesprochen, was sich seit dem Start des Telepanels verändert hat, sagte mir Willem Saris in einem Interview: »Es gibt [...] einen Unterschied zum heutigen Verfahren, nämlich dass wir versucht haben, eine Zufallsstichprobe aus der Bevölkerung zu ziehen. Wir wollten, dass sie repräsentativ ist, und das ist heute nicht mehr der Fall. Im Allgemeinen werden, von wenigen Ausnahmen abgesehen, nur Daten von Personen erhoben, die sich zu einer Teilnahme bereit erklärt haben« (Willem Saris, Interview [Übers. L.G.]). Diese Entwicklung markierte den Bruch mit dem seit den 1950ern etablierten Goldstandard, was erneut, wie in der Einleitung angedeutet, neue Praktiken der Grenzarbeit hervorbringt und möglicherweise einen Übergang von *Ready Made Science* zurück zu *Science in the Making* einläutet. Ähnlich wie vor 1934 laufen die neuen Konfliktlinien der gegenwärtigen Debatte um die Frage nach der Überlegenheit bewusster oder zufälliger Auswahl,

wenn auch vor dem Hintergrund neuer technischer Möglichkeiten. Ob sich die Zufallsstichprobe erneut durchsetzen wird, werden die nächsten Jahre zeigen.

4. Schluss

Ich möchte mit einer Dimension enden, die die aktuelle Kontroverse mitbeeinflusst, nämlich die größer werdende Rolle privater profitorientierter Unternehmen, welche die Autorität der akademischen Sozialwissenschaften in der Bestimmung des Wissens über das Soziale durch neue Formen der Datenerhebung herausfordert und damit neue Praktiken der Grenzziehung hervorbringt. In ihrem 2007 erschienenen Artikel *The Coming Crisis of Empirical Sociology* formulierten Savage und Burrows (2007) die Befürchtung, dass, während die empirische Sozialforschung in den Jahren zwischen 1950 und 1990 eine Reihe ausgeprägter methodischer Werkzeuge entwickeln und für sich beanspruchen konnte, im 21. Jahrhundert sozialwissenschaftliche Daten in gänzlich anderer Art und Weise gesammelt werden, wobei die Rolle der akademischen Sozialwissenschaften immer unklarer wird. Während diese noch vor 50 Jahren die Speerspitze des sozialwissenschaftlichen Forschungsapparats darstellten, nehmen sie heute eine zunehmend marginale Position ein. Savage und Burrows beschreiben, wie sich eine für akademische Sozialwissenschaftler*innen bisher unbekannte, jedoch ausgereifte Welt kommerzieller Sozialwissenschaften offenbarte und deren zentrale methodische Ressource, die Stichprobenerhebung, ihre glorreichen Jahre scheinbar hinter sich gelassen hat. Savage und Burrows halten es demnach für unwahrscheinlich, dass die Stichprobenerhebung in Zukunft eine wichtige Rolle als Forschungsinstrument spielen wird und dass diejenigen, die auf diese Methode setzen, möglicherweise der Marginalisierung und Redundanz ausgesetzt sein werden. Savages und Burrows' Argument lautet also, dass die innovative Kraft, welche die Sozialwissenschaften in der Nachkriegszeit bestimmte, nachlässt und die Innovation in den Forschungsmethoden primär nicht mehr in der Wissenschaft, sondern in profitorientierten Organisationen angesiedelt sein wird. Sieben Jahre später reflektierten dieselben Autoren (Burrows und Savage 2014) über ihre ursprünglichen Aussagen und schließen mit der Vermutung, dass diese Krise die Expert*innen nicht darin vereinen wird, neue Formen der Datenerhebung zu erforschen, sondern zu einer Polarisierung und Spaltung führen wird. Ein Blick auf Grenzarbeiten in der Geschichte und in zeitgenössischen Kontroversen innerhalb der Umfrageforschung zeigt in jedem Fall, dass die Disziplin selten mit einer Stimmsprache und dies auch in Zukunft wahrscheinlich auch nicht immer werden wird.

5. Literatur

- Bethlehem, J. 2009. The rise of survey sampling. (CBS Discussion Paper; No. 09015). *Statistics Netherlands*.
- Bowley, A. L. 1906. Address to the Economic Science and Statistics Section of the British Association for the Advancement of Science, York, 1906. *Journal of the Royal Statistical Society* 69: 540.
- Bowley, A. L. 1925. Measurement of the Precision attained in Sampling. *Bulletin de l'Institut International de Statistique* 22: 359–380.
- Bowley, A. L. 1913. Working-Class Households in Reading. *Journal of the Royal Statistical Society* 76: 672.
- Burrows, Roger, und Mike Savage. 2014. After the crisis? Big Data and the methodological challenges of empirical sociology. *Big Data & Society* 1: 1–6.
- Desrosières, Alain. 2005. *Die Politik der großen Zahlen*. Berlin/Heidelberg: Springer-Verlag.
- Gertenbach, Lars, und Henning Laux. 2019. *Zur Aktualität von Bruno Latour*. Wiesbaden: Springer Fachmedien Wiesbaden.
- Gieryn, Thomas F. 1983. Boundary-Work and the Demarcation of Science from Non-Science: Strains and Interests in Professional Ideologies of Scientists. *American Sociological Review* 48: 781.
- Gieryn, Thomas F. 1999. *Cultural Boundaries of Science*. The University of Chicago Press.
- IIS. 1895. *Bulletin de l'Institut international de statistique*. TOME IX. Den Haag: Institut international de statistique.
- IIS. 1926. *Bulletin de l'Institut international de statistique*. TOME XXII. Den Haag: Institut international de statistique.
- Jasanoff, Sheila. 1995. *Science at the Bar: Law, Science, and Technology in America*. Cambridge, MA: Harvard University Press.
- Jensen, Adolph, Konrad Saenger, und Arthur L Bowley. 1926. Rapport de M. Jensen sur l'Application de la méthode représentative. *Bulletin de l'Institut international de statistique* 22: 58–61.
- Keller, Felix. 2001. *Archäologie der Meinungsforschung*. Konstanz: UVK Verlagsgesellschaft mbH.
- Kruskal, William, und Frederick Mosteller. 1980. Representative Sampling, IV: The History of the Concept in Statistics, 1895–1939. *International Statistical Review/Revue Internationale de Statistique* 48: 169.
- Kuhn, Thomas S. 2012. *The structure of scientific revolutions*. University of Chicago press.
- Laplace, Pierre Simon. 1786. Sur les naissances, les mariages et les morts a Paris depuis 1771 jusqu'a 1784 et dans toute l'étendue de la France, pendant les an-

- nées 1781 et 1782. *Mémoires de l'Académie Royale des Sciences présentés par divers savans* 35–46.
- Latour, Bruno. 1987. *Science in Action – How to follow scientists and engineers through society*. Cambridge, Massachusetts: Harvard University Press.
- Likert, Rensis. 1948. Public Opinion Polls. *Scientific American* 179: 7–11.
- Lusinchi, Dominic. 2021. Kiær and the rebirth of the representative method: A case-study in controversy management at the International Statistical Institute (1895–1903). *Journal of the History of the Behavioral Sciences* 1–20.
- Marsh, Catherine. 1982. The survey method: the contribution of surveys to sociological explanation.
- Mouat, Frederic J. 1885. History of the Statistical Society of London. *Journal of the Statistical Society of London* 14–371.
- Neyman, Jerzy. 1934. On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection. *Journal of the Royal Statistical Society* 97: 558–625.
- Noelle-Neumann, Elisabeth, und Thomas Petersen. 2005. *Alle, nicht jeder*. Berlin/Heidelberg: Springer-Verlag.
- Osborne, Thomas, und Nikolas Rose. 1999. Do the social sciences create phenomena?: the example of public opinion research. *The British Journal of Sociology* 50: 367–396.
- Prietl, Bianca, und Armin Ziegler. 2016. Machtvolle Grenzen als konstitutive Momente des Sozialen. In *Handbuch Geschichte der deutschsprachigen Soziologie*, 1–16. Wiesbaden: Springer Fachmedien Wiesbaden.
- Quetelet, Adolphe. 1846. *Lettres à SAR le duc régnant de Saxe-Coburg et Gotha: sur la théorie des probabilités, appliquée aux sciences morales et politiques*. M. Hayez.
- Saris, W E, und W M De Pijper. 1986. Computer assisted interviewing using home computers. *European Research* 14: 144–150.
- Saris, Willem E. 1998. Ten Years of Interviewing Without Interviewers: The Telepanel. In *Computer Assisted Survey Information Collection*, Hg. Couper Mick P. et al. John Wiley & Sons.
- Savage, Mike, und Roger Burrows. 2007. The Coming Crisis of Empirical Sociology. *Sociology* 41: 885–899.
- Squire, Peverill. 1988. Why the 1936 Literary Digest Poll Failed. *Public Opinion Quarterly* 52: 125.
- Stigler, Stephen M. 1986. *The history of statistics: the measurement of uncertainty before 1900*. Cambridge, Mass.: Belknap Press of Harvard University Press.
- Vehovar, Vasja, und Katja Lozar Manfreda. 2017. Overview: Online Surveys. In *The SAGE Handbook of Online Research Methods*, Hg. Nigel G. Fielding, Raymond M. Lee und Grant Blank, 143–161. London: SAGE Publications Ltd.

