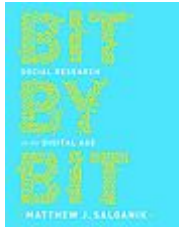


Das Beste aus zwei Welten?

von Janosch Schobin



Titel Bit by Bit

Untertitel Social Research in the Digital Age

Autor(en) Matthew J. Salganik

Land USA / Großbritannien

Erschienen Princeton, NJ / Oxford 2017: Princeton University Press

Umfang 448 S.

Preis \$ 35,00 | £ 27,00

ISBN 9780691158648

Matthew Salganik entwirft eine quantitative Soziologie für das digitale Zeitalter

Im Rahmen der Digitalisierungsdebatte stellt sich für die Soziologie virulent die Frage, ob es für eine zunehmend in digitalen Strukturen organisierte Gesellschaft nicht auch ein Repertoire eigener quantitativer Methoden – oder sogar eine ganze neue quantitative Methodologie – braucht. Bei aller in dieser Debatte herrschenden Uneinigkeit dürfte zumindest Folgendes Konsens sein: Eine Soziologie, die nicht über das nötige Handwerkszeug verfügt, um soziales Handeln im Rahmen digitaler

Strukturen angemessen quantitativ beobachten und beschreiben zu können, ist sowohl gesellschaftsdiagnostisch als auch gesellschaftstheoretisch zur Bedeutungslosigkeit verurteilt. Offen dagegen ist die Frage, wie eine *neue* quantitative Methodik oder gar *neue* Methodologie einer digital kompetenten Soziologie aussehen könnte.

Das Lehrbuch „Bit by Bit“ von Matthew Salganik verspricht vor diesem Hintergrund das Beste aus zwei Welten zusammenzuführen, ist es doch erklärtes Ziel des Autors, die ‚alten‘ Methoden der quantitativen Feldforschung mit den ‚neuen‘ Methoden der Data Science in ein produktives Ergänzungsverhältnis zu bringen (S. 5). „Bit by Bit“ ist dabei schon deshalb ein interessanter Versuch, weil es sich nicht mit abstrakten Problemen, sondern mit konkreten Studien aus der digitalen Gegenwart befasst. Anhand dieser exemplarischen Studien werden im Buch die ‚Best Practices‘ digitaltauglicher Methoden herausgearbeitet. Umrahmt wird das Buch von der methodologisch zu verstehenden These, dass mit dem digitalen Wandel nicht nur neue gesellschaftlich relevante Strukturen entstünden, die in ihrem eigenen Medium beobachtet werden wollen, sondern sich auch die Bedingungen der Erforschung aller sozialen Tatsachen – also analoger wie digitaler – extrem veränderten. Der zentrale Ausgangspunkt, zu dem Salganik dabei in der Behandlung seiner Beispiele immer wieder zurückkehrt, ist die Krise des repräsentativen Paradigmas in der quantitativen empirischen Sozialforschung: Seit Jahren steigt selbst in den besten repräsentativen Surveys die Rate der Befragungsverweigerung.[1] Salganik bemerkt richtigerweise, dass sich dieses Problem im Rahmen von auf digitalen Massendaten basierenden Umfragen weiter verschärft (S. 29 ff.). Denn während sich die Repräsentativität einer Befragung zumindest in der Theorie dadurch gewährleisten lässt, dass jedes Objekt aus der interessierenden Grundgesamtheit eine bekannte positive Chance hat, in die Stichprobe zu gelangen, ist eine solche Grundvoraussetzung bei Daten, die auf digitalen Plattformen automatisiert generiert werden, in der Regel nicht erfüllt. Tatsächlich sind selbst die besten sozialwissenschaftlichen Umfragen, die auf Zufallsstichproben beruhen, in der Praxis bereits weit von diesem Ideal entfernt. Was ihre Repräsentativität angeht, nähern sich analoge und digitale Daten also zunehmend aneinander an – und zwar indem sie sie vernachlässigen. Die Ursache für die Krise des repräsentativen Paradigmas, so kann Salganik gedeutet werden, besteht in einer Transformation der Aufmerksamkeitsökonomie der Befragten – und hiervon ist der Erfolg quantitativer empirischer Sozialforschung direkt abhängig. Die Subjekte sind immer weniger bereit, als passive Beobachtungsgegenstände adressiert zu werden: „If you want to collect data in the digital age, you need to realize that you are competing for people’s time and attention. The time and attention of your participants is incredibly valuable to you; it is the raw material of your research.“ (S. 356)

Im hier besprochenen Buch entwickelt Salganik zwei Strategien, mit denen er der Krise des repräsentativen Paradigmas begegnen will. Die erste dieser beiden Strategien besteht in der Anwendung und Weiterentwicklung neuer Methoden der „participant-centered data collection“ (S. 356). So soll etwa der Einsatz von

Spielelementen – auch Gamification genannt – dem Problem abhelfen, dass das Beantworten umfangreicher Fragebögen die Befragten langweilt oder gar nervt (S. 115). Die größte Chance des Einsatzes digitaler Spielelemente sieht Salganik darin, die Studienteilnehmer*innen primär nicht länger als Träger von Eigenschaften zu adressieren, aus denen Daten zu extrahieren sind, sondern sie vielmehr selbst als Kollaborateure in die Datenproduktion einzubinden. Dieser Idee widmet er unter dem Titel „Creating Mass Collaboration“ ein eigenes Kapitel (S. 231–280). An Beispielen wie *Foldit* – einem *Serious Game*, bei dem Spieler*innen Proteinketten falten und damit Forscher*innen bei der Lösung extrem hochdimensionaler Optimierungsprobleme unterstützen – wirkt diese Idee auf den ersten Blick plausibel (S. 249 ff). Am Ende scheitert die Darstellung jedoch daran, überzeugend auf die klassischen Probleme der soziologischen Feldforschung übertragen zu werden. Wie etwa soll das vorgeschlagene Verfahren bei einer Befragung funktionieren, bei der möglichst genau das Einkommen einer Person erhoben werden soll? Es ist zumindest fragwürdig, ob sich Befragungen, die dem Einzelnen notwendigerweise den Aufwand abverlangen, exakt über sich und die eigenen Lebensumstände Auskunft zu geben, stets in eine unterhaltsame „Spielform“ bringen lassen. Eine weitere Möglichkeit, dem Wandel der Aufmerksamkeitsökonomie auf methodischer Ebene mittels digitaler Datenerhebungstechniken zu begegnen, spürt Salganik in der Entwicklung von speziellen ‚mikroinvasiven‘ Techniken wie etwa den „ecological momentary assessments“ (S. 108 f.) auf. Sie gestatten es, einen klassischen, meist langen und potenziell ermüdenden Fragebogen aufzusplitten und ihn nur nach und nach in wohl dosierter Form an die Befragten zu bringen. Eine quantitative Umfrage kann durch ein solches Vorgehen besser in den Alltag integriert und an die Bedürfnisse der Befragten angepasst werden, was wiederum deren Bereitschaft zur Beantwortung erhöht. Eine dritte exemplarische Möglichkeit zur stärkeren Einbindung von Befragten in den Prozess der Datenerhebung sieht Salganik darin, die klassische Dichotomie zwischen offenen und geschlossenen Fragen durch neuartige digitale Frageerzeugungsmethoden wie beispielsweise Wiki Surveys aufzubrechen. Geschlossene Fragen sind bekanntlich Standard in quantitativen Umfragen – jedoch nicht, weil sie stets die sinnvollste Art der Erhebung sind. Vielmehr ist das Gegenteil der Fall: Sie stoßen die Befragten oft vor den Kopf, weil sie ihre eigenen (emischen) Konzepte in den fremden (etischen) von Forscher*innen vorgegebenen Kategorien nicht wiederfinden. Geschlossene Fragen werden vor allem deshalb präferiert, weil sie sich am einfachsten auswerten lassen. Offene Fragen hingegen ermöglichen es den Befragten zwar, sich weitaus genauer und idiosynkratischer zu äußern, sind in der Analyse aber sehr aufwendig und oft nur unter großen Schwierigkeiten in geschlossene Kategorienschemata zu überführen, die statistischen Analyseverfahren zugänglich sind. Wiki-Surveys setzen an diesem Problem an. Sie sind eine Art Hybrid aus offenen und geschlossenen Fragen und gestatten es den Befragten, nach und nach neue Kategorien für ein eingangs vorgegebenes geschlossenes Kategorienschema vorzuschlagen. Wird eine neue Antwortmöglichkeit (das Verfahren der Annahme/Ablehnung kann variieren) schließlich von Seiten der Forscher*innen angenommen, steht sie fortan auch allen anderen Befragten als Antwortkategorie im geschlossenen Frageschema zur Verfügung. Im Rahmen eines solchen Verfahrens entsteht durch die Partizipation

der Befragten im Laufe der Zeit eine empirisch gesättigte geschlossene Frage (S. 111-115).

Die zweite von Salganik vorgeschlagene Strategie im Umgang mit der Repräsentationskrise bei Befragungen könnte als ‚Flucht in die Kausalität‘ beschrieben werden: Denn dem Autor zufolge besteht einer der großen Vorteile bei der Verwendung digitaler Daten nicht nur darin, neuartige Feldexperimente durchführen zu können. Sie ermöglichen auch die Beobachtung natürlicher Experimente, die sich auf digitalen Plattformen ereignen. Das mag auf den ersten Blick verzweifelt wirken. Soziolog*innen betrachten die in den Wirtschafts- und Humanwissenschaften dominierenden Kausaltheorien –, die im Übrigen auch Salganik referiert, – in der Regel mit großer Skepsis. Das Ziel solcher Theorien[2] ist die Klärung der Frage, unter welchen theoretischen Bedingungen sich in empirischen Feldstudien kausale Faktoren isolieren und in ihrer Wirkgröße genau bestimmen lassen. In sozialen Systemen jedoch sind schon aus theoretischer Perspektive die anzunehmenden Kausalzusammenhänge meist so strukturiert, dass sich einzelne Kausalfaktoren selbst nach Maßgabe der entsprechenden Kausaltheorien nicht genau identifizieren lassen.[3] Noch schwerer wiegt, dass sich in den Daten empirischer Feldstudien selten ‚echte‘ bedingte statistische Unabhängigkeiten finden, die als Kriterien der Falsifikation kausaler Theorien herhalten könnten. Andrew Gelman bezeichnete das unlängst als „No-true-zeros“-Problematik.[4] Vor diesem Hintergrund hat die quantitative Soziologie im Gegensatz zu den Wirtschafts- und Humanwissenschaften von jeher ihre Aufgabe viel stärker in der Deskription von Populationen auf Basis beschränkter Beobachtungen gesehen. Auf den zweiten Blick ist der Experimentalismus bei Salganik jedoch etwas mehr als der bloße Ausdruck eines unreflektierten digitalen Neopositivismus. Es handelt sich um den Entwurf einer neuen Rolle für Feldforscher*innen, der im Kontext der These von einer Veränderung der Aufmerksamkeitsökonomie in sich digitalisierenden Gesellschaften gelesen werden muss: Die Subjekte wollen und können immer weniger als passiver Gegenstand einer Deskription adressiert werden. Um dem zu begegnen, nehmen Soziolog*innen in digitalen Studien und Feldexperimenten vermehrt die Rolle von Spielleiter*innen ein: Sie bauen oder modifizieren digitale Welten und ermöglichen den Studienteilnehmer*innen dadurch Handlungen und Kommunikationen, die dann wiederum zu Daten eines Feldexperiments oder Surveys und so zu empirischen Belegen für oder gegen eine soziologische Theorie werden.

Salganik reduziert seine Idee am Ende auf fünf methodische Prinzipien: „[...] motivate participants, leverage heterogeneity, focus attention, enable surprise, and be ethical“ (S. 265). Zu Ende gedacht ist diese Position sicher nicht. Das liegt vermutlich auch daran, dass es sich um ein Lehrbuch handelt, das sich an der Praxis orientiert und an eine Wissenschaftsgemeinde gerichtet ist, für die sowohl feldexperimentelle Forschung als auch der Versuch der Überprüfung kausaler Theorien üblich sind und daher keiner besonderen Rechtfertigung bedürfen. Das Buch soll also Wege aufzeigen, die mehr Praxis dieser Art möglich machen. Das Totschlagargument, nach dem es sich hier schlicht um einen neuen

Technopositivismus handelt, wird der Lehrmeinung des Buches jedoch nicht gerecht. Das wird besonders bei der Behandlung der unerwarteten Effekte digitaler Infrastrukturen auf die Datenproduktion deutlich: Salganik stellt schon ganz am Anfang seines Buches heraus, dass digitale Daten systematisch durch algorithmische Agenten koproduziert werden, die von ‚interessierten‘ Akteur*innen programmiert werden. Die Daten „driften“ (S. 33) also systematisch. Damit ist gemeint, dass sich die algorithmischen Erzeugungsbedingungen des Datenkorpus unter der Hand kontinuierlich ändern. Folglich sind Muster in den Daten nicht selten auf eine komplizierte Verzahnung zwischen Nutzer*innenfeedback und den Veränderungen der algorithmischen Datenproduktionsbedingungen digitaler Plattformen zurückzuführen (S. 33–38). Salganik behandelt diesen Umstand zugegeben durch das gesamte Buch hinweg stets ausschließlich als methodisches und nicht auch als methodologisches Problem. Der Autor versäumt zudem, derlei Aspekte auf die von ihm vertretene ‚Best Practice‘ zur Erstellung digitaler Welten, die der Datenerhebung dienen, zu beziehen. Aber von einem Datenpositivismus bleibt auch so schon nichts übrig, schreibt er doch: „In my experience, there is something like a ‘no free lunch’ rule for data: if you don’t put in a lot of work collecting it, then you are probably going to have to put in a lot of work think (sic!) about it and analyzing it.“ (S. 61)

Hinsichtlich der Bewertung des Buches sehe ich mich gezwungen, zweiteilig vorzugehen, da es sich zum einen um ein Lehrbuch, zum anderen aber auch um einen Praxisvorschlag für die Zukunft handelt. In ersterer Hinsicht ist das Buch tadellos: Es ist packend geschrieben und selbst bei der Erläuterung vergleichsweise komplizierter sozialwissenschaftlicher Methoden glänzt es durch ausgezeichnete Nachvollziehbarkeit. Besonders hervorzuheben ist der lesefreundliche Umgang mit mathematischem Ballast; dieser wird möglichst schlank und mit wenig Formalismus im Anhang präsentiert. Der einzige Wehrmutstropfen sind die gestellten Übungsaufgaben, die nahezu ausnahmslos sehr ambitioniert sind. Sie setzen gute multivariate Statistikkenntnisse und viel Erfahrung mit flexibel programmierbaren Datenanalyseinstrumenten wie etwa R oder Python voraus. Im deutschen Lehrbetrieb wird das Buch daher vermutlich nur an sehr wenigen Fakultäten und dann auch erst im Masterstudium zum Einsatz kommen. Für das Selbststudium von Doktorand*innen ist es jedoch sehr zu empfehlen – man muss allerdings eine Menge Zeit mitbringen.

Unter dem Gesichtspunkt einer Methodik für Morgen weist das Buch einige Schwachstellen auf, die bereits in der inhaltlichen Darstellung angeklungen sind: Zum einen gelingt die Übertragung neuer Teilnehmer*innen-zentrierter digitaler Erhebungsmethoden in eine exemplarische Praxis nicht immer überzeugend. Zum anderen ist die methodologische Position des Buches nicht ausgereift. Sie bietet dennoch kontinuierlich positive Denkanstöße, weil sie das methodologische Problem des in einer Krise befindlichen repräsentativen Paradigmas immer wieder mit der Praxis einer bereits existierenden digitalen Methodik kontrastiert und zu lösen sucht. Vor diesem Hintergrund kann das Buch auch Methodenkenner*innen empfohlen werden, die nach interessanten Anregungen für die Entwicklung einer

digitalen Methodologie suchen.

Fußnoten

[1] Bruce D. Meyer / Wallace K. C. Mok / James X. Sullivan, Household Surveys in Crisis, in: *The Journal of Economic Perspectives* 29 (2015), 4, S. 199-226.

[2] Zurzeit dürfte es sich dabei vor allem um die Theorie der kausalen Graphen von Judea Pearl sowie die Theorie möglicher Ergebnisse von Donald Rubin handeln (Judea Pearl, *Causality*, Cambridge 2009; Donald B. Rubin, Causal Inference Using Potential Outcomes, in: *Journal of the American Statistical Association* 100 (2005), 469, S. 322-331).

[3] Vgl. hierzu etwa die Bedingungen der Identifizierbarkeit kausaler Beziehungen in einfachen kausalen Systemen mit Feedback-Schleifen in: Judea Pearl, Rina Dechter, *Identifying Independencies in Causal Graphs with Feedback*, 1996.

[4] Andrew Gelman, Causality and Statistical Learning, in: *American Journal of Sociology* 117 (2011), 3, S. 955-966.