

Dirk Baecker

Metadaten. Eine Annäherung an Big Data*

Neue Dimensionen der Konnektivität

Der Umgang nicht nur mit der schieren Größe der Datenmengen, sondern auch mit den überraschenden und weitreichenden Verknüpfungen dieser Daten durch meist weder sichtbare noch nachvollziehbare Algorithmen stellt die heuristischen Fähigkeiten der Menschen auf eine neue Probe. Den Umgang mit einem sogenannten »information overload«, den Alvin Toffler mit seinem Buch über den Zukunftsschock populär gemacht hat (Toffler 1970), sind wir zwar gewohnt, seit wir es mit Menschen zu tun haben, die zu schnell zu viel reden, mit Texten, die zu lang und zu dicht geschrieben sind, mit Bibliotheken, in denen zu viele Bücher stehen, mit Organisationen, in denen auf zu viele Sachverhalte zu viele Rücksichten zu nehmen sind, mit Massenmedien, die zu viele Nachrichten in jeweils unzureichender Tiefe zu schnell hintereinander veröffentlichen, mit dem Internet, in dem auf jeder Seite zu viele Clicks zu schnell auf zu viele Abwege führen, und mit Geheimdiensten, die sammeln, was sie kaum noch auswerten können. Heute geht es allerdings nicht mehr nur um die Menge und Schnelligkeit der Publikationen, sondern um eine neue Dimension der Konnektivität.

Zwar kann man sagen, dass auch ein Geheimnis, ein Gerücht, eine Geschichte, eine Tageszeitung, eine Akte, eine Theorie oder ein Modell, ganz zu schweigen von einem Gedicht, einer Theateraufführung oder einem Tanz bereits eine Konnektivität von Sachverhalten, Einstellungen, Erinnerungen, Erwartungen, körperlichen Befindlichkeiten, emotionalen Anspielungen und ideologischen Schlagseiten enthalten, die kein Mensch, der mit ihnen dennoch meist unbekümmert Umgang hat, durchschauen kann. Doch geht der Verknüpfungsreichtum, mit dem wir es heute

* In: Heinrich Geiselberger und Tobias Moorstedt (Hrsg.), Big Data: Das neue Versprechen der Allwissenheit, Berlin: Suhrkamp, 2013, S. 156-186.

zu tun bekommen, darüber hinaus, weil wir jetzt nicht mehr nur mit Gehirnen, Institutionen, Leidenschaften, Interessen, Wut und Begeisterung rechnen, die uns mit überraschenden und überfordernden Verknüpfungen konfrontieren, sondern mit einem Rechner, der im Netzwerk weiterer Rechner darauf programmiert ist, Algorithmen zu befolgen, die aus den Spuren, die wir im Netz hinterlassen, Schlussfolgerungen ziehen, deren Konsequenzen uns an anderer Stelle ereilen, ohne dass wir dies voraussehen oder etwas dagegen tun könnten.

Diese Informationsflut ist nichts Neues. Toffler selbst hat sich von der Organisationstheorie anregen lassen, die gerade entdeckt hatte, dass die Informationsverarbeitungskapazitäten auch von Menschen, die weitreichende Entscheidungen treffen müssen, begrenzt sind und man es daher immer mit Situationen zu tun hat, in denen Entscheidungen getroffen werden müssen, obwohl man gleichzeitig damit zu rechnen hat, dass wichtige Informationen übersehen wurden (vgl. Simon 1957). Schon im Fall von Organisationen hat man herausgefunden, dass Menschen in solchen Situationen dazu neigen, die Problemstellungen in kleinere Formate zu zerlegen und in so kleinen Schritten zu bearbeiten, dass allfällige Fehler entdeckt und korrigiert werden können. Inkrementalismus nennt man das, wenn man sich anspruchsvoll ausdrücken will, Durchwurschteln, wenn man glaubt, es mit einem defizitären Entscheidungsmodus zu tun zu haben (was nicht der Fall ist). Seither rechnen Organisationsmodelle nicht mehr mit der vollkommenen Rationalität der Ökonomie, sondern mit der beschränkten Rationalität der Psychologie, auch wenn Rationalität hier nichts anderes heißt als klug mit Beschränkungen umzugehen.

Längst haben wir begonnen, unsere Probleme mit der Einführung der Computer und ihrer Netzwerke historisch zu relativieren und, wie oben angedeutet, zu entdecken, dass bereits die Einführungen der Sprache, der Schrift und des Buchdrucks ähnlich überfordernde Qualitäten aufwiesen. Die Sprache führt die Referenz auf Abwesendes, also im Moment nicht Überprüfbares ein; die Schrift führt Vergan-

genheiten ein, an die man sich erinnern, und eine Zukunft, auf die man sich jetzt schon festlegen kann; der Buchdruck zwingt uns dazu, ständig mit Leuten zu rechnen, die gelesen haben und daraus Schlussfolgerungen ziehen, die in keiner Wirklichkeit überprüft worden sind. Wir mussten lernen, mit diesen »Medienkatastrophen«, die zu Recht so heißen, weil sie das System der Gesellschaft im Sinne der Katastrophentheorie René Thoms (Thom 1980) zu einem disruptiven Wechsel in den Parametern seiner Reproduktion zwingen, fertig zu werden, und wir haben dies durchaus erfolgreich gelernt. Die Gesellschaft hat sich über diese Katastrophen hinweg gerettet, indem sie Stämme, soziale Schichten und Funktionssysteme eingeführt hat (vgl. Luhmann 1997; Baecker 2007). Anders als vor ihnen die Herden sprachloser Affen bilden Stämme Kulturen aus, um mit der mündlichen Sprache umgehen zu können. Soziale Schichtung variiert und kontrolliert die Reichweite möglicher Erinnerungen und Pläne. Und Funktionssysteme differenzieren zur Kontrolle von Kommunikation Medien wie Geld, Macht, Wahrheit, Glauben oder Kunst aus. Wir haben den Umgang mit den Katastrophen so erfolgreich gelernt, dass wir uns jetzt nicht mehr daran erinnern, wie uns das gelungen ist. Also müssen wir uns auf umfangreiche kulturwissenschaftliche Forschung und komplizierte soziologische Theorien einlassen, um dies mühsam wieder zu rekonstruieren. Dabei wissen wir noch nicht einmal, ob diese alten Medienkatastrophen und deren Bewältigung uns dabei helfen, auch auf den Computer und seine Netzwerke Antworten zu finden.

Was wir jedoch wissen, ist, dass wir die Beobachtung einer eher ökonomischen Theorie, die jedem neuen Verbreitungsmedium der Kommunikation eine Senkung der Transaktionskosten der Kommunikation, also Effizienzgewinne, zuschreibt, durch eine soziologische Beobachtung ergänzen müssen. Diese soziologische Beobachtung geht davon aus, dass neue Verbreitungsmedien die alten Strukturen überfordern, daher erst einmal abgelehnt werden (so wie Platon die Schrift ablehnt).

te, die Zeitgenossen von Gutenberg den Buchdruck und wir in den vergangenen Jahrzehnten auf mannigfaltige Weisen den Computer) und nur im Medium dieser Ablehnung vorsichtig und dann mit zunehmender Geschwindigkeit eingeführt werden können. Da Techniken aller Art nicht auf die Stirn geschrieben ist, wozu sie nützlich sind, muss dies in allen bekannten Fällen erst ausprobiert werden. Deshalb unterscheiden sich auch die Gründe, aus denen Techniken eingesetzt werden, immer von den Gründen, aus denen sie erfunden worden sind. Ich verzichte auf die bekannten Beispiele (das Auto, das Ford für Bauern erfunden hat, damit sie bequemer ihre Felder erreichen konnten, oder das Telefon, das von Bell für Hörgeschädigte entwickelt wurde).

Interessant ist jedenfalls, dass erst die Ablehnung zeigt, welche Gewohnheiten und Strukturen durch ein neues Medium herausgefordert werden. Und dass uns daher das Studium früherer Medienkatastrophen (das dank der Forschung von Marshall McLuhan, Walter J. Ong, Talcott Parsons, Niklas Luhmann und anderen möglich geworden ist) zum einen die gesellschaftlichen Strukturen vor Augen führt, die wir als vertraute Lösungen mittlerweile unvertrauter Probleme zuerst wieder begreifen müssen; und dass es zum anderen die Komplexität und Unvorhersehbarkeit eines gesellschaftlichen Strukturwandels deutlich macht, auf den wir uns heute angesichts einer neuen Medienkatastrophe einstellen müssen (vgl. Luhmann 1997; McLuhan 1964; Ong 1977; Parsons 1980). Mündliche Kommunikation lehnen wir ab, indem wir sie als nicht glaubwürdig betrachten. Schriftliche Kommunikation lehnen wir ab, indem wir darauf hinweisen, dass die aktuelle Situation andere Antworten erfordert. Und gedruckte Kommunikation lehnen wir ab, indem wir sie kritisch auf Kontexte rückbeziehen, die nicht die unseren sind. All dies geschieht zivil und multiperspektivisch, denn was jetzt nicht glaubwürdig ist, kann es gleich werden, was in der aktuellen Situation nicht passt, passt vielleicht wenig später, und ein Kontext, der jetzt nicht der unsere ist, behält dennoch für andere seine Gültigkeit. Aber all

diese längst institutionalisierten, also selbstverständlich gewordenen Formen müssen wir uns wieder anschauen, wenn wir wissen wollen, wie wir denn nun mit der neuen Medienkatastrophe des Computers und seiner Netzwerke umgehen sollen.

Was sind Metadaten?

Es sieht vielleicht nicht so aus, aber ich bin längst beim Thema dieses Aufsatzes, bei der Frage nach den Metadaten, die es uns erlauben, mit der Herausforderung Big Data nicht nur umzugehen, sondern sie auch produktiv und kreativ zu bewältigen. Das Konzept des Metadatum ist offenbar erstmals von Philip Bagley in einem 1968 erschienenen Buch über Konzepte zur Erweiterung von Programmiersprachen eingeführt worden (Bagley 1968). Diese Metadaten haben nichts mit denen zu tun, die man gegenwärtig in der Diskussion über die Datensauger der Geheimdienste, etwa PRISM, Tempora und so weiter, als solche bezeichnet, weil sie Adressen der Kommunikation in Netzwerken elektronischer Medien sammeln, ohne auch die Inhalte dieser Kommunikation zu erfassen. Diese »Metadaten« sind bloße Verknüpfungsdaten, die nach Bagleys Konzept der Metadaten ihrerseits mit geeigneten Filtern allererst sortiert und codiert werden müssen, um aussagekräftig zu werden.ⁱ

In der Informatik geht es bei Metadaten darum, Konzepte zu finden, die Daten sowohl zu unterscheiden als auch zu vergleichen, sowohl zu listen als auch zu gruppieren erlauben. Je nach Verwendungszusammenhang und Verwertungsinteresse können dabei der Typ des Datums (Text, Zahl, Bild, Programm), die möglichen Werte des Datums (Alphabet, Numerik, Kontraste, Funktionen), der Definitionsbereich des Datums (Gegenstandsbereich, Domain), die Zugänglichkeit des Datums (Code) und nicht zuletzt die möglichen Aktionen des Datums (Operationen) bestimmt und unterschieden werden (vgl. Cormen 2013; McCormick 2012).

Schon hier beginnt eine Wissenschaft, in der sich nichts von selbst versteht (schon gar nicht die gerade gewählte Unterscheidung und Beschreibung der Eigenschaften und Strukturen von Daten) und in der Lösungen für Detailfragen gefunden werden müssen, auf die man ohne die Problemstellung einer Algorithmisierung nicht aufmerksam geworden wäre (vgl. Warnke 2003). Nicht die kleinste Pointe hierbei wäre, dass man in sozialen Zusammenhängen und daher auch bei der sozialen Verwendung von Daten (und was wäre eine »nicht-soziale« Verwendung?) zwischen Daten und Metadaten immer nur fallweise, vorübergehend und beobachterabhängig unterscheiden kann. Schon im nächsten Moment, in einem anderen Fall oder für einen anderen Beobachter werden Metadaten zu Daten und können Daten als Metadaten befragt werden. Logiker, die geneigt sind, sich an Typentheorien, also stabile Objekt- und Metaebenenunterscheidungen à la Bertrand Russells und Alfred North Whiteheads »Principia Mathematica« zu halten, verwirrt dies. Netzwerktheoretiker, die es mit skalierbaren und fraktalen, also in ihrer Leistung steigerbaren und selbstähnlichen Phänomenen zu tun haben, fangen jedoch erst unter dieser Bedingung an, mit einem solchen Konzept zu arbeiten.

In unserem Zusammenhang ist nun wichtig, dass wir solche Metadaten im Umgang mit Big Data schon im ersten Abschnitt dieses Textes eingeführt haben. Wir haben einen Kontext eröffnet, der es erlaubt, die Sammlung, Verknüpfung und Auswertung sehr großer Datenmengen als ein Datum eigener Art aufzufassen und nach Metadaten zu fragen, die in der Lage sind, mit diesem Datum umzugehen. Die Sammlung, Verknüpfung und Auswertung dieser sehr großen Datenmengen ist neu, doch das muss nicht bedeuten, dass die Gesellschaft nicht ihrerseits bereits auf den Umgang mit diesem neuen Problem und dieser neuen Möglichkeit vorbereitet ist. Wir fragen, mit anderen Worten, danach, welche Kontexte, Rahmungen, Skripte, Institutionen, Verfahren, Begriffe und Ideen geeignet sein könnten, mit dem Faktum Big Data umzugehen, während dieses Faktum ohne Zweifel dieselben

Kontexte, Rahmungen, Skripte und so weiter auf eine nicht triviale Weise herausfordert.

Wir begreifen Big Data als Datum einer Gesellschaft, als Datum von Kommunikation, als Datum einer Theorie alter und neuer Verbreitungsmedien, als Datum eines Strukturwandels der Gesellschaft, den wir als Medienkatastrophe beschrieben haben, oder auch als Datum einer Bemühung um das Verständnis der Reichweite von Algorithmen. Wir verstehen die Kultur- und Medienwissenschaften, die Soziologie und andere Formen der Beobachtung von Gesellschaft als Sprachen des Entwurfs von Metadaten zur Kontrolle und zum Vergleich von Daten, die als diese Sprachen genau dann brauchbar sind, wenn es um eine Einschätzung der gesellschaftlichen und kulturellen Bedeutung eines Phänomens geht. Und wir kommen dabei nicht umhin, bereits scheinbar einfache Konzepte wie »Gesellschaft« oder »Kultur« als Metadaten zu verstehen, von denen nicht auf der Hand liegt, welche Listung, Sortierung, Gruppierung und Strukturierung welcher Daten sie jeweils leisten, ganz zu schweigen davon, dass man mit der Frage nach möglichen Unterschieden zwischen kulturwissenschaftlichen, medienwissenschaftlichen und soziologischen (und im Unterschied dazu: sozialwissenschaftlichen) Metadaten konfrontiert wird, die man nur dann geneigt sein kann zu beantworten, wenn kein Kollege, der es anders sehen könnte, in Reichweite ist.

Man versteht auf Anhieb, welcher Anregungsreichtum sich für die wissenschaftliche Heuristik ergibt, wenn man mit der Problemstellung von Big Data die Frage nach Metadaten verknüpft und derart ausgerechnet jene Ressourcen einer wissenschaftlichen Untersuchung und Erklärung variabilisiert, auf die man sich gerade noch verlassen wollte, um endlich herauszufinden, womit man es zu tun hat. Und man versteht, dass allfällige Vokabeln wie die der Problemorientierung, Praxisorientierung, Internationalisierung und Interdisziplinarisierung (von der Suche nach »sozialen Innovationen« gar nicht zu reden, auf die Universitäten neuerdings ver-

pflichtet werden) keine andere Bedeutung haben als die, für den erforderlichen Bewegungsspielraum bei der Produktion und Interpretation von Daten zu sorgen. Diese Vokabeln signalisieren einen Abstand zu akademischen Problemstellungen, nationalen Vorlieben und disziplinären Paradigmen, ohne dass man in vielen Einzelfällen bereits wüsste, ob nicht gerade Akademismen, nationale Wissenschaftskulturen oder Disziplinen die leistungsfähigeren Metadaten liefern würden. Das ist hier nicht zu entscheiden. Entscheidend ist einstweilen nur, dass Diskussionen dieser Art präzise Anlässe haben und dazu führen, Fragestellungen neu zu sortieren, die angesichts der unabsehbaren Folgen der Verfügbarkeit neuer Datenmengen und ihrer algorithmischen Verknüpfung neu sortiert werden müssen.

Zur Ontologie und Autologie der Metadaten

Auffällig ist nun, dass noch niemand den Versuch gemacht hat, Begriffe wie »Gesellschaft« oder »Kultur« als Metadaten in die Diskussion um die Kontrolle von Big Data einzuführen. Talcott Parsons und Niklas Luhmann sind zwar nicht die einzigen Soziologen, die – nicht zufällig auf der Grundlage ihrer Systembegriffe – explizit nach Begriffen einer soziologischen Theorie gesucht haben, die der modernen Medienunruhe und der sich ankündigenden Medienkatastrophe der Einführung elektronischer Medien auf die Spur zu kommen versuchen. Auf eine ähnliche Unruhe reagieren auch die Diskurstheorie Michel Foucaults, die Feldtheorie Pierre Bourdieus und die Netzwerktheorien Harrison C. Whites und Bruno Latours, die mehr oder minder genau darum wissen, wie viel sie älteren Problemstellungen der soziologischen Theorie, die sich bei Auguste Comte, Gabriel Tarde oder Georg Simmel finden, verdanken. Aber eine Diskussion mit Mathematikern und Informatikern darüber, wie Gesellschaftstheorien aussehen, die den Zusammenhang von Daten und Metadaten explizieren und variabilisieren können, fehlt bislang.

So oder so fallen mit der Frage danach, welche Heuristik wissenschaftlich (denn es gibt auch eine alltagspraktische, politische, ökonomische, pädagogische, ästhetische und religiöse Heuristik) geeignet sein könnte, die durch Big Data aufs Schlagwort gebrachte Herausforderung anzunehmen, zunächst Begrifflichkeiten in den Blick, die Zeitgenossen der Einführung des Computers und seiner Netzwerke sind, wie etwa die Begriffe des Systems, der Komplexität, des Netzwerks, des Programms. Vor dem Hintergrund dieser Begrifflichkeiten können dann allerdings auch die für die Moderne typischen Begriffe wie eben Gesellschaft und Kultur, aber auch Politik und Wirtschaft, Vernunft und Verstand, Emotion und Rationalität, Körper, Seele und Geist als Metadaten auffällig werden (vgl. z. B. Latour 1998), die allesamt die Funktion der Codierung, Rahmung und Verknüpfung von Daten haben. Es sind Begriffe, die es uns erlauben, die Welt zu verstehen, indem sie es uns erlauben, Verbindungen zwischen empirischen Sachverhalten zu ziehen, die erst dank dieser Verbindungen konstruiert werden können. Und es sind Sachverhalte, die materieller, temporaler, mentaler, sozialer oder artifiziieller Art sein können, eine Unterscheidung, die ihrerseits Metadatenstatus hat, denn was »sind« die Materie, die Zeit, der Geist, die Kommunikation oder das Artefakt anderes als ihrerseits hoch problematische Konstruktionen von möglicherweise zu ihnen passenden Daten?

Die wichtigste Eigenschaft eines Metadatum besteht darin, dass es zugleich ein Datum ist. Es ist uns als Ausgangspunkt zur Sortierung eines empirischen Materials gegeben, dem es selber angehört. Es trifft eine Aussage über die Welt, die in dieser Welt überprüfbar sein muss. Und es trifft diese Aussage anhand der Frage, ob es in der Lage ist, empirisches Material zu sortieren, dem es selber angehört. Wie die Idee in der Philosophie Hegels ist das Metadatum zugleich Begriff, Form, Substanz und deren operative Verknüpfung (vgl. etwa Hegel 1975 [1830], § 213). Es entwirft eine Ontologie und eine Autologie.

So ist der Begriff der Komplexität selber komplex, der Begriff der Kommunikation kommuniziert, der Begriff des Systems reproduziert und differenziert, der Begriff der Evolution evoluiert, der Begriff des Netzwerks vernetzt, der Begriff der Kooperation kooperiert, der Begriff des Schwarms schwärmt, der Begriff der Handlung handelt, und der Begriff der Form formt. Und spätestens, wenn man sich eingesteht, dass für jeden dieser Begriffe zugleich gilt, was für alle anderen Begriff gilt, also dass der Begriff der Komplexität nicht nur komplex ist, sondern auch kommuniziert, reproduziert und differenziert, evoluiert, vernetzt, kooperiert, schwärmt, handelt und formt und so weiter für jeden der genannten Begriffe, ahnt man, welche architektonischen Möglichkeiten in Begriffen stecken, die als Metadaten zugleich die Qualität eines Datums haben. Man ahnt allerdings auch, dass sich die Arbeit an einer Theorie, die die Verknüpfung der Metadaten untereinander ausbaut und kontrolliert, schnell zu einer nichttrivialen Angelegenheit entwickelt, die auch dadurch nicht erleichtert wird, dass der Begriff der Nicht-Trivialität seit Heinz von Foersters Entwurf einer Kybernetik zweiter Ordnung (von Foerster 2003) selbst zu den hier einschlägigen Metadaten gehört – vom Begriff der Arbeit, wie man seit Sigmund Freud weiß, der das »Durcharbeiten« zur Königsdisziplin der Therapie erklärt hat (vgl. Freud 1992 [1914]), ganz zu schweigen.

Wir beschränken uns auf die gerade genannte Liste und behaupten, dass es sich beim Metadatum, das seinerseits in der Liste als solcher, unvollständig wie sie ist, enthalten ist, um seinerseits um eine Datensammel-, -sortier- und -filtertechnik handelt, die geeignet ist, es mit der Herausforderung durch Big Data aufzunehmen. Die Begriffe der Komplexität, der Kommunikation, des Systems, der Evolution, des Netzwerks, der Kooperation, des Schwarms und der Handlung, aber nicht nur diese sind geeignet, der ihrerseits nichttrivialen Komposition sehr großer Datenmengen und ihrer nachvollziehbare algorithmischer Bearbeitung durch interessierte Behörden, Unternehmen, Institute und Vereine mit einer Einschätzung entgegen-

zutreten, welche Art von Erkenntnissen mit welchen Reichweiten diese Akteure wohl zu erreichen vermögen. Selbst die größte Datensammlung und der schnellste und leistungsfähigste Algorithmus müssen nach Kriterien sortiert und mit Einschränkungen programmiert werden, die prinzipiell dafür sorgen, dass das Ergebnis mit Allwissenheit und umfassender Vorhersehbarkeit nichts zu tun hat. Im Gegenteil, Big Data wird uns umso gefährlicher, je präziser diese Einschränkungen gesetzt werden. Die genannten Begriffe zeichnen sich unter anderem darin aus, dass sie eine solche kritische, die Bedingungen der Möglichkeit überprüfende Beobachtung anleiten können.

Natürlich wird auch die Beobachtung von Big Data von den Datenbanken und Algorithmen, die Big Data ausmachen, beobachten. Spätestens jetzt wird das Datenuniversum, in dem wir uns bewegen, hyperkomplex (es beobachtet sich selbst als komplex) und nicht-trivial. Das System insgesamt, wenn man hier von einem System sprechen möchte, bezieht sich laufend und unberechenbar auch auf sich selber. Dieses System ist jedoch kein gigantischer technischer Apparat oder »Komplex«, wie man früher gesagt hätte, sondern es ist die Gesellschaft im Medium ihrer technischen Möglichkeiten, fasziniert beobachtet und damit betrieben von gegenwärtig (online und offline) mehr als sieben Milliarden Menschen und einer unbekannten Zahl von Algorithmen.

Eine Liste von Metadaten

(1) Das wichtigste Metadatum von allen ist sicherlich das Datum der *Komplexität*. Es bedeutet, dass wir es auch im Umgang mit Big Data mit Phänomenen zu tun haben, die aus zu vielen Faktoren bestehen, um kausal, und aus zu unterschiedlichen Elementen, um statistisch erklärt werden zu können (vgl. Weaver 1948). Stattdessen bestehen Big-Data-Phänomene aus einer großen Zahl unterschiedlicher Ele-

mente und Verknüpfungen, die immer nur selektiv wahrgenommen werden können und sich im Zeitablauf verändern.

Ob Big-Data-Phänomene ihre eigenen Komplexität zu bewältigen vermögen, indem sie wie ein Gehirn, ein Bewusstsein oder die Gesellschaft zur Selbstorganisation fähig werden, ist gegenwärtig eine der offenen Fragen. Sicherlich hängt die Antwort auf diese Frage auch davon ab, ob man annehmen kann, dass »die Technik« sich zu einem System zu formieren vermag. Ich würde das bezweifeln. Ich würde allenfalls annehmen, dass zu den Datenmengen und Algorithmen auch die Nutzer und Programmierer zu zählen wären, die die Datenbanken füttern und die Algorithmen entwerfen, um diese Elementen in laufender Auseinandersetzung miteinander als sich selbst organisierendes System beobachten zu können. Notwendig ist dies für die Anwendbarkeit des Begriffs der Komplexität jedoch nicht.

Interessanter ist so oder so, dass der Begriff der Komplexität, verstanden als Einheit einer Vielfalt, Phänomene beschreibt, deren wichtigstes Merkmal eine unreduzierbare Differenz ist. Eine minimale Komplexität ist ein Paar von Zahlen, Formeln, Aussagen oder anderen Einheiten, die zugleich aufeinander angewiesen und nicht aufeinander reduzierbar sind. So setzt Big Data mindestens voraus, dass jede einzelne Datenspur in jedem Moment inkommensurabel auf jemanden zurückgeführt werden kann, der sie legt, und jemand anderen, der sie liest. Und vom Legen und Lesen von Spuren reden wir grundsätzlich unter dem Gesichtspunkt einer Unentscheidbarkeit der Frage, ob wir es mit Menschen oder Algorithmen zu tun haben.

Diese Unentscheidbarkeit hindert uns nicht etwa am Umgang mit Big Data, sondern ist selbst ein produktives Motiv dieses Umgangs, da sie ihrerseits die Frage offen zu halten erlaubt, ob das Datum, mit dem wir es konkret in einer spezifischen Situation zu tun haben, das Datum eines auffindbaren Elements oder einer Verknüpfung von Elementen ist. Letztlich muss das eine sich auch als das andere

bewähren, das heißt die Verknüpfung muss ein Element sein und jedes Element eine Verknüpfung darstellen. Deswegen ist die Frage, ob Mensch oder Maschine einen Impuls gesetzt haben, irrelevant.

(2) *Kommunikation* soll heißen, dass wir es bei jedem Big-Data-Ereignis mit einer Selektion in einem Auswahlbereich von Möglichkeiten zu tun haben, die als diese Selektion ihrerseits ein Datum für eine weitere Selektion ist, für die dasselbe gilt (vgl. Shannon/Weaver 1963 [1948]). Diese Akzentuierung von Selektion in einem Auswahlbereich von Möglichkeiten unterscheidet das Metadatum der Kommunikation vom Metadatum der Kausalität, das die Wissenschaft zumindest wissenschaftstheoretisch, also dogmatisch, aber nicht unbedingt heuristisch, zweitausend Jahre lang dominiert und auf die Suche nach möglichst eindeutigen Verknüpfungen von Ursachen und Wirkungen festgelegt hat. An die Stelle dieser eindeutigen Verknüpfungen treten mehrdeutige Verknüpfungen, die den Vor- und den Nachteil haben, dass sie jemanden voraussetzen, der jeweils eine Deutung, eine Interpretation, vornimmt (vgl. Peirce 1983).

Mit dem Metadatum der Kommunikation wird somit zugleich das Metadatum des *Beobachters* eingeführt, so dass wir es bei Kommunikation grundsätzlich und ausschließlich mit unbedingbaren Setzungen bedingter Beziehungen, oder auch: mit Beziehungen der Abhängigkeit zwischen unabhängigen Einheiten, zu tun haben (vgl. Baecker 2013). Das ist insofern ein Nachteil, als wir nun mit unberechenbaren Selektivitäten rechnen müssen. Und es ist insofern ein Vorteil, als diese unberechenbaren Selektivitäten als Subjekte im kantschen Sinne der Selbstsetzung gelten können und wir uns selbst zu diesen Subjekten zählen dürfen. Big Data ist als ein kommunikatives Phänomen durchsetzt mit Setzungen, die sich voneinander unabhängig auf Abhängigkeitsbeziehungen einlassen. Der Gesamtzustand dieses Phänomens kann allenfalls mit dem Stichwort und Metadatum der funktionalen Turbulenz beschrieben werden, wenn »Turbulenz« heißt, dass jedes Element laufend

mit überraschenden Rückkopplungen eigener Aktionen konfrontiert wird, und »funktional« heißt, dass keine dieser Rückkopplungen nicht ihrerseits auf situativ bestimmte und kontextuell unbestimmte Art und Weise spezifische Bedingungen setzt und Effekte auslöst – und sei es die Produktion von Abfall, der jedoch im nächsten Moment zu einer unschätzbaren Datenquelle werden kann.

(3) Mit dem Metadatum des *Systems* wird die Annahme und Beobachterdirektive formuliert, dass Big Data unter drei und nur drei Perspektiven untersucht werden kann: als Ereignis in einem System, als Struktur eines Systems oder als Phänomen in der Umwelt eines Systems. Interessant ist dies schon deswegen, weil alle drei Systemreferenzen sich voneinander unterscheiden können, ohne dass sie deswegen füreinander irrelevant wären. Im Gegenteil, sie formieren sich nichtsummativ zur bereits genannten funktionalen Turbulenz, innerhalb derer jedes Ereignis einen Unterschied machen kann, aber nicht muss.

Big Data kann ein Ereignis in einem System sein, das wir Gesellschaft nennen können und für das wir, wenn wir einen hinreichend weiten historischen Bogen spannen wollen, sagen können, dass Big Data eine Singularität, eine evolutionäre Universalie darstellt, nach deren Auftreten nichts mehr ist, wie es vorher war. Das kann natürlich vielerlei heißen. Es kann sein, dass wir feststellen werden, dass Big Data der Moment war, der dieses System unter seiner eigenen Komplexität zusammenbrechen ließ, so dass sich danach und nach den damit verbundenen Katastrophen nur noch technikfeindliche Fundamentalistengruppen auf der Erde halten konnten, die patriarchalisch oder matriarchalisch in ihren Häusern sitzen, Schafe hüten und Wolle spinnen. Es kann aber auch sein, dass sich nach diesem Ereignis alle gewohnten Institutionen der Verwaltung von Macht, Geld, Glauben und Wahrheit auflösen und wir es nur noch mit piratenhaften Verknüpfungen von Kompetenzen zu tun haben, wie man sie für die sogenannte Wissensgesellschaft beschrieben hat.

Etwas weniger dramatisch geht es zu, wenn wir Big Data als Struktur eines Systems betrachten. Strukturen dienen der Verknüpfung von Ereignissen und müssen selbst als Ereignisse oder zumindest Ereignisaspekte auftreten, um dies leisten zu können. Auch das System, das wir uns im Hinblick auf Big Data als Struktur anschauen, können wir Gesellschaft nennen, doch jetzt können wir genauer fragen, welche alten Strukturen durch diese neue Struktur bedrängt, ersetzt oder auch ergänzt werden und welche neuen Verknüpfungen welcher Ereignisse durch die neue Struktur ermöglicht werden. Im Grunde genommen ist dies die Leitfrage, unter der der vorliegende Text geschrieben ist. Das heißt, wir nehmen an, dass sich in der Tat alles ändert, seit Big Data auftritt, aber alles nur ein wenig. Die Gesellschaft hat es nach wie vor mit Politik und Wirtschaft, Kunst und Erziehung, Recht und Religion, Technik und Kultur zu tun, muss jedoch zusätzlich mit Datenmengen und Algorithmen, von den durch sie beeindruckten Programmierern, Hackern, Bloggern und Surfern ganz zu schweigen, fertig werden, die weder am Datum noch am Typ der Verknüpfung von Daten feststellen können, ob sie es mit Politik oder Wirtschaft, Kunst oder Erziehung, Recht oder Religion, Technik oder Kultur zu tun haben. Wieder tritt jene funktionale Turbulenz auf, die jedoch zusätzliches Material aus Daten gewinnt, an denen sich die Ordnung der modernen Gesellschaft, der Buchdruckgesellschaft, orientiert hatte und die erkennbar zu einer so nicht mehr erkennbaren Ordnung der Gesellschaft beitragen.

Dazu passt dann auch unsere dritte Möglichkeit, die Untersuchung von Big Data als Phänomen in der Umwelt von Systemen, zu denen wir alle möglichen einschlägigen Kandidaten zählen können.ⁱⁱ Interaktionssysteme im Sinne von Gesprächskreisen aller Art (Partytalk, Seminargruppen, Gremiensitzungen, Konferenzen, Parteitage ...) regen sich darüber auf, dass sie nicht wissen, was ihnen mit Big Data bevorsteht, füttern jedoch zugleich die Social Media, die ihnen wiederum darüber Auskunft geben, worüber sie sich aufregen können. Organisationssysteme wie Be-

hörden, Unternehmen, Kirchen, Krankenhäuser, Armeen und Vereine wehren sich mit Händen und Füßen gegen die Subversion ihrer Kompetenzprofile durch Datenspeicher, können aber mit ihrer Konkurrenz nur mithalten, wenn sie diese Subversion zugleich mitbetreiben.

Funktionssysteme wie die der modernen Gesellschaft (also Politik und Wirtschaft, Kunst und Erziehung, Recht und Religion, Wissenschaft und Massenmedien) verlieren schließlich im selben Maße die Anschlusssicherheit ihrer Operationen, wie sie sich zur Bewahrung und Steigerung dieser Anschlusssicherheit auf die Unterstützung ihrer Kalküle durch Datenbestände und Algorithmen einlassen. Die Macht muss erleben, dass allenfalls noch Terroristen und Notenbanken ihrer Willkür bedürfen, weil sich alle anderen politischen Maßnahmen viel legitimer und durchgriffssicherer in Datenbanken ausrechnen lassen. Das Kapital muss erleben, dass selbst ständige Liquiditätsspritzen unter Niedrig-, Null- und Negativzinsbedingungen nichts daran ändern, dass man über Kosten und Gewinne nur vor Ort und nur anhand hochreagibler Entwickler-, Lieferanten- und Kundenaktionen entscheiden kann. Die Kunst verliert ihren Status der kreativ-genialen Autonomie, seit sie schneller katalogisiert und archiviert wird, als sie produziert werden kann, und jeder Kunststudent in der Lage ist, sich in diesen Archiven seine eigenen Ideen zusammensuchen, für die sich milieu- und nicht zuletzt bildungsspezifisch auch fast immer ein Publikum findet. Die Erziehung beobachtet fassungslos die Abschaffung des Lehrpersonals, seit sich herumspricht, dass kein Lerneffekt verlässlicher ist als der des Zeigefingers auf einem Tabletcomputer, der Datenbestände aufeinander bezieht, die dann entweder Korrektheit oder Irrtum oder, am besten, die nächste Herausforderung signalisieren. Das Recht wird nur noch von denen in Anspruch genommen, die mit der Schlichtung ihrer Streitigkeiten öffentliche Aufmerksamkeit gewinnen wollen (und sei es in der Vertretung des Staats durch Staatsanwälte), um so neue Rechtspositionen zu begründen, die ohne Öffentlich-

keit neue Daten schaffen. Die Religion verflüchtigt sich in Pfingstbewegungen, die nur noch vom sprichwörtlich fluiden Geist irgendeine Art der Erlösung erhoffen. Die Wissenschaft transformiert sich in eine Ansammlung von Daten mit Impact-Faktoren, die unterschiedslos durch Texte, Autoren, Institute, Universitäten oder ganze Disziplinen hinterlegt sein können und nichts anderes signalisieren als die Menge weiterer Daten, die auf die erstgenannten Daten Bezug nehmen. Und die Massenmedien sind vielleicht das einzige Funktionssystem, das seine Grenzen scharf halten kann, weil es früher als alle anderen gelernt hat, Information von Nicht-Information zu unterscheiden (vgl. Luhmann 1996) und weil diese Unterscheidung sich anschiebt, zum Supercode der sogenannten Wissensgesellschaft zu werden. Aber das hat den Preis, dass die Monopole aller Redaktionen sich auflösen und jeder Nutzer zum Redakteur seiner Datenwolke wird. Eine Gesellschaft von Buribunken, um es mit Carl Schmitt zu formulieren (Schmitt 1917/1918).

Big Data als Phänomen in der Umwelt eines Systems sichert und gefährdet die Differenzierung und Reproduktion dieses Systems im selben Maße, in dem es dem System noch gelingt und nicht mehr gelingt, sich von diesem Phänomen zu unterscheiden und im Medium der Störung durch dieses Phänomen zu erhalten. Wetten auf das Schicksal der Funktionssysteme in einer nicht länger modernen Gesellschaft werden nicht mehr angenommen.

(4) Unter dem Metadatum der *Evolution* wollen wir keine Perspektive verstehen, die Daten daraufhin untersucht, wie es ihnen gelingt, sich im Überlebenskampf gegen andere Daten als besonders anpassungsfähig oder störunanfällig und damit robust zu behaupten – obwohl auch das eine interessante Fragestellung ist. Von wie vielen Daten hört man kurz nach ihrer Produktion nie wieder etwas? Welche Daten erwerben auf welche Weise jene Prominenz, die den Nutzern die freie Aussicht auf andere Daten verdirbt? Trotz dieser durchaus wichtigen Fragen wollen wir das Metadatum Evolution hier neodarwinistisch verstehen (vgl. dazu etwa Campbell

1969), das heißt drei evolutionäre Mechanismen unterscheiden (Variation, Selektion und Retention) und nicht vom Überleben der vermeintlich Tüchtigsten reden. Eine Untersuchung von Big Data bekommt durch den Einbezug einer evolutionären Perspektive einen Zugriff auf eine zeitliche Dynamik, die darin ihre Pointe hat, dass die Zeit hier nicht als passiver Maßstab irgendwelcher auf einer Zeitskala abzutragenden Ereignisse, sondern als aktives Moment der Produktion von Ereignissen gewürdigt werden kann.

Die Evolutionstheorie ist eine Theorie, die zeigt, wie die Zeit Unterschiede macht. Big Data tritt als Variation auf und zwingt andere Systeme, psychische wie soziale, zu einer entweder negativen oder positiven Selektion. Die negative Selektion grenzt das Ereignis der Variation zwar aus dem aktuellen Relevanzbereich des Systems aus, doch lauert es anschließend jenseits der Grenze dieses Relevanzbereichs darauf, dass ein anderes System es positiv selektiert und dem ersten System die weitere Exklusion unmöglich wird. Die evolutionären Effekte verstärken sich, wenn Systeme klug genug sind, die Variation nicht nur positiv zu selektieren, sondern wenn sie die Evaluation dieser Variation zugleich zu einer eigenen Selektionsebene ausgestalten. Lassen wir doch Big Data mit Big Data fertig werden, lautet die dahinterstehende Überlegung: Wenn uns schon Datenmengen und Algorithmen behelligen, sollten wir diese Datenmengen mit eigenen Algorithmen auf auch für uns nützliche Datenmengen hin auswerten. Zumeist merkt man zu spät, dass man auf diese Art und Weise nicht etwa ein technisches Problem technisch behebt, also schlicht Gelegenheiten in Instrumente verwandelt, die von den gewohnten Instanzen souverän gehandhabt werden können, sondern dass man unangenehmerweise mit diesen Instrumenten neue Medien bereitstellt, in denen sich andere Instanzen, unbekümmert um bisherige Ansprüche, ihre eigene Souveränität aufbauen. Die Selektionsebenen machen sich selbständig.

Und spätestens jetzt kommt der dritte evolutionäre Mechanismus, die Retention oder Restabilisierung, ins Spiel, weil es jetzt entweder gelingt, das positiv selektierende System mit den neuen Selektionsinstanzen in ein neues operatives (Un-)Gleichgewicht zu bringen, oder weil dies misslingt.

(5) Der *Netzwerk*begriff kann sich gegenwärtig dank des Internets und mit ihm verbundener Turbulenzen in der Gesellschaft über mangelnde Prominenz nicht beschweren, deswegen beschränken wir uns hier auf eine soziologische Variante dieses Begriffs, die das Netzwerk mit Harrison C. White als Ungewissheitskalkül versteht (vgl. White 1992). Auch Datenhaufen sind ein Netzwerk, in dem jedes Datum und jede algorithmische Verknüpfung zwischen Daten jederzeit unter dem Druck alternativer Interessen und Algorithmen stehen. Wer glaubt, sich in diesem Netzwerk konsumtiv, produktiv und lukrativ bewegen zu können, schafft damit für andere Akteure Gelegenheiten der Verknüpfung, die unter Umständen exakt diejenigen Ressourcen entwerten, wegen derer man sich gerade noch engagiert hat, oder wiederum anderen Akteuren Daten liefern, die den eigenen Spielraum einschränken. Identitäten stehen in diesem Netzwerk laufend neu auf dem Spiel. Und Kontrolle ist nur möglich, wenn man sich klar macht, dass man die Kontrolle in dem Moment aufgegeben hat, in dem man jemandem Daten liefert, die nicht nur für diesen im Rahmen der eigenen Beziehung, sondern auch für unbekannte Dritte aufschlussreich sein können.

(6) Das Metadatum der *Kooperation* reagiert bereits, ohne es unbedingt zu wissen, auf das Metadatum eines Ungewissheitskalküls im Netzwerk. Das Metadatum der Kooperation ist der rote Faden einer sich im Medium von Datenmengen und Rechnern neu erfindenden Sozialwissenschaft, in der Ökonomen, Physiker, Informatiker, Psychologen und Soziologen, natürlich streng interdisziplinär, nach Regeln oder auch nur Mustern der Verknüpfung von Nutzerspuren suchen. Dieser auch »computational social science« genannte Wissenschaftszweig tritt mit dem Charme

auf, so zu tun, als wüsste man nichts mehr von den historisch bewährten Problemstellungen der beteiligten Disziplinen und könne mit einem »clean slate« neu beginnen, Logiken der Produktion, Reproduktion und Verknüpfung von Daten zu entdecken (vgl. Conte et al. 2012). Statistische Korrelationen, Regressionen und Frequenzen werden auf ihre Chancen der Emergenz, aber auch des Zerfalls von wiedererkennbaren Mustern hin untersucht.

Als Leitfaden dient dabei ein Begriff der Kooperation, der seit Robert Axelrod auf spieltheoretische Überlegungen dazu abstellt, wie Akteure mit einem möglichst minimalen Schatz an Erfahrungen, das heißt an behindernden Lerneffekten, zu möglichst stabilen Mustern der wechselseitigen Anpassung und Orientierung finden (vgl. Axelrod 1984). Vier Regeln scheinen sich dabei in Simulationen sogenannter Multi-Agenten-Systeme dann zu bewähren, wenn man damit rechnet, einem Mitspieler wieder zu begegnen: (a) der andere darf um seine Erfolge nicht beneidet werden, denn das vergiftet das Spiel; (b) man darf nicht der erste sein, der sich absetzt, denn dann kommt man nicht wieder hinein; (c) jede Kooperationsbereitschaft oder Absetzung muss sofort mit Gleichem erwidert (TIT FOR TAT) und dann vergessen werden, das heißt: »sei weder im Positiven noch im Negativen nachtragend; beides macht dich blind«; und (d) versuche nicht, zu klug zu sein, denn das läuft voraussichtlich nur auf eine Selbsttäuschung hinaus.

Die vierte Regel schützt vor Selbstreferenz, infiniten Regressen, Paradoxien und nicht zuletzt vor dem Problem der doppelten Kontingenz, das die Spieltheorie seit Oskar Morgenstern heimsucht. Und die ersten drei Regeln kann man eventuell auf die eine des Improvisationstheaters reduzieren, nach der es gilt, jede Aktion eines Mitspielers zu bejahen und anschließend nach eigenen Vorstellungen fortzusetzen (vgl. Johnstone 1993). Im Umgang mit Big Data bedeutet das, dort anzuknüpfen, wo eigene Anschlüsse möglich scheinen, alles andere auf sich beruhen zu lassen und nur darauf zu achten, dass man nicht angegriffen wird. Es liegt auf der Hand,

dass Angriff und Verknüpfung vor allem kurzfristig nicht voneinander zu unterscheiden sind, so dass man laufend mit der Unruhe derer rechnen muss, die glauben, Zug um Zug einen Angriff abwehren zu müssen, und dabei übersehen, dass dies auf eine Art und Weise erwidert werden kann, die sich unter Umständen rasch zu einer Welle der Empörung aufschaukelt, die aber möglicherweise genauso schnell wieder verschwindet.

(7) Von hier ist es nicht weit zum Metadatum des *Schwarms*, das aber nicht nur ein besonders robustes Format der Kooperation einführt, sondern zugleich die Perspektive wechselt. Neben der Frage nach den Regeln auf der individuellen Ebene kooperativer Spiele wird jetzt auch die Frage nach Verhaltensdynamiken auf der emergenten Ebene des Schwarms gestellt. Und man sieht, dass unter den Bedingungen der Verstärkung gewünschter Verhaltensweisen, der Abschwächung nicht-gewünschter Verhaltensweisen und einer individuellen Toleranz für kleinere Abweichungen bereits minimale Fluktuationen in den wechselseitig aufeinander Rücksicht nehmenden Verhaltensweisen zu unvorhersehbaren Verhaltensänderungen des Schwarms insgesamt führen können (vgl. Bonabeau et al. 1999).

Big Data ist sicherlich selbst ein Schwarmphänomen, und dies nicht nur im Hinblick auf ein sich selbst verstärkendes Interesse am Phänomen selber, sondern auch im Hinblick auf die beiden Fragen, welche Datentypen hier auf ein besonderes Interesse stoßen und mit welchen Algorithmen in welchen Programmiersprachen jeweils gearbeitet wird. Wenn die Schwarmtheorie festhält, dass Schwärme geeignet sind, zunächst homogene Medien mehr oder minder hochgradig zu differenzieren, so gilt dies auch für das Medium der großen Datenmengen, innerhalb dessen eben nicht Datum gleich Datum, geschweige denn Datenbank gleich Datenbank und Algorithmus gleich Algorithmus sind. Selbst die größte Datenmenge ist selektiv, und selbst die schnellsten Algorithmen sind von Voraussetzungen abhängig, unter denen sie arbeiten können, so dass Big Data in seiner Erfassung und Errechnung

weltweit interessanter Daten Schlagseiten, Lücken und Tendenzen aufweist, die zum einen dafür verantwortlich sind, dass wir auf viele wichtige Fragen dennoch keine Antwort finden, und zum anderen genau deswegen Aktionen auslösen, die neue Datentypen definieren, andere Datenbanken anlegen und mit anderen Algorithmen arbeiten. Der Schwarm ist in Bewegung, und er bleibt in Bewegung. Man denke beispielhaft nur an die Fluktuationen des Schwarms in Fragen der Entwicklung eines alternativen Geldes, alternativer Kapitalsammelstellen und alternativer Kreditvergabeplattformen.

(8) Quer zu Schwärmen, obwohl selbst als Schwarmphänomen auftretend, ist jüngst wieder ein verstärktes Interesse an einem Metadatum zu verzeichnen, dem man vielleicht am besten den Namen *Handlung* gibt, seit der Name des Subjekts nicht zuletzt dank Slavoj Žižeks Dekonstruktion des cartesianischen Subjekts deutlich an Resonanzfähigkeit verloren hat (vgl. etwa Žižek 2001). Die Handlung muss an die Stelle des sich selbst setzenden Subjekts treten, wobei man etwa im Anschluss an eine Frühschrift Michail Bachtins zur Philosophie der Handlung oder an Michel de Certeaus Interesse an einer aus dem Alltag gewonnenen Kunst des Handelns die Gelegenheit nutzt, transzendente Referenzen auf die Vernunft durch empirische Referenzen auf die Erfahrung zu ersetzen und dennoch die Autonomie der Handlung zu betonen (vgl. Bachtin 2011 [1919]; de Certeau 1988).

Im Umgang mit Big Data ist dies ein doppelt interessantes Metadatum. Denn zum einen kann nur die Idee der Autonomie der Handlung deutlich machen, warum man überhaupt ein Interesse an der Sammlung großer Datenmengen haben kann. Allein im Hinblick auf eine Autonomie der Handlung kann die Idee einer weitgehenden Unabhängigkeit der Daten aufrechterhalten werden, da andernfalls jedes Datum aus jedem anderen errechnet werden könnte und man es statt mit hoher Varietät mit großer Redundanz menschlichen Handelns zu tun hätte. Big Data verliert jedoch umso mehr von seiner Relevanz, je höher diese Redundanz ist. Big Da-

ta lebt davon, der Varianz auf die Spur zu kommen. Big Data lebt davon, gerade auf dem Feld der Varianz sich selbst und den Gegenstand überraschende Prognosen zu Verkehrsflüssen, Kundeninteressen und selbstorganisierten Formen der Vernetzung von arbeitsteiliger Arbeit machen zu können.

Und zum anderen begegnet das Metadatum der Handlung diesem Interesse an einem Versuch, überraschendes Handeln überraschend vorhersagen zu können, indem es gegen jede Berechnung des Verhaltens dessen Unberechenbarkeit betont. Nicht ohne Ironie werden den Datensaugern von Big Data die Inkommensurabilitäten situativer Erfahrung, körperlicher Befindlichkeit, mentaler Idiosynkrasie und sozialer Performanz gegenübergestellt, die eine ihrem Namen gerecht werdende verantwortliche Handlung eines Subjektes kennzeichnen, dieses Subjekt und seine Handlung jedoch zugleich in neue Datentypen zerlegen, die zueinander nicht passen wollen. Es geht um den Einwand des Lebens gegen die Daten und damit vielleicht um andere, aber eben auch um neue Daten.

Aus der Einheit der Differenz von Varianz und Inkommensurabilität, wenn ich das so formulieren darf, bezieht David Weinberger ein starkes Argument gegen alle Versuche, auf eine Redundanz der formalen oder logischen Kohärenz aller Daten im Big-Data-Netzwerk zu setzen (Weinberger 2011). Sicherlich gibt es diese Kohärenz. Man braucht nur hinreichend stark zu generalisieren, dann ist jedes Datum redundanterweise nichts anderes als ein Datum. Und man braucht nur hinreichend stark zu spezifizieren, dann ist redundanterweise kein Datum mit irgendeinem anderen Datum identisch. Tatsächlich liegen jedoch alle Fragen, die für uns Menschen im fragilen Netzwerk einer Gesellschaft auf einem noch einigermaßen ökologisch robusten Planeten interessant sind, auf mehreren Ebenen zwischen dieser zu starken Generalisierung und zu starken Spezifizierung. Sie stellen sich lokal auf der Ebene von Praktiken, regional auf der Ebene von Medien und global auf der Ebene von Kontexten, die jeweils andere Evidenzen, Potenziale und Kritikalitäten

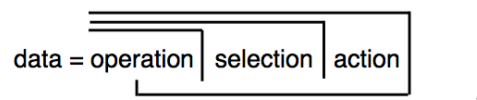
(im Sinne empfindlicher Schwellenwerte) aufweisen und in dieser Kombination, das ist das Argument von Weinberger, allen Versuchen widersprechen, im Umgang mit komplexen Systemen vorhersagefähig zu werden. Stattdessen, so Weinberger, kommt es auf die Entwicklung von Modellen der verteilten Produktion von Wissen an (Creative Commons), die ebenso rhizomatisch wie selbstähnlich – man denke an die Plateaus von Gilles Deleuze und Felix Guattari (Deleuze/Guattari 1980) –, immer nur Verknüpfungen begrenzter Reichweite vornehmen, weil nur diese gegen die zu rasche und fatale Diffusion von Irrtümern immun sind.

(9) Schließlich darf auf unserer unvollständigen Liste möglicher Metadaten zur Bewältigung der Herausforderung von Big Data der Begriff der *Form* nicht fehlen, wie ihn George Spencer-Brown mit seinem Formkalkül propagiert (Spencer-Brown 2008 [1969]). Das Metadatum der Form erlaubt es, wie Tina Piazzzi und Stefan M. Seydel gezeigt haben (Piazzzi/Seydel 2010), Formen der Unruhe zu bearbeiten, die auf der einen Seite Formen der Kommunikation in den Medien des Computers, des Internets sowie des Big-Data-Phänomens sind, auf der anderen Seite Formen der kreativen Bewegung in diesen Medien. Das hat viel mit der gerade erwähnten Dialektik der Handlung zu tun, geht jedoch insofern darüber hinaus, als der Formbegriff zugleich die nach Belieben differenzierungsfähige Innenseite einer Handlung und die radikale Offenheit der Außenseite derselben Handlung zu betonen vermag.

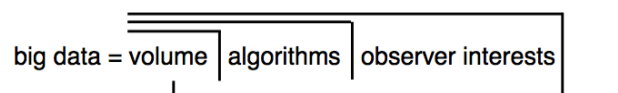
Der Begriff der Form kann darüber hinaus dazu herangezogen werden, die Operation, Struktur und Verknüpfung der Daten im Rahmen von Big Data selber genauer zu untersuchen. Denn »Form« heißt, dass jedes Datum einen Unterschied macht, sofern es überhaupt auftritt, mit diesem Unterschied auf einen Beobachter aufmerksam macht, der so und nicht anders unterscheidet und beobachtet und damit einen Raum weiterer Unterscheidungen und damit Daten aufruft, die in einem nichtbeliebigen Zusammenhang zum ersten Datum stehen. Wenn man so will,

betont dieses Metadatum eine Gegenperspektive zu jener der Handlung, die ihrerseits versucht, sich vom Schwarm abzusetzen. Das Metadatum der Form formuliert die formalen und damit operationalen und funktionalen Bedingungen, unter denen Daten in einem komplexen, kommunikativen, systemischen, evolutionären und vernetzten Raum unter Bedingungen der Kooperation, Schwarmbildung und Handlung unterschiedlich und vergleichbar zugleich sein können. Diese Analyse hat noch niemand durchgeführt, aber es wäre interessant, Big Data daraufhin zu untersuchen, ob und wenn ja welche Artikulation heterogenen Daten zugrunde liegt und welche Art von Zugriffen (Codierungen) welche Art von Algorithmen in welchen Datenmengen vornehmen um welche Art von Ordnung herzustellen.

Offener können wir die hier interessierende Fragestellung nicht formulieren, aber nur so wird man der Frage auf die Spur kommen, wie die Komplexität, Selektivität und Rekursivität von Big Data beschaffen und aufeinander bezogen sind. Wenn jedes einzelne Datum (*data*) mindestens die Form des Tripels von Operation (*operation*), Selektion (*selection*) und Aktion (*action*) hat,



das heißt eine Codierung vornimmt, dabei einer bestimmten Struktur gehorcht und auf eine bestimmte Art und Weise adressiert und gelesen werden kann, dann fragt sich, ob man auch Big Data die Form eines Tripels zuordnen kann. Die These, dass wir es mit einem Tripel von Datenmengen (*volume*), Algorithmen (*algorithms*) und Beobachterinteressen (*observer interests*) zu tun haben,



ist zwar nur eine erste These, aber sie möge uns an dieser Stelle als Zusammenfassung unserer Überlegungen dienen, ein weiteres Beispiel für die hier nicht weiter erläuterebedürftige Notation einer Form liefern, für die Formulierung von Tripeln werben, wie sie für den Qualitätssprung vom Internet zum semantischen Netz vorgeschlagen wird (vgl. Berners-Lee et al. 2001), und ein offenes Forschungsprogramm definieren.

(10) Und nicht zuletzt müssen wir das Metadatum des *Metadatum*s selber nennen.ⁱⁱⁱ Denn es versteht sich alles andere als von selbst, dass wir glauben, im Umgang mit Big Data Metadaten bereitstellen zu können. Wie jeder andere Anspruch an Theorien, Modelle, Beschreibungen und Erklärungen formuliert auch die Idee des Metadatum eine Ordnungsvermutung, die schneller auf den Beobachter zurückfallen und sich auf diesen reduzieren kann, als es diesem lieb sein wird. Umgekehrt jedoch halten wir hier gerade wegen dieser Zurechnung auf einen Beobachter an der Idee des Metadatum fest. Wir bringen uns damit selbst ins Spiel. Wir formulieren Metadaten genauso wie Theorien, Modelle, Beschreibungen und Erklärungen als Setzungen und Unterscheidungen eines Beobachters, der damit weitere Daten für andere Beobachter produziert.

Wir müssen das Metadatum des Metadatum schon deswegen offen legen, weil auch hier wieder mehrere Optionen je unterschiedlicher Beobachter möglich sind. Die Annahme, dass die Welt eine Ordnung aufweist, die sich auf den Begriff eines Begriffs bringen lässt, hört in der Statistik auf den Namen der Korrelation, in der Ökonomie auf den Namen der Suboptimalität, in der Theologie auf den Namen der Schöpfung, in den Geisteswissenschaften auf den Namen des Sinns, in der Mathematik auf den Namen der Zahl oder der Kategorie und so weiter und so fort. In der Soziologie hört sie auf den Namen der Gesellschaft, worunter die Annahme verstanden wird, dass der Statik und Dynamik gesellschaftlicher Phänomene, dem Verhältnis der Menschen untereinander und dem Begriff der Gesellschaft selber

die Einheit einer Funktion zugrunde liegt, die diese Elemente der Statik, der Dynamik, des Verhältnisses und des Begriffs als einen nichtbeliebigen Zusammenhang definiert, so schwer es auch fallen mag, diese Funktion auszubuchstabieren (so etwa Adorno 2003 [1954]; Luhmann 1992).

In diesem Sinne haben wir in diesem Text Big Data zum Datum des Metadatum der Gesellschaft erklärt und damit den Anspruch festgehalten, dass die Soziologie Erkenntnisse zu diesem Datum beitragen kann, die darüber hinausgehen, nun ihrerseits Modelle zu entwerfen, die es erlauben, mit den Daten zu rechnen, die Big Data bereitstellt. Denn zugleich wird es darum gehen, Modelle zu entwerfen, in denen Big Data selbst als ein Datum vorkommt, das historisch variiert, sozial selektiert und kulturell unter Umständen diskriminiert.

Literatur

Theodor W. Adorno (2003 [1954]), »Gesellschaft: Erste Fassung eines Soziologischen Exkurses«, in: Rolf Tiedemann (Hg.), *Frankfurter Adorno Blätter VIII*, München: edition text + kritik, S. 143-150.

Robert Axelrod (1984), *The Evolution of Cooperation*, New York: Basic Books 1984.

Dirk Baecker (2013), *Beobachter unter sich. Eine Kulturtheorie*, Berlin: Suhrkamp.

Dirk Baecker (2007), *Studien zur nächsten Gesellschaft*, Frankfurt am Main: Suhrkamp.

Michail M. Bachtin (2011 [1919]), *Zur Philosophie der Handlung*, Berlin: Matthes & Seitz.

Philip Bagley (1968), *Extension of Programming Language Concepts*, Philadelphia: University City Science Center 1968.

Tim Berners-Lee/James Hendler/Ora Lassila (2001), »The semantic web. A new form of web content that is meaningful to computers will unleash a revolution of new possibilities«, in: *Scientific American* 284/5, S. 34-43.

Eric Bonabeau/Marco Dorigo/Guy Theraulaz (1999), *Swarm Intelligence. From Natural to Artificial Systems*, New York: Oxford University Press.

Donald T. Campbell (1969), »Variation and selective retention in socio-cultural evolution«, in: *General Systems* 14, S. 69-85.

Michel de Certeau (1988), *Kunst des Handelns*, Berlin: Merve.

Rosaria Conte et al. (2012), »Manifesto of computational social sciences«, in: *European Physical Journal Special Topics* 214, S. 325-346.

- Thomas A. Cormen (2013), *Algorithms Unlocked*, Cambridge, MA: MIT Press.
- Gilles Deleuze/Felix Guattari (1980), *Mille Plateaux. Capitalisme et schizophrénie 2*, Paris: Les éditions de Minuit.
- Heinz von Foerster (2003), *Understanding Understanding. Essays on Cybernetics and Cognition*, New York: Springer.
- Sigmund Freud (1992 [1914]), »Erinnern, Wiederholen und Durcharbeiten. Weitere Ratschläge zur Technik der Psychoanalyse II« (1914), in: ders., *Zur Dynamik der Übertragung. Behandlungstechnische Schriften*, Frankfurt am Main: Fischer, S. 85-95.
- Georg Wilhelm Friedrich Hegel (1975 [1830]), *Enzyklopädie der philosophischen Wissenschaften im Grundrisse*, herausgegeben von Friedhelm Nicolin und Otto Pöggeler, 7., durchgesehene Auflage, Hamburg: Meiner 1975, § 213.
- Keith Johnstone (1993), *Improvisation und Theater*, Berlin: Alexander Verlag.
- Udo Kittelmann/Claudia Dichter (Hg.) (2013), *George Widener. Secret Universe 4*, Köln: Walther König 2013.
- Bruno Latour (1998), *Wir sind nie modern gewesen. Versuch einer symmetrischen Anthropologie*, Frankfurt am Main: Suhrkamp.
- Niklas Luhmann (1997), *Die Gesellschaft der Gesellschaft*, Frankfurt am Main: Suhrkamp.
- Niklas Luhmann (1996), *Die Realität der Massenmedien*, Opladen: Westdeutscher Verlag.
- Niklas Luhmann (1992), »Die Systemtheorie zwischen Involution und Normativität. Ein Interview von Gerald Breyer und Niels Werber«, in: *Symptome* 10, S. 45-56.
- John McCormick (2012), *Nine Algorithms That Changed the Future. The Ingenious Ideas That Drive Today's Computers*, Princeton: Princeton University Press.
- Marshall McLuhan (1964), *Understanding Media. The Extensions of Man*, New York: McGraw/Hill.
- Walter J. Ong SJ (1977), *Interfaces of the Word. Studies in the Evolution of Consciousness and Culture*, Ithaca: Cornell University Press.
- Tina Piazzzi/ Stefan M. Seydel (2010), *Die Form der Unruhe*, Band 2: *Die Praxis*, Hamburg: Junius.
- Talcott Parsons (1980), *Zur Theorie der sozialen Interaktionsmedien*, herausgegeben von Stefan Jensen, Opladen: Westdeutscher Verlag.
- Charles Sanders Peirce (1983), *Phänomen und Logik der Zeichen*, herausgegeben und übersetzt von Helmut Pape, Frankfurt am Main: Suhrkamp.
- Carl Schmitt (1917/1918), »Die Buribunken«, in: *Summa* 1, S. 89-106.
- Claude E. Shannon/Warren Weaver (1963 [1948]), *The Mathematical Theory of Communication*, Urbana: University of Illinois Press.
- Herbert A. Simon, *Models of Man Social and Rational: Mathematical Essays on Rational Human Behavior in a Social Setting*, New York: John Wiley & Sons 1957.
- George Spencer-Brown (2008 [1969]), *Laws of Form*, Leipzig: Bohmeier.
- René Thom (1980), *Modèles mathématiques de la morphogenèse*, zweite, durchgesehene und erweiterte Auflage, Paris: Christian Bourgeois éditeurs.

Alvin Toffler, *The Future Shock*, New York: Random House 1970.

Martin Warnke (2003), »Daten und Metadaten: Online-Ressourcen für die Bildwissenschaft«, in: *zeitenblicke* 2/1, online verfügbar unter: {<http://www.zeitenblicke.de/2003/01/warnke/>} (Stand Juli 2013).

Warren Weaver 1948, »Science and complexity«, in: *American Scientist* 36, S. 536-544.

David Weinberger (2011), »The machine that would predict the future, in: *Scientific American* 305, S. 32-37.

Harrison C. White (1992), *Identity and Control: A Structural Theory of Action*, Princeton: Princeton University Press.

Slavoj Žižek (2001), *Die Tücke des Subjekts*, Frankfurt am Main: Suhrkamp.

Anmerkungen

ⁱ Insofern müsste die aktuelle Diskussion sich nicht nur anschauen, welche Daten von den Geheimdiensten gesammelt werden, sondern auch mithilfe welcher Konzepte sie ausgewertet werden. Nicht nur in der Sammlung, sondern auch in diesen Konzepten stecken politisch diskussionswürdige und parlamentarisch zu kontrollierende Kriterien. Und nicht nur in der Sammlung, sondern auch in den Konzepten verbergen sich Entscheidungen, die Geheimdienste ihrer Natur nach geheim zu halten geneigt sein werden. – Hat nicht Ulrich Oevermann sein Verfahren der objektiven Hermeneutik in den siebziger Jahren dazu verwendet, Horst Herolds Rasterfahndung so zu verfeinern, dass sie unter der Überlast ihrer Ergebnisse zusammenbrach?

ⁱⁱ Wir beschränken uns hier auf soziale Systeme. Dass jedoch auch in psychischen Systemen der Umgang mit sehr großen Datenmengen und Algorithmen vom Ereignis in der Umwelt zum Ereignis und zur Struktur im System werden kann, dokumentiert George Widener mithilfe einer künstlerischen Bearbeitung seines Asperger-Syndroms (vgl. Kittelmann/Dichter 2013).

ⁱⁱⁱ Ich folge einer Anregung von Athanasios Karafillidis.