

# 2

## Quantitative empirische Sozialforschung

Der Neurophysiologe Wolf Singer schrieb in einem Buchbeitrag mit dem Titel „Verschaltungen legen uns fest: Wir sollten aufhören, von Freiheit zu sprechen“ unter der Kapitelüberschrift „Wie Wissen in den Kopf kommt“: „Alles Wissen, über das ein Gehirn verfügt, residiert in seiner funktionellen Architektur, in der spezifischen Verschaltung der vielen Milliarden Nervenzellen“ (Singer 2013, S. 54). Dies mag als Beispiel für eine Aussage dienen, die klar von einem materialistisch-deterministischen Welt- und Menschenbild (vgl. Abschn. 2.1) ausgeht, welches keineswegs selbstverständlich ist. Der Politikwissenschaftler Alexander Wendt beispielsweise würde auf dem Hintergrund der Erkenntnisse der Quantenphysik als Alternative vorschlagen, den Menschen als „walking wave

function“ zu begreifen (vgl. Wendt 2015, S. 3, 37, 154 usw.) – und von der Nicht-Lokalität dessen, was wir unter „geistige Prozesse“ subsumieren, ausgehen (vgl. Wendt 2015, S. 95–108). Der Biologe und Hirnforscher Gerhard Roth hat offenbar genau diesen Punkt im Auge, wenn er vorsichtiger formuliert: „Das Dogma der heutigen Neurobiologie lautet, dass *alle* Leistungen des Gehirns aus den geschilderten Integrationsleistungen einzelner Nervenzellen resultieren“ (Roth 1998, S. 46; Hervorhebung im Original – wobei ich an dieser Stelle zusätzlich das Wort „Dogma“ herausheben möchte). Hier wird das Dogma als solches benannt.

Die genannten Beispiele mögen illustrieren, dass auch die quantitativ-empirische Forschung auf nicht weiter hinterfragten Axiomen aufbaut (vgl. hierzu auch Bunge und Mahner 2004, S. 64) – und aufbauen muss, wie in Kap. 5 zu zeigen sein wird. Es gilt in der quantitativ-empirischen Sozialforschung das Postulat der Einheitswissenschaft (vgl. hierzu auch Kelle 2008, S. 58, 81; Flick 2010, S. 23–24 oder Bohnsack 2014, S. 16–17), nach dem sich „... die verschiedenen Erfahrungswissenschaften (z. B. Naturwissenschaften wie Physik, Chemie, Biologie oder Sozialwissenschaften wie Ökonomie, Psychologie, Soziologie) lediglich durch den *Gegenstand*, mit dem sie sich befassen, [unterscheiden,] nicht dagegen in der *Art ihres Vorgehens*“ (Kromrey 2009, S. 17; Hervorhebungen im Original). Diese Aussage findet sich fast wortgleich auch bei anderen Autoren, z. B. bei Häder (2010, S. 68), Seipel und Rieker (2003, S. 47), bei Bortz und Döring (2016, S. 48) oder auch schon bei Festinger (2012, S. 13). Als Ausblick

auf Kap. 6 sei auch die Fortführung des Kromrey-Zitats erwähnt: „Die letztere Annahme wird von anderer – von ‚qualitativ‘ orientierter – Seite heftig bestritten“ (Kromrey 2009, S. 18). Hierzu später mehr.

Eine klare Verbalisierung der dem Gedanken einer Einheitswissenschaft zugrunde liegenden Sichtweise findet sich zum Beispiel schon bei Skinner (1973, zit. nach Pervin et al. 2005, S. 433):

Wir können den Weg, den Physik und Biologie eingeschlagen haben, nachvollziehen, wenn wir uns an die Verbindungen des Verhaltens mit der Umwelt halten, dagegen geistige Zustände, die angeblich vermittelnd wirken, bewusst hintenansetzen. Fortschritte in der Physik werden nicht erzielt, weil man sich näher mit dem Frohlocken eines fallenden Körpers befasste, genauso wenig wie es mit der Biologie vorwärtsging, weil sie sich mit der Natur von Lebensgeistern auseinandersetzte; aus demselben Grund aber ist es für eine wissenschaftliche Verhaltensanalyse unnötig, herausfinden zu wollen, in welcher Ordnung Persönlichkeit, Geisteszustände, Empfindungen, Wesenszüge, Pläne, Absichten, Intentionen ... wirksam sind.

## 2.1 Glaubenssätze des materialistisch-deterministischen Weltbildes

Worin bestehen die wichtigsten Glaubenssätze oder Axiome (vgl. hierzu z. B. Schnell et al. 2013, S. 50 oder Bunge und Mahner 2004, S. 4, 18–21, 64, 211, 234), auf denen das materialistisch-deterministische Weltbild gründet? Zunächst

ist da die Unterstellung einer „objektiv existierenden Wirklichkeit“ – unabhängig vom Betrachter. Behnke und Behnke (2006, S. 21) schreiben hierzu treffend:

Die Unterstellung einer objektiv existierenden Wirklichkeit ist durchaus nicht trivial und wird von vielen ernst zu nehmenden Philosophen und Wissenschaftstheoretikern bestritten. Denn schließlich können wir nicht wissen, ob es diese Wirklichkeit gibt, geschweige denn, wie sie aussieht. Daran hindert uns der unvermeidliche Übersetzungsmechanismus, durch den wir die Wirklichkeit wahrnehmen. Alles, was wir von der Wirklichkeit wissen, ist bereits durch unsere Wahrnehmung, Beobachtung, Messinstrumente oder ähnliches gefiltert, die wiederum durch unser Wissen, unsere Theorien und unseren spezifischen Blickwinkel beeinflusst sind. Wir arbeiten also lediglich mit Abbildern der Wirklichkeit, niemals mit der Wirklichkeit selbst.

In den Kap. 3 und 5 wird hierzu mehr zu sagen sein. Vorerst mag die Feststellung genügen, dass es sich bei der Annahme einer „objektiv existierenden Wirklichkeit“, von der im Rahmen der genannten Einheitswissenschaft ausgegangen wird, um einen *Glaubenssatz*, ein Axiom, handelt (vgl. z. B. Schnell et al. 2013, S. 46, Bunge und Mahner 2004, S. 6 oder auch Kanitscheider 2007, S. 24). Der Solipsismus ist empirisch nicht widerlegbar (vgl. z. B. Zeilinger 2005, S. 215, 228).

Zur Demonstration der Zentralität dieses Glaubenssatzes ein Zitat von Helmut Kromrey: „Die in der Literatur zu standardisierten Methoden dominierende Sichtweise, dass ‚Daten‘ Informationen sind, die einer *universell existierenden Realität* unter Einsatz geeigneter Instrumente im Sinne

eines Messvorgangs abzugewinnen sind, wird in der qualitativ-empirischen Sozialforschung überwiegend abgelehnt“ (Kromrey 2009, S. 243; Hervorhebung: Sch.). Relevant ist hier nur der erste Teil des Zitates – zum Standpunkt der qualitativen Forschung später mehr.

Weitere Kernannahmen betreffen die Eigenschaften der objektiv existierenden Wirklichkeit: Sie setzt sich zusammen aus Materie und Energie<sup>1</sup>, die sich gegenseitig beeinflussen und – nach neueren Erkenntnissen – ineinander umwandeln können (vgl. z. B. Greene 2006a, S. 70–71; Greene 2006, S. 398–399 oder auch Zeilinger 2005, S. 135–144). Alexander Wendt (2015, S. 283) spricht von: „... classical mechanics, which describe a purely material world of matter and energy“. Dabei herrschen über die Zeit konstante, kausale Gesetze<sup>2</sup>. Diese Sichtweise

---

<sup>1</sup>Für die hier angestellten Überlegungen aus Sicht der klassischen Physik mag diese grobe – und durchaus hinterfragbare (vgl. z. B. Bunge und Mahner 2004, S. 33) – Vorstellung genügen. Bernulf Kanitscheider etwa präsentiert die These (sic!): „Die Natur ist auf der fundamentalen ontologischen Ebene aus Teilchen, Feldern und Raumzeit aufgebaut, wobei die drei strukturell miteinander verflochten sind und dynamisch interagieren“ (Kanitscheider 2007, S. 67, Hervorhebung im Original). Er betont den „hypothetischen Charakter der philosophischen These des Naturalismus“ (Kanitscheider 2007, S. 68), was sich etwa dadurch zeige, dass durchaus offen sei, „ob Raumzeit wirklich als fundamental zu gelten [habe]“ (Kanitscheider 2007, S. 67), ob Information „möglicherweise in die Liste der Grundbausteine aufgenommen werden [müsse]“ (Kanitscheider 2007, S. 68) oder auch Strings (Kanitscheider 2007, S. 68). Derartige Überlegungen – jenseits der klassischen Physik – spielen allerdings als Grundannahmen der quantitativen, empirischen humanwissenschaftlichen Forschung meines Erachtens kaum eine Rolle.

<sup>2</sup>Vgl. auch – über den Bereich der klassischen Physik hinausgehend – Kanitscheider (2007, S. 68–69, 90) oder Bunge und Mahner (2004, S. 211). Nicht-kausale Gesetze – Bunge und Mahner (2004, S. 97) nennen als Beispiele hierfür etwa die Gesetze der Planetenbewegung – seien in diesem Zusammenhang ausgeklammert, da sie in der humanwissenschaftlichen Forschung kaum eine Rolle spielen.

bezeichne ich im Folgenden als „materialistisch-deterministisches Weltbild“. Danach kann die Entwicklung des Kosmos bis an den Urknall (nach Greene, 2006a, S. 105 und 400 vor etwa 15 Mrd. Jahren; vgl. hierzu auch Greene, 2006, S. 311–312 – ansonsten wird meist von 13.8 Mrd. Jahren ausgegangen) zurückverfolgt werden<sup>3</sup>, wobei die Singularität eine Grenze für wissenschaftliche Forschung darstellt. Über sie hinaus können keine Aussagen gemacht werden. Keine Aussage kann auch über die Entstehung der herrschenden Naturgesetze gemacht werden. Näheres zur Materie und zu den Naturkräften findet sich zum Beispiel in Greene (2006a, S. 21–28) oder zum Thema „Materie und Information“ in Zeilinger (2007, S. 70–73). Für die nachfolgenden Erörterungen in Bezug auf die humanwissenschaftliche Forschung genügt jedoch die oben dargestellte Skizze.

Alexander Wendt sieht die Situation ähnlich: „The dominant model of man in naturalistic social science today is materialist, ontologically deterministic, and mechanistic“ (Wendt 2015, S. 267). Einen damit verbundenen Punkt, welcher seitens der qualitativen Forschung an zentraler

---

<sup>3</sup>Soweit zumindest die bis vor einigen Jahren vorherrschende Ansicht. In neuerer Zeit werden zunehmend jenseits der klassischen Physik alternative theoretische Vorstellungen vorgeschlagen, welche allerdings – soweit ich dies zu beurteilen vermag – für die hier angestellten Überlegungen (zur klassischen Physik) kaum Relevanz aufweisen; allenfalls wäre, wie in Kap. 3 angesprochen, ein „objektiver Zufall“ in Rechnung zu stellen, jedoch keinesfalls – das ist der zentrale Punkt – ein irgendwie gearteter „freier Wille“. Aus diesem Grund bleibt die nachfolgende Darstellung beim „Urknall-Bild“. Generell wirft die Urknall-Vorstellung aus materialistisch-deterministischer Sicht insofern ein Problem auf, als nach dieser Ansicht „nichts aus dem nichts entsteht und nichts ins Nichts verschwindet“ (Bunge und Mahner 2004, S. 211).

Stelle thematisiert wird, stellt er sehr klar und pointiert heraus: „... materialism is a philosophy that privileges death. For materialists dead matter is the norm, the baseline in terms of which life must be explained; the living, in effect, are nothing but assemblages of the dead“ (Wendt 2015, S. 282). Ähnlich argumentiert Carl Rogers (Näheres in Abschn. 6.1): „Analog ließe sich sagen, dass die Wissenschaft eine Autopsie der toten Ereignisse der Therapie durchführen, aber aufgrund gerade ihrer Natur nie in die lebendige Physiologie der Therapie eindringen kann“ (Rogers 2016, S. 210) oder, bezüglich der Terminologie „independent-intervening-dependent variable“: „... diese Begriffe scheinen mir für die Autopsie besser geeignet als für die lebende Physiologie der wissenschaftlichen Arbeit auf einem neuen Gebiet“ (Rogers 2016a, S. 19). Die genannte Sicht der Dinge hat Konsequenzen: Wendt fährt fort: „Not only does that blur the distinction between life and non-life philosophically, it also raises questions about what happens when such a view permeates society, like it does today, such that it becomes »under the ontological dominance of death« ... “ (Wendt 2015, S. 282).

Eine andere Konsequenz: Die in der empirischen Sozialforschung meist mehr oder weniger selbstverständlich als „gegeben“ vorausgesetzte Freiheit des menschlichen Willens ist in diesem Weltbild kaum zu rechtfertigen (vgl. z. B. Singer 2015, S. 12 oder das Themenheft 55(4), 2004 der Psychologischen Rundschau). Nur selten wird dies klar ausgesprochen bzw. zugegeben – wie von dem Psychologen und Kognitionswissenschaftler Wolfgang Prinz (2013, S. 22): „Die Idee eines freien menschlichen Willens ist mit wissenschaftlichen

Überlegungen prinzipiell nicht zu vereinbaren. Wissenschaft geht davon aus, dass alles, was geschieht, seine Ursachen hat und dass man diese Ursachen finden kann. Für mich ist unverständlich, dass jemand, der empirische Wissenschaft betreibt, glauben kann, dass freies, also nicht determiniertes Handeln denkbar ist“. Sofern man dies akzeptiert, ergeben sich gravierende argumentative Konsequenzen. So wäre etwa letzten Endes der Entschluss einer Forscherin, sich zu einem bestimmten Zeitpunkt mit einer bestimmten Frage zu beschäftigen, nicht ihre freie Willensentscheidung, sondern bereits seit unmittelbar nach dem Urknall festgelegt – nebst sämtlicher Rahmenbedingungen und den Ergebnissen ihrer Forschung (vgl. hierzu z. B. Greene 2006a, S. 31–32, 393–394 oder Markowitsch 2004, S. 164). Dieser Punkt wird in Kap. 4 näher erörtert.

Nach dem skizzierten materialistisch-deterministischen Weltbild stellt sich die Aufgabe der quantitativ-empirischen Forschung wie folgt dar: Zu untersuchen sind bestimmte, per Definition abgegrenzte materielle Einheiten – in den Humanwissenschaften also in erster Linie Menschen, aber auch situative Gegebenheiten, denen sie ausgesetzt sind (und die sie eventuell ihrerseits beeinflussen). Diese „Einheiten“ weisen bestimmte Merkmale (oder: Variablen) in einer bestimmten Ausprägung auf. Aufgabe der empirischen Wissenschaft ist zunächst, diese Merkmalsausprägungen möglichst reliabel und valide zu messen und damit ein möglichst genaues Abbild „der Realität“ in den betrachteten Ausschnitten zu erstellen (vgl. hierzu z. B. Bunge und Mahner 2004, S. 211). In der Physik beispielsweise gelingt das relativ gut, in den Humanwissenschaften



hingegen ist dieses Unterfangen mit gravierenden Problemen verbunden, wie der nachfolgende Abschn. 2.2 zeigen wird. Die Ergebnisse der Messungen dienen in erster Linie dazu, die (Kausal-)Gesetze zu erforschen, nach denen die interessierenden „abhängigen Variablen“ eine bestimmte Ausprägung annehmen – etwa die „Stärke des politischen Interesses“ oder die „Intention zur Wahl einer bestimmten politischen Partei“ (vgl. hierzu auch z. B. Flick 2010, S. 23–24 oder Bunge und Mahner 2004, S. 95–96).

Bortz und Döring (2016, S. 48) formulieren an herausgehobener Stelle (zu Beginn des Kapitels „Deduktiv-nomologische Erklärungen der Realität“): „Wissenschaftlicher Erkenntnisgewinn vollzieht sich gemäß Kritischem Rationalismus in den Natur- wie den Sozialwissenschaften durch das Überprüfen von Theorien, die sich auf Ursache-Wirkungs-Relationen (Wenn-Dann-Relationen) in der Wirklichkeit beziehen“. Dass anstelle deterministischer Hypothesen „nur“ probabilistische überprüft würden, liege an der Komplexität des Forschungsgegenstandes und dem damit verbundenen Mangel an Information:

In der empirischen Forschung ... haben wir es jedoch mit *probabilistischen (wahrscheinlichkeitstheoretischen) Gesetzen bzw. Theorien* zu tun. Probabilistische Erklärungen erheben nicht den Anspruch, zwingend für jeden Einzelfall zu gelten, sondern lediglich mit bestimmter Wahrscheinlichkeit zuzutreffen. ... Eine sozialwissenschaftliche Theorie kann in der Regel nicht alle Einflussfaktoren auf menschliches Erleben und Verhalten im Zusammenhang mit dem

untersuchten sozialen Phänomen vollständig aufklären. Indem sie einige wichtige Ursachenfaktoren betrachtet, andere dagegen vernachlässigt, kann sie das Phänomen nur zum Teil erklären und muss abweichende Fälle hinnehmen (Bortz und Döring 2016, S. 49–50; Hervorhebungen im Original).<sup>4</sup>

Zum Menschenbild des materialistisch-deterministischen Weltbildes äußert sich Wendt (2015, S. 189) folgendermaßen:

No model of human beings is complete that does not have room for the *experience* of being human, of what it is like to be you or me. This feeling, consciousness, is such an essential feature of the human condition that a life without it would be hardly worth living at all. ... Yet as suggested in Chapter 1, for fear of Cartesian dualism the main currents of twentieth-century social theory, both mainstream and critical, have run away from experience, seeking to reduce, displace or otherwise marginalize it in their models of man. Human beings are rendered instead into machines or zombies, both ultimately material systems which are able to think and behave but not to feel – transformed, in short, from subjects into objects (Hervorhebung im Original).

---

<sup>4</sup>Positiv hervorzuheben ist, dass anschließend der Gedanke, der Mensch könne über einen „freien Willen“ verfügen, kurz andiskutiert, dann allerdings in keiner Weise weiterverfolgt wird: „Zudem wird auch argumentiert, dass menschliches Erleben und Verhalten möglicherweise gar nicht vollständig erklärbar ist, sondern schon allein die Wirksamkeit des freien Willens zu individuellen Abweichungen von deterministischen Gesetzmäßigkeiten führen würde. Dagegen wird wiederum eingewandt, dass auch vermeintlich freie Willensentscheidungen möglicherweise nach bestimmten Gesetzmäßigkeiten entstehen“ (Bortz und Döring 2016, S. 50). Zu dieser Sicht der Dinge mehr in Kap. 4.

Neben einem Ausblick auf die von qualitativer Seite zu äußernde Kritik zeigt dieses Zitat nochmals sehr pointiert das vorstehend skizzierte Menschenbild.

## 2.2 Charakteristika quantitativer Forschung und interne Kritik

In Abschn. 2.2 geht es nicht darum, Schritt für Schritt den Ablauf quantitativ orientierter empirischer Sozialforschung darzustellen und zu kommentieren. Dies wird als bekannt vorausgesetzt und kann bei Bedarf in diversen Lehrbüchern, etwa Schnell et al. (2013), Diekmann (2009) oder auch Schumann (2012) nachgelesen werden. Vielmehr werden für das Anwendungsbeispiel einer Befragung im Rahmen der Umfrageforschung (Erhebung der Daten mittels standardisierter Interviews) solche Punkte im Forschungsprozess thematisiert, die auch unter *Anerkennung* der in Abschn. 2.1 thematisierten Glaubenssätze Probleme bergen und entsprechend (unter anderem) aus Richtung der qualitativen Forschung kritisiert werden können. In Ergänzung zu dieser in erster Linie internen Kritik, welche die zugrunde liegenden Glaubenssätze unberührt lässt, wird an späterer Stelle (Kap. 7) zusätzlich externe Kritik vorgestellt werden, welche sich genau auf diese Glaubenssätze oder Axiome bezieht. Daneben sind in Abschn. 2.2 solche Punkte des Forschungsprozesses dargestellt, welche sich fundamental vom qualitativ orientierten Vorgehen unterscheiden, wie später noch zu zeigen sein wird.

Zunächst zum Ablauf quantitativ-empirischer Forschung: Dieser wird üblicherweise in Entdeckungs-, Begründungs- und Verwertungszusammenhang unterteilt. Der „eigentliche

Forschungsprozess“ spielt sich dabei im Begründungszusammenhang ab (vgl. z. B. Krebs und Menold 2014, S. 425), für den das Postulat der Wertfreiheit gilt: „Die Beschreibung und Erklärung von Tatsachen soll objektiv, d. h. wertfrei und damit für jedermann, der über das Fachwissen der Disziplin verfügt, nachvollziehbar sein. Wissenschaftliche Aussagen über die Realität dürfen nicht durch die Wunschvorstellungen des Wissenschaftlers beeinflusst sein“ (Schnell et al. 2013, S. 88). Für den Entdeckungs- und den Verwertungszusammenhang gilt das Postulat der Wertfreiheit hingegen nicht. Dieses Bild deckt sich gut mit dem bereits angesprochenen Gedanken einer „Einheitswissenschaft“ (vgl. Kromrey 2009, S. 17): Der Forscher hat natürlich Motive für seine Forschungstätigkeit und vielleicht auch Vorstellungen darüber, wie seine Forschungsergebnisse zu verwerten seien, ansonsten jedoch versucht er, als „externer Beobachter“ möglichst objektiv die für seinen Forschungsgegenstand relevanten Gesetze zu erforschen. Kromrey (2009, S. 21) spricht von der „... nach empirischen Gesetzmäßigkeiten fahndenden Forscherin und ihren Kollegen“ (vgl. hierzu auch Schnell et al. 2013, S. 53). Die Vorstellungen eines „externen Beobachters“ sowie der „Objektivität“ werden in den Kap. 3 und 5 wieder aufgegriffen.

Wie sehen die im Rahmen des Begründungszusammenhangs zu prüfenden Hypothesen aus? In aller Regel handelt es sich um *Kausalhypothesen* (vgl. z. B. Kühnel und Dingelstedt 2014, S. 1017). Die Forschung geht „... prinzipiell von der Determiniertheit aller natur- und sozialwissenschaftlichen Phänomene [aus]“ (Schnell et al. 2013, S. 55; vgl. dort auch S. 63). Baumgartner und Graßhoff zitieren in diesem Zusammenhang Max Planck mit den Worten: „...

das wissenschaftliche Denken verlangt nun einmal nach Kausalität, insofern ist wissenschaftliches Denken gleichbedeutend mit kausalem Denken, und das letzte Ziel einer jeden Wissenschaft besteht in der vollständigen Durchführung der kausalen Betrachtungsweise“ (Max Planck 1934, zit. nach Baumgartner und Graßhoff 2004, S. 13). Sie fahren fort:

... Max Planck vertritt in diesem Zitat die weit reichende These der Identität von wissenschaftlichem und kausalem Denken. Wie er hat etwa auch Albert Einstein die herausragende Bedeutung kausaler Fragestellungen für wissenschaftliche Tätigkeit stets betont. Doch selbst Wissenschaftler, die Plancks Überzeugung von der Identität wissenschaftlichen und kausalen Denkens nicht teilten (oder teilen), räumten (und räumen) kausalen Analysen eine zentrale Stellung im Rahmen wissenschaftlicher Untersuchungen ein ... (Baumgartner und Graßhoff 2004, S. 13).

Fazit: Kausalhypothesen stehen im Mittelpunkt der quantitativen empirischen Sozialforschung (vgl. z. B. auch Gangl und DiPrete 2006, S. 396) – ganz im Gegensatz zur qualitativen Forschung, wie zu zeigen sein wird.

Ein Problem bei der Formulierung von Hypothesen und der Umsetzung entsprechender Forschung im humanwissenschaftlichen Bereich besteht darin, dass „Ursache“ und „Wirkung“ zu Forschungszwecken oft schwer zu isolieren sind, da es sich über weite Strecken um rückgekoppelte Prozesse handelt und nicht um eine punktuell auftretende „Ursache“, welche eine „Wirkung“ nach sich zieht. Letzteres wäre zum Beispiel eine notwendige (wenn auch nicht hinreichende) Bedingung für eine experimentelle Prüfung

– der potentesten Möglichkeit, Kausalzusammenhänge zu untersuchen. Wie aber soll etwa geprüft werden, ob „politisches Interesse“ *kausal* zu einer erhöhten Neigung, sich an Wahlen zu beteiligen, führt?

Ein weiterer Aspekt sei, auch als Ausblick auf die Ausführungen zur qualitativen empirischen Sozialforschung, anhand eines Zitates von Philip Zimbardo (2008, S. 228) angesprochen: „In der Psychologie sind diejenigen, die am stärksten in der humanistischen Tradition verwurzelt sind, die nachdrücklichsten Verfechter des Grundsatzes, dass menschliche Würde Vorrang haben muss vor den erklärten Zielen der Disziplin, nämlich menschliches Verhalten vorherzusagen und zu beherrschen“. An dieser Stelle ist lediglich der letzte Teil des Zitates relevant. Die Erforschung der Gesetze im jeweiligen Gebiet (etwa der Rechtsextremismusforschung) wird in der Regel kaum aus „reiner Neugier“ erfolgen, sondern mit der Absicht, Eingriffsmöglichkeiten und die damit verbundenen Effekte zu untersuchen (vgl. auch z. B. Rogers 2016, S. 372, 376–377). Michael Häder (2010, S. 68) bringt es auf den Punkt: „Die Aufgabe der Wissenschaft ist es, diese Prinzipien und Regeln aufzudecken und damit letztlich die Welt beherrschbar zu machen“. So besagt etwa bei einer multiplen Regression ein (unstandardisierter), kausal interpretierter Regressionskoeffizient, welche Veränderung für die abhängige Variable bei Erhöhung der Ausprägung der betreffenden unabhängigen Variablen um eine Einheit zu erwarten ist. Carl Rogers (Näheres in Abschn. 6.1) spitzt das Argument zu, wenn er schreibt: „Wissenschaft führt in ihrer letzten Konsequenz zur Manipulation“ (Rogers 2016, S. 211–212; vgl. hierzu auch Rogers 2016, S. 203, 358, 373 und Rogers 2014, S. 119).

Die „Eingriffsmöglichkeiten“ dürften dabei im Wesentlichen darin bestehen, das vom Menschen wahrgenommene (und großenteils medial vermittelte) Bild „der Realität“ zu beeinflussen (vgl. hierzu auch Kap. 5). Die Spannweite der Charakterisierung dieser Beeinflussung mag dabei von „Manipulation“ im Sinne von „Geraderücken einer – wie auch immer definiert – verzerrten Wahrnehmung“ auf der einen Seite bis hin zu „Manipulation“ im Sinne von „Propaganda“ reichen. Veränderungen im Bereich persönlicher Eigenschaften im engeren Sinne dürften dagegen schwer zu erzielen sein, wie in dem Buch von Gerhard Roth: „Persönlichkeit, Entscheidung und Verhalten. Warum es so schwierig ist, sich und andere zu ändern“ (Roth 2013a) ausführlich dargelegt ist.

Die nun folgenden Abschnitte beschäftigen sich mit der Operationalisierung und damit der Messung im weitesten Sinne – einschließlich der Konzeptspezifikation: Zunächst zur Definition des Begriffs „Messung“: Man versteht darunter die strukturtreue Abbildung eines empirischen Relativs in ein (meist) numerisches Relativ (vgl. z. B. Schnell et al. 2013, S. 128–129). Vorausgesetzt wird dabei – nach Abschn. 2.1 als „Glaubenssatz“ – ein „empirisches Relativ“, d. h. die Existenz einer „objektiv existierenden Wirklichkeit“ mit bestimmten Eigenschaften, die wir obendrein bereits *vorab* als „bekannt“ betrachten müssen (um zu wissen, was unter „strukturtreu“ zu verstehen sei).

Letzteres ist keineswegs trivial, wie ein anschauliches Beispiel aus den Naturwissenschaften zeigt: Für die Größe „Temperatur“ veränderte sich mit der Entdeckung des absoluten Nullpunkts die Vorstellung vom empirischen Relativ. Statt Intervall- wird nun Ratioskalenniveau angenommen,

was zur Folge hat, dass strukturtreue Messungen in Kelvin durchgeführt werden und Messungen in „Grad Celsius“ oder „Grad Fahrenheit“ nicht mehr als strukturtreu zu betrachten sind (vgl. z. B. Schumann 2012, S. 24).

Im „empirischen Relativ“ finden sich „Merkmalsträger“ – meist Menschen – mit Merkmalen in unterschiedlicher Ausprägung. Via Messung soll die Ausprägung derjenigen Merkmale, die für den Forschungsprozess relevant sind, strukturtreu – meist durch die Zuordnung von Zahlenwerten – erfasst werden. „Strukturtreu“ bedeutet dabei, dass sich die im empirischen Relativ zu untersuchenden Relationen auch im numerischen Relativ, also den Zahlenwerten (oder sonstigen Messwerten) finden.

Das klingt zunächst trivial, entpuppt sich jedoch auf den zweiten Blick in mehrerlei Hinsicht als ausgesprochen komplex und problembehaftet, was sich bereits bei der einfachsten Form der Messung, nämlich der Beobachtung, zeigt. Nehmen wir eine der „Königsdisziplinen“ der politikwissenschaftlichen Forschung, die quantitativ-empirische Wahlforschung: Sie ist in aller Regel nicht in der Lage, ihre zentrale abhängige Variable – das Wählerverhalten – direkt zu messen. Beobachtungen in der Wahlkabine durchzuführen, ist nun einmal nicht möglich. Ersatzweise wird im Rahmen der Umfrageforschung mit Proxies gearbeitet – wie der (hypothetischen) Sonntagsfrage: „Wenn nächsten Sonntag Bundestagswahl wäre ...“, der Frage nach der Wahlintention (Wie haben Sie vor, sich zu entscheiden ...?) oder einer Recallfrage (Wie war das bei der letzten Bundestagswahl ...?). Die so ermittelten Messwerte beruhen auf *Selbstausskunft* der befragten Personen, womit



man sich vom eingangs erwähnten Postulat der Einheitswissenschaft deutlich entfernt. Kein Physiker fragt eine Kugel, wie schnell sie sich bewege, bewegen würde, bewegen werde oder sich bewegt habe.

In den Humanwissenschaften werden generell nur in seltenen Fällen Messungen im Sinne einer „externen Beobachtung“ durchgeführt – sei es aufgrund restriktiver Rahmenbedingungen (wie dem Wahlgeheimnis oder eines in Grenzen zu haltenden Aufwands) oder aus prinzipiellen Gründen, wie der Tatsache, dass sehr oft nicht unmittelbar beobachtbare „hypothetische Konstrukte“ zu untersuchen und die Ausprägung entsprechender latenter Variablen zu messen sind. Beispiele für eine Vielzahl solcher latenter Variablen aus dem Persönlichkeitsbereich finden sich in Schumann und Schoen (2005). Deren hypothetischen Charakter unterstreicht ein Zitat von Philip Zimbardo (2008, S. 7): „Obwohl Sie vermutlich von sich selbst glauben, Sie hätten eine über Zeit und Raum gleich bleibende Persönlichkeit, trifft dies wahrscheinlich nicht zu. Sie sind nicht dieselbe Person, wenn Sie allein arbeiten oder in einer Gruppe; bei einer romantischen Verabredung oder im Hörsaal; wenn Sie mit guten Freunden zusammen sind oder in einer anonymen Menschenmenge; oder wenn Sie auf Reisen sind statt im vertrauten Zuhause“. Beispiele für latente Variablen aus anderen Forschungsbereichen wären politische Einstellungen, Werthaltungen oder allgemeine Eigenschaften einer Person (wie deren „politisches Interesse“).

Die hypothetischen Konstrukte sind sehr oft über Situations-Reaktions-Zusammenhänge definiert, welche prinzipiell beobachtbar wären. Aufgrund des Aufwands, der damit verbunden wäre, muss jedoch auch hier meist

auf eine Selbstauskunft ausgewichen werden. Da sich die Hypothesen, die auf Grundlage der Messergebnisse geprüft werden sollen, normalerweise auf die hypothetischen Konstrukte selbst beziehen (z. B. auf die Ausprägung des „politischen Interesses“) und nicht auf entsprechende Selbsteinschätzungen (wie die Antwort auf die Frage: „Wie stark interessieren Sie sich für Politik ...“), liegt hier ein Problem. Es ist vorstellbar, dass eine Person, die – nach der jeweiligen Definition – als „politisch desinteressiert“ einzustufen ist, sich selbst als „politisch interessiert“ sieht und – in ihren Augen völlig wahrheitsgemäß – entsprechend antwortet. Selbsteinschätzungen können von den Wahrnehmungen „externer“ Beobachter durchaus erheblich abweichen (vgl. z. B. Diekmann 2009, S. 575 oder Stemmler et al. 2011, S. 111–118). Von einer „struktur-treuen Abbildung“ zu sprechen, erscheint im genannten Beispiel nicht unbedingt einleuchtend. Da Messung im Rahmen der Umfrageforschung in aller Regel auf Selbstauskünften basiert, stellt dieser Punkt ein Problem für die Umfrageforschung allgemein dar – es sei denn, man ist aus theoretischen Gründen explizit an Selbstauskünften interessiert. Nach dem „Thomas-Theorem“ beispielsweise sind *subjektive* Wahrnehmungen und Einschätzungen handlungsrelevant und nicht die Gegebenheiten einer objektiv existierenden (und als solche feststellbaren) Wirklichkeit. Allerdings passt auch dies nicht zum Postulat der Einheitswissenschaft. Eine Kugel entwickelt keine subjektive Wahrnehmung über ihr Gewicht oder ihre Geschwindigkeit.

Eine zusätzliche Schwierigkeit beim Erfragen von Selbsteinschätzungen zur Messung der Ausprägung „tatsächlich“ (im Sinne einer objektiv existierenden Wirklichkeit)

gegebener Merkmale besteht im Streben des Menschen nach „kognitiver Konsonanz“, mit dem sich – aufbauend auf der Arbeit von Leon Festinger (1957; deutsch 2012) – ein ganzer Forschungszweig der Sozialpsychologie beschäftigt (vgl. hierzu z. B. Schumann 2001, S. 71–78). Menschen streben offenbar danach, dass (nach ihrem Empfinden) möglichst wenig „Spannungen“ zwischen kognitiven Elementen bestehen. Zu den „kognitiven Elementen“ zählen unter anderem die von ihnen vertretenen Einstellungen, ihre Werthaltungen, ihr (subjektives) Wissen über ihre Verhaltensweisen und ihre Grundüberzeugungen über die Beschaffenheit der sie umgebenden Welt. Zum Abbau solcher Spannungen oder „kognitiven Dissonanzen“ stehen eine Reihe von Möglichkeiten zur Verfügung – unter anderem an zentraler Stelle die „Anpassung“ von kognitiven Elementen (Festinger 2012, S. 30–31, 42). Genau nach diesen wird jedoch bei Messungen über Selbstberichte gefragt. Dies könnte beispielsweise folgendes Resümee von Haddock und Maio erklären, nach dem – im Vergleich zu anderen Forschungsbereichen – in der empirischen Wahlforschung erstaunlich starke Zusammenhänge zwischen Einstellungen und Verhalten zu verzeichnen sind: „Ein *Extrembereich* sind die Einstellungen gegenüber einer politischen Partei; hier ist der Zusammenhang zwischen Einstellungen und Verhalten gewöhnlich sehr hoch“ (Haddock et al. 2007, S. 216; Hervorhebung: Sch.). Wie bereits erwähnt, wird in der empirischen Wahlforschung neben der Einstellung auch das (Wahl-)Verhalten über Selbstauskünfte gemessen, während andere Bereiche der Sozialpsychologie eher auf die Beobachtung des Verhaltens setzen. Die Gefahr der systematischen Überschätzung von

Zusammenhängen ist unter diesem Gesichtspunkt nicht von der Hand zu weisen, worauf z. B. auch Fishbein und Ajzen (2010, S. 48–49) – gestützt auf Metastudien-Ergebnisse – hinweisen.

Abgesehen vom oben Gesagten ist mit der Messung von hypothetischen Konstrukten noch ein weiteres Problem verbunden – und zwar insbesondere dann, wenn ein hypothetisches Konstrukt als „unabhängige Variable“ zur Erklärung menschlichen Verhaltens betrachtet wird. Wie gesagt, werden hypothetische Konstrukte sehr oft über Situations-Reaktions-Zusammenhänge definiert, womit man sich auf einer *deskriptiven* Ebene bewegt. Es wird lediglich beschrieben, dass Menschen sich in bestimmten Situationen auf bestimmte Art und Weise verhalten. Eine derartige Deskription kann jedoch nichts zusätzlich erklären. Sowohl in der Sozialpsychologie als auch in der Persönlichkeitspsychologie wird dieser Gedanke vertreten (vgl. z. B. Stemmler et al. 2011, S. 284, Pervin 2005, S. 48 oder Schumann 2012a, S. 44, 78, 273). Um bei der Prüfung von Kausalhypothesen als „Erklärungsvariable“ dienen zu können, muss eine irgendwie geartete „biophysische Existenz“ der Merkmale bzw. ihrer Ausprägungen gegeben sein. In den Humanwissenschaften wird diese jedoch meist nur *unterstellt*. Hans-Jürgen Eysenck steht für eine der wenigen Ausnahmen. Er hat beispielsweise für das Persönlichkeitsmerkmal „Extraversion“ eine solche biophysische Basis beschrieben (vgl. hierzu Stemmler et al. 2011, S. 284–292). Die entsprechenden Forschungen charakterisieren Stemmler et al. (2011, S. 294–295) wie folgt: „... die Theorie zur Extraversion/Introversion hat sich ... als außerordentlich fruchtbar erwiesen und Anlass zu einer

Fülle von Untersuchungen gegeben, die von hochspeziellen experimentellen Aufgaben über Kurz- und Langzeitgedächtnis zu sozialen und politischen Einstellungen bis hin zu pharmakologischer und psychotherapeutischer Reagibilität reichen ...“. Es liegt auf der Hand, dass entsprechende Hypothesen mit einem lediglich beschreibenden Extraversions-Begriff nicht aufgestellt werden können.

Ein Wort noch zu „Methodeneffekten“: Es existiert eine Fülle von Forschungsergebnissen zum Thema „Messung“, welche bei einer „strukturtreuen Abbildung eines empirischen Relativs in ein numerisches Relativ“ nicht auftreten dürften. Hierzu gehören der Halo-Effekt (Fragen „strahlen auf nachfolgende Fragen aus“), ganz allgemein Positions- und Reihenfolgeeffekte (z. B. Primacy und Recency-Effekte), Interviewereffekte (Geschlecht, Alter etc. der Interviewer beeinflussen die Antworten), Effekte der Interviewsituation (z. B. Anwesenheit Dritter), Verzerrungen aufgrund „sozialer Erwünschtheit“ oder aufgrund einer „Ja-Sage-Tendenz“ (Akquieszenz) und vieles mehr (vgl. z. B. Schnell et al. 2013, S. 345–348). Selbst die Ausstattung der Räumlichkeiten, in denen eine Befragung stattfindet, kann die Messwerte beeinflussen – etwa dann, wenn nach der Zufriedenheit mit der persönlichen Wohnsituation gefragt wird (vgl. Diekmann 2009, S. 468). Ferner kann die Formulierung einer Frage wie der Antwortvorgaben die Messergebnisse beeinflussen – bis hin zu dem Effekt, dass sich für unterschiedliche Formulierungen selbst bei *logischer Äquivalenz* die Messergebnisse unterscheiden (vgl. hierzu z. B. Schumann 2012a, S. 207–208). Dies gilt selbst für vergleichsweise leicht und präzise zu beobachtende Größen, wie etwa die Dauer des täglichen

Fernsehkonsums. Schwarz et al. (1988, S. 22) konnten zum Beispiel zeigen, dass 16,2 % ihrer Befragten einen täglichen Fernsehkonsum von mehr als zweieinhalb Stunden berichteten – oder 37,5 %, je nachdem, wie die vorgegebenen Kategorien zur Beantwortung der entsprechenden Frage aussahen (vgl. hierzu auch Schwarz 1995, S. 159).

All dies spricht dafür, die Messung in Umfragen als „Aufgabe“ aufzufassen, vor welche die betreffenden Personen gestellt werden und die sie im Falle der Kooperation zu lösen versuchen. Tourangeau et al. (2007, S. 7–16) haben beispielsweise die kognitiven Prozesse, die hierbei ablaufen können, beschrieben (vgl. hierzu auch z. B. Groves et al. 2009, S. 217–257). Mit dieser Auffassung entfernt man sich allerdings sehr weit vom Postulat der Einheitswissenschaft. Keine Physikerin stellt eine Kugel vor die Aufgabe, herauszufinden, wie ihre „Anfrage“ aufzufassen sei und anschließend (im Falle der korrekten Interpretation der Frage) zu berichten, wie schnell sie sich bewege – oder (im Falle einer „Fehlinterpretation“) irgendetwas anderes.

Zwei weitere Methodeneffekte (im weitesten Sinne) stellen die Veränderung des Messgegenstandes durch die Messung sowie die „Konstruktion“ von Antworten dar. Die Veränderung des Messgegenstandes durch die Messung ist auch aus der klassischen Physik bekannt. Insofern „passt“ hier die Vorstellung einer Einheitswissenschaft. Wird etwa die Geschwindigkeit eines Fahrrads (auf altmodische Weise) über ein Messinstrument gemessen, das über ein Rädchen mit dem Vorderrad verbunden ist, so bremst dieses Rädchen bei der Messung das Vorderrad, sodass die Geschwindigkeit des Gesamtsystems gemessen

wird und nicht die Geschwindigkeit des Fahrrads „an sich“, ohne dass ein Messvorgang stattfände. Moderne Messinstrumente können diese Diskrepanz minimieren, jedoch prinzipiell nicht aufheben. Spätestens wenn man die Messung aus Sicht der befragten Person als „Lösen einer Aufgabe“ betrachtet, liegt das Problem der Veränderung des Messgegenstandes durch die Messung auch für die empirisch-quantitative Umfrageforschung auf der Hand. Um mit Hans-Peter Dürr zu sprechen: „Wir können nicht eine Versuchsanordnung einrichten und der Testperson sagen: ‚Du wiederholst jetzt tausend Mal exakt dasselbe!‘ Das geht nicht, weil die Testperson bei jedem Versuch eine andere ist!“ (Dürr 2012, S. 64). Natürlich ist die Person nach dem Lösen der Aufgabe (und *vor* der Abgabe der Antwort!) nicht mehr dieselbe wie vorher. Dieses Problem mag in bestimmten Fällen zu vernachlässigen sein. Eine Politikwissenschaftlerin dürfte zu den momentan amtierenden Spitzenpolitikern in aller Regel eine „fertige“ Meinung haben, welche bei der Frage: „Was halten Sie ganz allgemein von ...“ im Wesentlichen nur noch „abgerufen“ werden muss. Insofern hält sich in diesem Fall die „Aufgabe“ – und damit der Einfluss ihrer Lösung auf das Messergebnis – wohl in engen Grenzen. Wird dagegen eine Durchschnittsbürgerin gefragt: „Was halten Sie ganz allgemein vom Bundesverfassungsgericht“, dann dürfte für sie im Normalfall die Lösung der Aufgabe wesentlich anspruchsvoller sein und die damit verbundenen Effekte größer. Empirische Forschungsergebnisse deuten in diese Richtung (vgl. z. B. Abold et al. 2009, S. 149).

Eine weitere, in der Forschungspraxis auftretende Herausforderung für die Vorstellung von „Messung“ als „strukturetreuer Übertragung eines empirischen in ein numerisches Relativ“ stellen offensichtlich konstruierte Antworten dar. Studien haben beispielsweise gezeigt, dass auf die Frage: „Was halten Sie ganz allgemein von ...“ auch bei Vorgabe *nicht existierender* Politiker diese von einem nennenswerten Teil der Bevölkerung eingeschätzt werden. Selbst wenn explizit danach gefragt wird, ob solche „Phantom-Politiker“ bekannt sind, treten entsprechende Effekte (allerdings in verringerter Größenordnung) auf. Hierzu ein Zitat aus Diekmann (2009, S. 451):

Dieter Köstritz, Staatssekretär in der sächsischen Staatsregierung, hatte auf Anhieb Erfolg. Bei einer Meinungsumfrage im Auftrag der Regierung kannten ihn 8, unter denen mit höherem Bildungsgrad gar 18 Prozent der Befragten. Nur: Den Mann gibt es gar nicht. ‚Wir wollten einfach mal wissen‘, begründete die stellvertretende Regierungssprecherin Heidrun Müller das listig lancierte Umfragephantom, ‚wie verlässlich Meinungsumfragen sind‘ (Aus: «Der Spiegel» 36, 1993).

So viel zur „Messung“. Angemerkt sei lediglich noch, dass Messung nicht theorielos erfolgen kann. Dies zeigt schon die Fülle sehr unterschiedlicher Ansätze und Methoden, die zur Verfügung stehen. Das Likert-Verfahren beispielsweise lässt sich rekonstruieren im Lichte der klassischen Testtheorie (vgl. Schumann 2012, S. 28–45), die Guttman-Skalierung basiert je nach Sichtweise auf einer deterministischen (vgl. Schnell et al. 2013, S. 181) oder



probabilistischen (vgl. Bortz und Döring 1995, S. 191–192)<sup>5</sup> Testtheorie, die Rasch-Skalierung auf einer probabilistischen Testtheorie, die Magnitude-Skalierung baut auf experimentellen Erkenntnissen der Psychophysik auf, die Thurstone-Skalierung macht sich Expertenwissen zunutze usw. (vgl. zu den Verfahren z. B. Schnell et al. 2013, S. 169–196).

Nach erfolgter Messung ist die Güte der Messung bzw. des Messinstruments zu untersuchen – insbesondere die Reliabilität und die Validität. Die Reliabilität beschäftigt sich mit der *Genauigkeit* der Messung – ungeachtet des erfassten Inhalts. Nach der klassischen Testtheorie kann die Reliabilität definiert werden als das Quadrat der Korrelation zwischen Messwert und „wahrem Wert“ (womit sich deren gemeinsame Varianz ergibt). Diese elegante Definition kann empirisch nutzbar gemacht werden, da die genannte Größe aus empirischen Daten geschätzt werden kann, nämlich aus der Korrelation zwischen den Messwerten zweier paralleler Tests oder anderer Verfahren, die auf diesem Grundgedanken aufbauen (vgl. Schumann 2012, S. 29–42). Das Problem, dass ein „wahrer Wert“ empirisch nicht zu ermitteln ist, ist damit weitgehend „umschifft“. Angemerkt sei allerdings, dass die Existenz eines „wahren Wertes“ (genauer: eines „wahren Messwerts“) *vorausgesetzt* wird – womit die Argumentation auch in diesem Fall auf einem Glaubenssatz beruht (vgl. Abschn. 2.1 und Kap. 5).

---

<sup>5</sup>In der 5. Auflage des Buches von 2016 werden für die Guttman-Skala (auf S. 272–274) keine entsprechenden Angaben mehr gemacht.

Das Kriterium der Validität beschäftigt sich mit dem im Rahmen der Reliabilität „ausgeklammerten“ *Inhalt* der Messung; genauer gesagt damit, ob gemessen wird, was gemessen werden soll. Eine Möglichkeit, die Validität eines neu entwickelten Messinstruments zu untersuchen, besteht darin, ein bereits bekanntes und bewährtes Instrument als „Außenkriterium“ heranzuziehen und die Korrelation dessen mit dem neu entwickelten Instrument zu betrachten. An dieser Stelle übt Philipp Mayring Kritik. Für ihn wird hier offensichtlich, „... dass sich die Katze in den Schwanz beißt. Eine Untersuchung, ein Messinstrument will besser sein als Vorhergehendes, will exakt ihren Gegenstand treffen und nicht ähnliche Gegenstände, Korrelate. Also *muss* sie von alternativen Messinstrumenten, von Korrelaten abweichen, wenn sie gut sein will, und nicht umgekehrt“ (Mayring 2016, S. 141; Hervorhebung im Original). Dieser Kritikpunkt erscheint berechtigt, sofern das eingesetzte Messverfahren prinzipiell unverändert bleibt und die Entwicklung eines Instruments, das unter Validitätsgesichtspunkten „besser“ ist als ein bereits vorhandenes, beabsichtigt ist. Neue Instrumente können jedoch auch aus anderen Gründen entwickelt werden, etwa um Parallelversionen herzustellen, um gleiche Inhalte mit neuen Messmethoden zu erfassen oder um die Konstruktvalidität zu verbessern. Im letzteren Fall dienen als „Validierungskriterium“ die möglichst theoriekonformen Zusammenhänge der mit dem neuen Instrument erzielten Messwerte in einem Netzwerk anderer Variablen.

Als drittes zentrales Gütekriterium eines Messinstruments (oder: Tests) wird meist die Objektivität (in anderer

Bedeutung als bisher<sup>6</sup>; vgl. hierzu auch Kap. 5) genannt. Lienert und Raatz schreiben: „Unter *Objektivität* eines Tests verstehen wir den Grad, in dem die Ergebnisse eines Tests unabhängig vom Untersucher sind. Ein Test wäre demnach vollkommen objektiv, wenn verschiedene Untersucher bei denselben Pbn [Probanden; Sch.] zu gleichen Ergebnissen gelangen. Man spricht deshalb auch von ‚interpersoneller Übereinstimmung‘ der Untersucher“ (Lienert und Raatz 1994, S. 7; Hervorhebung im Original). Die Untersucher müssen also austauschbar sein und können bzw. dürfen damit nicht mit dem Untersuchungsgegenstand interagieren – zumindest nicht auf eine für sie charakteristische Art und Weise. Die damit verbundene Vorstellung eines „externen Beobachters“, der – ansonsten unbeteiligt – lediglich ein Instrument (oder mehrere) einsetzt, wird in den Kap. 3 und 5 wieder aufgegriffen.

Nun noch einige Anmerkungen zum Prozedere der Datenerhebung. Die untersuchten Personen werden dabei ausschließlich als „Merkmalsträger“ und „Datenlieferanten“ betrachtet. Werden sie mit einer Person (z. B. einer Interviewerin) konfrontiert, soll deren alleinige Funktion darin bestehen, ein Messinstrument (oder mehrere) anzuwenden und auf diese Weise die Daten zu erheben, welche anschließend der Forschung zur Verfügung gestellt

---

<sup>6</sup>Zur Unterscheidung schreiben Bunge und Mahner (2004, S. 121–122): „‚Objektiv‘ im ontologischen Sinne heißt so viel wie ‚außerhalb und unabhängig von unserem Denken existierend‘. ‚Objektiv‘ im erkenntnistheoretischen Sinne bedeutet indes, dass die Wahrheit oder Gültigkeit einer Erkenntnis nicht vom erkennenden Subjekt, nicht von der Person abhängt“. In letzterem Sinne wird „Objektivität“ hier gebraucht.

werden. Dies gilt sowohl für die Umfrageforschung als auch für das Vorgehen im Bereich der quantitativ-empirischen Sozialforschung allgemein. Dass diese Idealvorstellung nur in möglichst guter Annäherung erreicht werden kann, versteht sich von selbst. Eine Sozialforscherin (bzw. Interviewerin) *muss* in soziale Interaktion mit der befragten Person treten, wenn sie ein Interview durchführt. Sie kann sich lediglich bemühen, diese Interaktion auf ein Minimum zu beschränken; dass dennoch entsprechende Interviewer-Effekte nachweislich auftreten, wurde bereits erwähnt. Eine Physikerin dagegen tritt nicht in soziale Interaktion mit einer Kugel, wenn sie deren Masse oder Geschwindigkeit misst.

Die bei der Untersuchung eingesetzten Instrumente sind in aller Regel so gestaltet, dass mit den gewonnenen Daten bereits *fertig vorliegende* Hypothesen geprüft, modifiziert oder – im Falle ihrer „Nicht-Bewährung“ – verworfen werden können. Die eingesetzten Instrumente dürfen (und können) im Rahmen der Datenerhebung nicht verändert werden. In diesem Punkt unterscheiden sich die qualitative und die quantitative Vorgehensweise fundamental, wie noch zu zeigen sein wird (vgl. hierzu z. B. auch Brüsemeister 2008, S. 44).

Ein Problem, das insbesondere die Umfrageforschung betrifft, hängt mit der Stichprobenziehung zusammen. Möchte man Aussagen über alle Wahlberechtigten der Bundesrepublik machen, zieht man aus dieser Grundgesamtheit eine Zufallsstichprobe. Sind die entsprechenden Daten erhoben, können mittels inferenzstatistischer Verfahren Aussagen über die Grundgesamtheit abgeleitet werden. So können etwa Verteilungen betrachtet und Konfidenzintervalle berechnet oder Zusammenhangsmaße ermittelt und hierfür

Signifikanztests durchgeführt werden. Soweit die Theorie. Der Haken: Die inferenzstatistische Argumentation geht nicht nur von einer Zufallsstichprobe aus, sondern auch von einer Ausschöpfungsquote von 100 % (vgl. z. B. Brick und Williams 2013, S. 36). Keine Kugel in einer Urne weigert sich, per Zufall gezogen (und anschließend analysiert) zu werden. Zufällig für eine Untersuchung ausgewählte Personen weigern sich hingegen sehr wohl in nennenswertem Umfang, an dieser teilzunehmen.

Die Ausschöpfungsquoten für die methodisch sehr sorgfältig durchgeführten bundesweiten ALLBUS-Bevölkerungsumfragen liegen derzeit um die 30 bis 40 % – 2012 bei knapp 38 % (vgl. Wasmer et al. 2014, S. 58) und 2014 nach Gesis-Angabe bei 35 % (vgl. hierzu auch Blohm und Koch 2015, S. 87 und Proner 2011, S. 39) – also weit entfernt von den eigentlich benötigten 100 Prozent. Dieser Sachverhalt wird im Vorwort zu dem 2015 erschienenen Buch „Nonresponse-Bias“ (Schupp und Wolf 2015, S. 7) allgemein bestätigt: „Anspruchsvolle Umfragen mit genauer Überwachung des Feldes sowie kontrollierten Bruttostichproben erreichen derzeit selten eine höhere Ausschöpfung als etwa 35 %“ ist dort zu lesen (vgl. zur Größenordnung z. B. auch Pötter 2006, S. 172).

Dieses Problem wird in der angewandten Forschung selten thematisiert. Beispielsweise findet sich auf den 1090 Textseiten des „Handbuchs der sozialwissenschaftlichen Datenanalyse“ (Wolf und Best 2010) unter dem Suchbegriff „Ausschöpfung“ nur ein einziger Eintrag auf Seite 50 des Inhalts: „Im Zusammenhang mit Telefonumfragen werden verschiedene Ergebnisse aus Methodenstudien berichtet. Demnach sind auch in Telefonbefragungen komplexe Fragestellungen möglich, jedoch sollten Telefonbefragungen

eine dreiviertel Stunde in Bevölkerungsumfragen nicht überschreiten und es sind hier ähnlich hohe *Ausschöpfungsquoten* zu erreichen wie mit persönlichen Interviews (Diekmann 2007, S. 430 f.)“ (Hervorhebung: Sch.). Der Suchbegriff „Rücklauf“ ergibt acht Treffer, wobei allerdings die Höhe der erreichbaren Rücklaufquote im Text nicht thematisiert wird. Lediglich in einer Fußnote auf S. 44 findet sich ein Verweis auf ein Buch zum Thema „Diskussion der Rücklaufquote als Qualitätsmerkmal“. Generell gilt: Wer in Veröffentlichungen jedweder Art zu Umfrageuntersuchungen neben den Forschungsergebnissen auch eine Angabe zur Ausschöpfungsquote sucht, der wird in vielen Fällen feststellen, dass nur die schiere Anzahl der untersuchten Fälle berichtet ist, die eben *keinen* Schluss auf die Ausschöpfungsquote zulässt.

Das Problem wäre relativ gering, wenn die „ausgefallenen“ etwa 60 bis 70 % der eigentlich für die Befragung ausgewählten Personen sich nicht systematisch von den Teilnehmern an der Untersuchung unterscheiden würden. In diesem Fall wären keine größeren Verzerrungen der Ergebnisse zu erwarten. Leider zeigen empirische Untersuchungen sowohl bei der Betrachtung der Verteilung sozialstruktureller Merkmale im Vergleich zu den Angaben amtlicher Statistiken als auch aufgrund von „Nachfassaktionen“ (bei denen Personen, welche die Teilnahme an einer Befragung „eigentlich“ verweigern, doch noch einige Informationen entlockt werden oder sie sogar – etwa über Incentives – doch noch zur Teilnahme bewogen werden), dass man hiervon nicht unbedingt ausgehen kann. In einer Studie von Weinhardt und Liebig (2015) anhand der deutschen Stichprobe des European Social Survey sind zum Beispiel Personen mit maximal

Hauptschulabschluss deutlich unterrepräsentiert (vgl. Weinhardt und Liebig 2015, S. 57–58; 75–76). Das gleiche Ergebnis zeigte sich in einer Non-Response-Studie zum ALLBUS 1986 (vgl. Erbslöh und Koch 1998, S. 38, 42) sowie zumindest tendenziell in einer Non-Response-Studie zum ALLBUS 2008 (vgl. Blohm und Koch 2015, S. 109). Ferner nehmen politisch weniger Interessierte offenbar seltener an der Umfrage teil als politisch Interessiertere; gleiches gilt für politisch Unzufriedene (Schupp und Wolf 2015, S. 76; vgl. auch z. B. Proner 2011, S. 287; Kaspar 2009, S. 95). Dies legt – erstens – nahe, dass Personen, die an Befragungen teilnehmen, sich durchaus zumindest in bestimmten Punkten systematisch von solchen unterscheiden (können), die eine Teilnahme verweigern.

Zweitens wird ein „weißer Fleck auf der Landkarte“ sichtbar. So schreibt etwa Hanna Proner am Ende ihrer Nachfassstudie zum ALLBUS 2008: „Über das letzte Drittel Zielpersonen, die auch durch den enormen beschriebenen Aufwand nicht zu einer Auskunft bewegt werden konnten, lässt sich auf der Basis der vorliegenden Daten keine Aussage treffen“ (Proner 2011, S. 289). Man könnte zwar annehmen, dass das verbleibende Drittel der eigentlich zu befragenden Personen, die man als „hartnäckige Verweigerer“ bezeichnen könnte, politisch noch desinteressierter bzw. unzufriedener sei als die Personen, die sich schließlich doch noch zur Teilnahme motivieren oder überreden ließen. Sicher ist dies jedoch keineswegs. Die Daten der Umfrage sagen – wie bereits betont – einfach nichts über diese Gruppe aus.<sup>7</sup> Folgendes Zitat bringt die Sache auf den Punkt:

---

<sup>7</sup>Zumindest nicht ohne Zusatzinformationen. Diese liegen jedoch nur in seltenen Fällen vor und dann meist für Größen, welche für die humanwissenschaftliche Forschung von eher untergeordneter Bedeutung sind.

Die Richtigkeit der hier gezogenen Schlussfolgerungen hängt selbstverständlich davon ab, dass sich die Richtung und die Stärke der Zusammenhänge zwischen Sozialstruktur und politischen Einstellungen bzw. Verhaltensweisen bei Nichtteilnehmern von Wahlumfragen nicht von den Beziehungen unterscheiden, die bei Teilnehmern solcher Umfragen zu beobachten sind. Da gesicherte Daten über die Gruppe der Nichtteilnehmer fehlen, kann hierfür jedoch nicht gebürgt werden. Dennoch wird in der Praxis genau von dieser Annahme ausgegangen ... (Maier und Schneider 2009, S. 326).

Die möglichen Konsequenzen hiervon lassen sich am Beispiel der Nichtwähler-Forschung demonstrieren. Der Anteil der Nichtwähler wird im Rahmen der Umfrageforschung regelmäßig unterschätzt, was traditionell als Effekt der sozialen Erwünschtheit gedeutet wurde. Es widerspricht einerseits (kulturell) der gängigen „Wahlnorm“ und macht andererseits (situational) dem Interviewer gegenüber „einen schlechten Eindruck“, wenn man sich zum Nichtwählen bekennt. Also lüge die befragte Person gegebenenfalls (bei der Recallfrage) oder sie „schöne“ die von ihr konstruierte Antwort (z. B. bei der – hypothetischen – Sonntagsfrage). Das mag in vielen Fällen zutreffen, es könnte jedoch auch sein, dass im Extremfall keine einzige der befragten Personen lügt oder ihre Antwort schönt – und der Nichtwähleranteil trotzdem chronisch unterschätzt wird. Voraussetzung hierfür wäre, dass die Personen, welche eine Teilnahme an der Umfrage verweigern, seltener zur Wahl gingen als die Teilnehmer – etwa, weil sie politisch desinteressierter wären als diese (vgl. hierzu z. B. Weinhardt und Liebig 2015, S. 63, 66–67). Sie wären dann unter den Befragten



unterrepräsentiert und somit der Anteil der „Wähler“ in der Stichprobe überrepräsentiert. Die genannte Annahme ist keineswegs unplausibel, was deutlich macht, wie problematisch es ist, aus den Ergebnissen von Umfragen insbesondere auf *Verteilungen* in der Grundgesamtheit zu schließen – etwa auf den Anteil der mit der Demokratie Zufriedenen unter den Wahlberechtigten (besser: derer, die eine entsprechend formulierte Frage in diesem Sinne beantworten) oder – unter der Verwendung von Gewichtungszprozeduren – auf das „tatsächliche Wahlverhalten“ der Wahlberechtigten, wie es sich im amtlichen Wahlergebnis niederschlägt. An diesem Punkt wird in der politikwissenschaftlichen Forschung meines Erachtens sehr häufig deutlich überinterpretiert. Hier könnte beispielsweise auch eine wichtige Ursache dafür liegen, dass immer wieder dramatische Fehlprognosen im Bereich der empirischen Wahlforschung zu verzeichnen sind.

Häufiger als auf Verteilungen richtete sich das wissenschaftliche Interesse auf (Kausal-)Zusammenhänge. Nach dem Falsifikationsprinzip wird entsprechenden Hypothesen im Rahmen der Forschung Gelegenheit dazu gegeben, sich zu bewähren oder „an der Realität zu scheitern“. Bewährt sich eine Hypothese, so wird sie vorläufig akzeptiert – auch wenn eine Verifikation grundsätzlich ausgeschlossen ist. Scheitert sie, wird sie verworfen – so zumindest die Grundargumentation (vgl. z. B. Schumann 2011, S. 10–13). In den Worten von Seipel und Rieker (2003, S. 74) „... wird die Suche nach inhaltlicher Erkenntnis und die Suche nach Wahrheit nicht aufgegeben. Aufgegeben wird die Idee, ... dass es absolute Gewissheit gibt“. Ralf Bohnsack (2014, S. 18) formuliert noch weitgehender: „Der Wissenschaftler *weiß* nichts, er

rät nur. Er kennt keine *wahren*, nur *bewährte* Aussagen. Indem er falsifizierte Hypothesen umformuliert, ihren Falschheitsgehalt eliminiert, sie also einschränkend neu formuliert, ... kommt er allmählich zu bewährten Aussagen und tastet sich in diesem Sinne allmählich an die Wahrheit heran“ (Hervorhebungen im Original).

Vorausgesetzt wird damit eine objektiv existierende Wirklichkeit (die angesprochene „Realität“), wobei allerdings die in ihr geltenden *Gesetze* nicht 1:1 abgebildet werden können. Was die untersuchten (Kausal-) Zusammenhänge betrifft, deckt sich damit in der Forschungspraxis die Sichtweise der quantitativ-empirischen Sozialforschung weitgehend mit derjenigen des in Kap. 5 zu thematisierenden radikalen Konstruktivismus. Paul Watzlawick liefert mit folgendem Bild meines Erachtens eine treffende Beschreibung des Falsifikationsprinzips:

Ein Kapitän, der in dunkler, stürmischer Nacht eine Meerenge durchsteuern muss, deren Beschaffenheit er nicht kennt, für die keine Seekarte besteht und die keine Leuchfeuer oder andere Navigationshilfen besitzt, wird entweder scheitern oder jenseits der Meerenge wohlbehalten das sichere, offene Meer wiedergewinnen. Rennt er auf die Klippen auf und verliert Schiff und Leben, so beweist sein Scheitern, dass der von ihm gewählte Kurs nicht der richtige Kurs durch die Enge war. Er hat sozusagen erfahren, wie die Durchfahrt nicht ist. Kommt er dagegen heil durch die Enge, so beweist dies nur, dass sein Kurs im buchstäblichen Sinne nirgends anstieß. Darüber hinaus aber lehrt ihn sein Erfolg nichts über die wahre Beschaffenheit der Meerenge; nichts darüber, wie sicher oder wie nahe an der

Katastrophe er in jedem Augenblicke war: er passierte die Enge wie ein Blinder. Sein Kurs passte in die ihm unbekannten Gegebenheiten; er stimmte deswegen aber nicht, wenn mit stimmen das gemeint ist, was von Glasersfeld darunter versteht: dass der gesteuerte Kurs der wirklichen Natur der Enge entspricht. Man kann sich leicht vorstellen, dass die wahre Beschaffenheit der Meerenge vielleicht wesentlich kürzere, sicherere Durchfahrten ermöglicht (Watzlawick 2016, S. 14–15/Einleitung).

Ein Problem bei der Anwendung des Falsifikationsprinzips, dessen elegante Grundidee sich leicht anhand einer beliebigen deterministischen Hypothese: „Immer wenn ..., dann ...“ demonstrieren lässt, besteht darin, dass in der quantitativ-empirischen Sozialforschung solche deterministischen Hypothesen praktisch nicht vorkommen. Selbst wenn sie aufgestellt würden, würde ein einziger nicht hypothesenkonformer Fall zur Falsifikation führen und sie damit „verschwinden“ lassen. Stattdessen sind die Hypothesen meist stochastisch formuliert in der Form: „Je ... desto ...“ – allerdings nicht deterministisch. Die Zusammenhänge lassen sich nun mit geeigneten Maßzahlen – zum Beispiel Korrelationskoeffizienten – beschreiben, allerdings steht kein klares Falsifikationskriterium mehr zur Verfügung. Die Festsetzung, ab welchem Betrag des Korrelationskoeffizienten eine Zusammenhangshypothese als „bewährt“ im Sinne von „bislang nicht falsifiziert“ einzustufen ist, basiert auf einer subjektiven Entscheidung (vgl. z. B. Opp 2010, S. 30), die nur „mit guten Argumenten“ [sic!] gerechtfertigt werden kann und von Forscher zu Forscher durchaus unterschiedlich ausfallen kann.

Ein weiteres Problem bei der Anwendung des Falsifikationsprinzips ist als „Basissatzproblem“ bekannt (vgl. z. B. Schnell et al. 2013, S. 75–79). Basissätze behaupten, dass zu einer bestimmten Zeit an einem bestimmten Ort etwas „der Fall ist“. Die Argumentation beim Falsifikationsprinzip steht und fällt mit der empirischen Wahrheit der verwendeten Basissätze – welche letztlich nicht garantiert werden kann, da unsere Messungen und Wahrnehmungen theorieabhängig sind. Sie beruhen damit auf Konventionen (vgl. Popper 1994, S. 73). Aufgrund empirisch falscher Basissätze können Hypothesen mit anderen Worten *zu Unrecht* als „vorläufig bewährt“ bzw. „falsifiziert“ betrachtet werden.

Darüber hinaus stellt das Basissatzproblem eine gewisse Herausforderung für die Definition des Messens dar. Wäre es zweifelsfrei möglich, ein empirisches Relativ strukturtreu in ein numerisches Relativ abzubilden, dürfte das Basissatzproblem kaum auftreten. Oder anders gewendet: Wir können nach der Argumentation beim Basissatzproblem nie sicher sein, eine „Messung“ im Sinne der genannten Definition durchgeführt zu haben. Die genannte Definition kann lediglich als „Zielvorstellung“ dienen, was sich auch in der Forderung zeigt, die Reliabilität und die Validität der eingesetzten Instrumente einer Prüfung zu unterziehen.

Ein drittes Problem bei der Anwendung des Falsifikationsprinzips resultiert in der Umfrageforschung aus der Tatsache, dass – abgesehen von der generellen Schwierigkeit, Kausalhypothesen zu prüfen (vgl. Baumgartner und Graßhoff 2004; Opp 2010; Schnell et al. 2013, S. 53; Backhaus et al. 2015, S. 66) – Kausalhypothesen anhand

eines *Ex-Post-Facto-Designs* untersucht werden, wobei der Einfluss von Drittvariablen zu Methodenartefakten führen kann (vgl. z. B. Schumann 2012, S. 119–127). Insbesondere können „Scheinkorrelationen“ auftreten (die treffender als „scheinbare Kausalzusammenhänge“ zu bezeichnen sind). Hier treten empirisch Korrelationen auf, die nicht auf kausale Einflüsse unter den betrachteten Variablen zurückzuführen sind. Die betreffenden Variablen sind voneinander kausal unabhängig und die Korrelation zwischen ihnen entsteht aufgrund ihrer gemeinsamen Beeinflussung durch eine (oder mehrere) Drittvariablen. Solche Konstellationen sind nicht „per se“ problematisch – beispielsweise können formal gesehen Scheinkorrelationen als Ausgangsbasis für Faktorenanalysen dienen. Allerdings besteht die Gefahr, dass Kausalhypothesen zu Unrecht als „vorläufig bewährt“ eingestuft werden, wenn solche Korrelationen im Rahmen der Hypothesenprüfung *kausal* interpretiert werden. Auch der umgekehrte Fall ist denkbar: Bei „scheinbaren Nonkorrelationen“ ist zwischen kausal voneinander *abhängigen* Variablen dennoch empirisch keine Korrelation feststellbar, womit sich die Gefahr ergibt, dass Kausalhypothesen zu Unrecht als „falsifiziert“ gelten (vgl. hierzu z. B. Schumann 2012, S. 119–127).

Im Überblick und zusammenfassend zeigt Abschn. 2.2 in vielerlei Hinsicht in der Tat das Bemühen, den eingangs erwähnten Gedanken einer „Einheitswissenschaft“ in die Forschungspraxis umzusetzen. Dabei treten Probleme auf, die auch aus den Naturwissenschaften bekannt sind, sowie solche, die dort nicht zu finden sind. Letztere hängen in erster Linie mit der „Natur des Messgegenstandes“ in den Humanwissenschaften zusammen.

Unterstellt wurde bisher das in Abschn. 2.1 angesprochene, auf den Vorstellungen der klassischen Physik beruhende materialistisch-deterministische Weltbild. Kap. 3 wird nun einen Schritt weiter gehen und zeigen, dass grundlegende Annahmen der klassischen Physik nach all unserem Wissen als „empirisch falsch“ zu betrachten sind. Der Kapitän ist insofern – um im weiter oben zitierten Bild zu bleiben – bereits auf die Klippen gerannt und hat Schiff und Leben verloren. Brian Greene spricht explizit vom Zusammenbruch der klassischen (Newtonschen) Physik (vgl. Greene 2006a, S. 50, 114) und führt hierzu an anderer Stelle aus:

Die wissenschaftlichen Umwälzungen, die sich ungefähr zwischen 1900 und 1930 ereigneten, mündeten in einen vehementen Angriff auf Intuition, gesunden Menschenverstand und die allgemein anerkannten Gesetze, die von der neuen Garde schon bald als ‚klassische Physik‘ bezeichnet wurden – ein Begriff, der das Gewicht und den Respekt gegenüber einem altehrwürdigen, unmittelbar verständlichen, befriedigenden und vorhersagekräftigen Bild der Realität widerspiegelt (Greene 2013, S. 239).

Anton Zeilinger (2005, S. 134) ist „... der Überzeugung, dass die ... notwendigen Änderungen unseres Weltbildes so groß sein werden, dass alle möglichen technischen Konsequenzen im Vergleich dazu klein erscheinen“ und Hans-Peter Dürr (2010, S. 14–15) schreibt:

... die Quantenphysik ... hat in den letzten siebenzig Jahren seit ihrer Ausdeutung einen beispiellosen Triumphzug durch alle Gebiete der Physik angetreten und sich bis zum

heutigen Tag unangefochten bewährt. ... Der Bruch, den die neue Physik erfordert, ist tief. Er bezeichnet nicht nur einen Paradigmenwechsel, wie dies von Thomas Kuhn in seinem Buch »The Structure of Scientific Revolutions« 1962 beschrieben worden ist. Deutet diese neue Physik doch darauf hin, dass die Wirklichkeit, was immer wir darunter verstehen, *im Grunde keine Realität* im Sinne einer *dinghaften* Wirklichkeit ist. ... Auf dem Hintergrund unserer gewohnten, durch das klassische physikalische Weltbild entscheidend geprägten Vorstellungen klingt dies ungeheuerlich, eigentlich unannehmbar (Hervorhebungen im Original).

Quantitative und qualitative empirische Forschung

Ein Diskussionsbeitrag

Schumann, S.

2018, XII, 179 S., Softcover

ISBN: 978-3-658-17833-8