

Werner Nienhüser

Sieben Versuchungen in der empirischen Personal- und Organisationsforschung

Lehrstuhl für Allgemeine Betriebswirtschaftslehre,  
insb. Arbeit, Personal und Organisation  
Fakultät für Wirtschaftswissenschaften  
Universität Duisburg-Essen  
Universitätsstraße 12  
45117 Essen  
e-mail: [werner.nienhueser@uni-due.de](mailto:werner.nienhueser@uni-due.de)  
homepage: <http://www.personal.wiwi.uni-due.de>

Essen 2011

*1. Fassung – bitte nicht ohne Zustimmung des Autors zitieren*

## 1 Einführung

Die Wissenschaftstheorie und die Arbeiten zu den Methoden der empirischen Sozialforschung formulieren explizit oder implizit Regeln, wie man „richtig“ forscht. Selbstverständlich gibt es keinen umfassenden Konsens darüber, welches die richtigen Regeln sind. Gleichwohl kommen z.B. bei der Begutachtung von Anträgen auf Forschungsgelder, in Review-Verfahren wissenschaftlicher Publikationen, aber auch bei der Bewertung von Qualifizierungsarbeiten derartige Regeln zur Anwendung. Regeln werden benötigt, um Verhalten zu lenken. Verhalten ist lenkungsbedürftig, weil Menschen Versuchungen ausgesetzt sind. Im Folgenden soll es weniger um Regeln wissenschaftlichen Arbeitens gehen, sondern um die Versuchungen, denen Wissenschaftler ausgesetzt sind, und die – wenn man ihnen nachgibt – letztlich dazu führen, dass Geltungsansprüche (Habermas 1973) schlecht begründet oder schwer nachprüfbar und diskutierbar<sup>1</sup> sind.

Es wird folgendermaßen vorgegangen: Zunächst werden sieben Versuchungen mit ihren Folgen beschrieben. Hierin liegt der Schwerpunkt des Beitrages. Darüber hinaus sollen aber auch Ansatzpunkte dafür skizziert werden, wie man entweder verhindern kann, dass den beschriebenen Versuchungen nachgegeben wird oder wie die entstandenen Fehler reduziert werden können.. Solche Maßnahmen können auf der individuellen Ebene liegen, jedoch ist kaum zu erwarten, dass Individuen sich allein durch Aufklärung und Appelle davon abhalten lassen, diesen Versuchungen nachzugeben. Wichtiger ist daher die Entwicklung von Regeln, auf deren Einhaltung z.B. Herausgeber von Zeitschriften und Gutachter ihr Augenmerk richten sollten. Darüber hinaus werden in diesem Beitrag auch Hinweise über die Entwicklung der Forschung auf der Systemebene gegeben; so wird z.B. argumentiert, dass Theorievergleiche ein Gegenmittel gegen bestimmte Versuchungen bzw. daraus entstehende Folgen sein können.

Zur Logik und Basis der Argumentation des Beitrages (i.w.S.: Methodik dieses Beitrages). Die Aussagen werden durch unsystematische Beobachtungen, zum Teil Selbstbeobachtungen, illustriert.<sup>2</sup> Man könnte kritisch und zu Recht einwenden, dass Behauptungen über Verhaltensweisen aufgestellt werden, deren Geltung noch zu prüfen ist. Letztlich sind es „Es-gibt“-Behauptungen

---

<sup>1</sup> Ich unterstelle hier, dass Vertreter sehr unterschiedlicher wissenschaftstheoretischer Positionen es als Problem ansehen, wenn Geltungsansprüche schlecht begründet oder nicht nachprüfbar sind. (Man könnte sagen, es geht hier um die Kardinaltugenden (die üblicherweise Versuchungen gegenübergestellt werden) der Forschung.) Über die Verfahren der Begründung und Nachprüfung dürfte allerdings keine Einigkeit bestehen. – Bereits an dieser Stelle ist darauf hinzuweisen, dass bei Empfehlungen (normativen Aussagen) Wertungen eingeführt werden müssen. Ich gebe hier auch Empfehlungen, die – darauf komme ich zurück - allein durch Deskriptionen nicht begründet sind. Unter Zuhilfenahme dieser Wertungen sollte sich eine logisch korrekte Argumentationsskizze entwickeln lassen.

<sup>2</sup> Ich nenne hier keine konkreten Beispiele aus Publikationen und keine Namen, weil es mir problematisch erscheint, einen bestimmten Beitrag allein deswegen herauszugreifen, weil er eine Versuchung gut zu illustrieren vermag. Es ist ein Gebot der Fairness, bei Namensnennungen eine weitergehende, systematische Analyse vorzunehmen, die ich bisher nicht geleistet habe.

tungen, es werden keine Aussagen darüber gemacht, wie häufig die genannten Versuchungen sind bzw. wie häufig den Versuchungen nachgegeben wird.

### 1.1 Begriff und eine „Theorie der Versuchung“

Zunächst zum Begriff der Versuchung. Intuitiv wissen wir vermutlich, was eine Versuchung ist. Jeder und jede war bereits Versuchungen ausgesetzt und wird ihnen ausgesetzt sein. Was ist ein sinnvoller, systematischer Begriffsgebrauch des Wortes Versuchung? Erstens: Versuchungen sind Anreize und haben grob gesprochen etwas mit Bedürfnissen zu tun, die man zu befriedigen trachtet. Zweitens: Eine Versuchung ist eine Relation zwischen einem Objekt, von dem die Versuchung „ausgeht“ und einem Menschen, der ihr unterliegt und ihr nachgeben kann oder nicht. Drittens: Wenn wir von Versuchungen sprechen, unterstellen wir im Allgemeinen zudem, dass Normen existieren, die besagen, dass man etwas nicht tun soll. Viertens nehmen wir an, dass diese Normen dem oder der Versuchten bekannt sind.

Hier ist nicht der Raum, die verhaltenstheoretischen Grundlagen einer „Theorie der Versuchungsgründe“ auszuarbeiten (siehe dazu auch Koboldt 1995, der eine „Ökonomik der Versuchung“ entwickelt), das heißt, die Ursachen und Mechanismen von Versuchungen darzustellen und zu begründen. Gleichwohl soll eine durch eine knappe Skizze angedeutet werden, wie eine solche theoretische Begründung aussehen könnte. Grundsätzlich gehe ich hier von beschränkt rationalem Verhalten aus (Simon 1972; Taylor 1975). Die entsprechenden Annahmen bedeuten u.a., dass wir bei Entscheidungen diejenige Alternative bevorzugen, deren Nutzen wir im Vergleich zu anderen und gemessen an unserem Anspruchsniveau höher einschätzen; dass wir aber die Folgen unseres Handelns nicht vollständig überblicken; dass wir Gelegenheiten ausnutzen; dass wir gezwungen sind, mit Vereinfachungen, mit „Definitionen der Situation“ (Esser 1996), ggf. mit Vorurteilen zu operieren etc. Hinzu kommt, dass wir nach „Sinn“ suchen und kognitive Dissonanz zu vermeiden oder zu reduzieren trachten (Festinger 1957). Schließlich würde eine ausgearbeitete Theorieanwendung auch berücksichtigen, dass Menschen sich an anderen orientieren, insbesondere dann, wenn sie einer Gruppe angehören wollen, deren Mitgliedschaft sie als erstrebenswert ansehen. (vgl. für viele Shaw 1981).

Im Kontext der empirischen Personal- und Organisationsforschung – vermutlich in allen Bereichen der empirischen Forschung – ist unser Verhalten durch folgende Aspekte geleitet:

(1) Aufgrund unseres intendiert (beschränkt) rationalen Verhaltens geben wir ungerne etwas auf, das wir uns mühsam angeeignet haben. Dies kann man aus zwei Perspektiven sehen: Zum einen investieren Menschen in Alternativen und möchten die Erträge aus diesen Investitionen nicht verlieren. Zum anderen: Wenn wir etwas getan haben, müssen wir die Gründe für unser Handeln mindestens vor uns selbst rechtfertigen (um kognitive Dissonanz zu vermeiden). (2) Wir haben Vorurteile, die wir lieber bestätigt als widerlegt sehen. (3) Wir neigen zu Verallgemeinerungen, zu Schlussfolgerungen über den Einzelfall hinaus. Dies ist vermutlich ähnlich wie die Entstehung von Vorurteilen durch beschränkte Rationalität, aber auch evolutionär erklärbar. Populationen mit Vor-

urteilen und der Neigung zu verallgemeinern, hatten (und haben vielleicht noch) eine höhere Überlebenschance und konnten sich daher schneller verbreiten. (4) Wir wollen praktisch gestalten. Auch wenn es große interindividuelle Unterschiede geben mag, sind wir doch alle grundsätzlich praktisch veranlagt. Wir tendieren dazu, praktische Überlegungen anzustellen, das bedeutet auch, auf Basis von Forschungsergebnissen praktische Gestaltungsmaßnahmen zu entwickeln. Dies kann man ebenfalls evolutionär begründen, aber auch durch Nutzenstreben und Vermeidung kognitiver Dissonanz.

Es geht also im Folgenden um Versuchungen, die mit dem Festhalten an Investitionen, an bestimmten Vorstellungen (ggf. Vorurteilen) etc. zu tun haben. Welche Versuchungen sollen behandelt werden? Ich bezeichne die zu behandelnden Versuchungen mit folgenden Begriffen:

- Theoriebestätigungsversuchung,
- Methodenwiederholungsversuchung,
- Datensatz-Versuchung,
- Versuchung der Ignoranz von Effektstärken,
- Versuchung der einseitigen Hypothesendiskussion,
- Versuchung der Verallgemeinerung ohne hinreichende Basis,
- Versuchung des Schlusses vom Sein auf das Sollen.

## 2 Versuchungen

Die ersten drei Versuchungen – die Theoriebestätigungsversuchung, die Methodenwiederholungsversuchung und die Datensatz-Versuchung - haben alle damit zu tun, dass wir an etwas hängen, in das wir viel Zeit investiert haben, etwa uns eine Theorie anzueignen oder sogar zu entwickeln, eine Methode, z.B. der Datenanalyse, zu erlernen oder einen Datensatz aufzubauen oder mit einem vorhandenen Datensatz umgehen zu können. Dies mag man mit Investitionen erklären, auf deren Erträge man nicht verzichten möchte oder auch mit kognitiver Dissonanz. Die ersten zwei Versuchungen werden in der Literatur (vgl. Onwuegbuzie/Daniel 2003 und die dort angegebene Literatur)<sup>3</sup> häufig als Problem genannt, die dritte wird weniger angesprochen.

### 2.1 Theoriebestätigungsversuchung

Wer eine bestimmte Theorie präferiert, wird sie immer wieder anzuwenden und Phänomene ausschließlich oder vorrangig aus dieser Perspektive sehen. Damit wird er in gewisser Weise „konstruieren“, indem bestimmte Aspekte eines Phänomens in den Vordergrund gerückt werden und andere in den Hintergrund treten. Wer sich z.B. die Luhmannsche Systemtheorie angeeignet hat, wird sie so schnell nicht wieder loswerden wollen oder sogar können. Solch eine Theorie-rezeption ist recht mühsam und zeitaufwendig. Und je komplizierter und verzweigter (vermutlich auch: je schwerer verständlich) eine Theorie ist, desto stärker wird die Bindung an diese sein. Ein zweiter Grund neben spezifischen Investitionen und Effekten der kognitiven Dissonanz sind Grup-

---

<sup>3</sup> Die Arbeit von Onwuegbuzie/Daniel (2003) beschränkt sich allerdings nahezu ausschließlich auf die erziehungswissenschaftliche Forschung in den USA.

peneffekte. Wenn es notwendig ist, eine bestimmte Theorie und damit auch eine spezifische Sprache zu verwenden, um einer Gruppe anzugehören, dann erzeugt dies ebenfalls eine Bindung an die Theorie, und zwar in dem Maße, in dem man bei Distanz oder gar Kritik an der Theorie als jemand gesehen würde, der nicht dazu gehört.

Wozu führt nun eine solche Theoriebindung, was ist hier die Versuchung? Zwar mag man es auch Versuchung nennen, die Realität mit dem „Scheinwerfer“ einer Theorie zu beleuchten. Dies ist aber unproblematisch und hier nicht gemeint. Auf der individuellen Ebene ist es notwendig oder unvermeidlich, einen Sachverhalt aus einer spezifischen theoretischen Perspektive zu sehen bzw. zu konstruieren.

Die Theoriebestätigungsversuchung besteht vielmehr darin, die Theorie beizubehalten, auch wenn empirische Daten nicht mit ihr vereinbar sind. Ein Beispiel: Ich erinnere mich an einen Vortrag, in dem es um den Wettbewerb im Telekommunikationssektor ging und der Referent ein Modell entwickelte, das vollständige Konkurrenz unterstellte. Ein Informatiker – ein „Fachfremder“ – wandte in der anschließenden Diskussion ein, es gäbe doch aber marktbeherrschende Unternehmen. Die Antwort des Referenten war kurz – dies dürfe natürlich nicht sein, in der Theorie (oder seiner Variante) gebe es ja idealerweise keine Monopole. Er war nicht einmal bereit, zu diskutieren, welche Implikationen es für sein Modell hätte, wenn man die Annahme der vollständigen Konkurrenz aufgeben bzw. einschränken würde.

Was tun? Das Problem ist auf der kollektiven Ebene zu lösen. Wir benötigen viel mehr *Theorievergleiche und aktive Theoriekonkurrenz* als wir bisher haben. Es gibt kaum konzeptionell angelegte Theorievergleiche, schon gar nicht theorievergleichende empirische Untersuchungen. Mehr und kritischere Diskussionen über Theorien sind nötig, mehr Konferenzen, auf denen sich beispielweise Foucault-Anhänger mit Vertretern der Rational Choice-Theorie streiten und auf denen sich nicht lediglich die Anhänger der jeweiligen Theorien gegenseitig bestätigen. Dies setzt sicher auch Kenntnis der jeweiligen anderen Theorien voraus.

Für wissenschaftliche Zeitschriften würde dies auch bedeuten, dass theorievergleichende Arbeiten als ein spezifischer Typ von Publikationen akzeptiert werden und dass empirische Forschung im engeren Sinne zwar im Vordergrund stehen sollte, aber nicht andere Veröffentlichungstypen verdrängen darf, eine Entwicklung, die sich nach meinem Eindruck derzeit abzeichnet.

## 2.2 Methodenwiederholungsversuchung

Methoden der Datenerhebung und –auswertung liefern Lösungen für ganz bestimmte Probleme. Wer eine bestimmte Lösung bevorzugt, kann leicht versucht sein, Probleme so zu selektieren oder zuzuschneiden, dass die gegebene Lösung passt, also anwendbar ist.

Wenn man beispielsweise eine Präferenz für die Methode der Regressionsanalyse hat, muss man für eine sinnvolle Anwendung eine Trennung in unabhängige und abhängige Variablen vornehmen. Dies mag aber dem Problem nicht angemessen sein, wenn bestimmte Sachverhalte eher als Konfi-

gurationen, technisch gesprochen als Cluster von Variablen oder Fällen, konzeptionalisiert werden sollten. So ist es etwa bei den Arbeiten zu den „Wirkungen“ von HRM-Variablen (z.B. Entlohnungs-, Personalauswahl- und Qualifizierungssysteme, Bindung des Personals) auf den Unternehmenserfolg insb. bei der Nutzung von Querschnittsdaten angemessener, nicht zwischen abhängigen und unabhängigen Variablen zu trennen, weil Unternehmenserfolg aufwendige Personalpraktiken erst ermöglicht oder zumindest fördert. Im Übrigen könnte man über solche klassifizierenden Methoden wie Cluster- oder Faktorenanalyse, multidimensionale Skalierung oder Korrespondenzanalyse auch das Problem lösen, dass bei den nicht selten bei Regressionsanalysen vorkommenden langen Listen von unabhängigen Variablen mögliche Interaktionseffekte mit erfasst werden.<sup>4</sup>

In der qualitativen Forschung findet man eine deutliche Präferenz für gruppierende bzw. typisierende Verfahren. Diese mögen ebenfalls bestimmten Fragestellungen angemessen sein. Wenn es aber eben um Fragen der Kausalität geht, sind diese nicht angemessen. Insofern wird man schon aufgrund der Bevorzugung bestimmter Verfahren der Datenanalyse weniger Kausalanalysen finden.

Vergleichbares wie für die Verfahren der Datenanalyse lässt sich für Methoden der Datenerhebung anführen. Wer schriftliche Befragungen bevorzugt, ist versucht, das Problem so zu konstruieren, dass die Methode zum Einsatz kommen kann und blendet damit bestimmte Fragestellungen aus. So dürfte es kaum möglich sein, etwas über illegale Beschäftigung mittels schriftlicher Befragungen zu erfahren. Wer qualitative Inhaltsanalysen<sup>5</sup> bevorzugt, wird in allem „Texte“ oder Dokumente sehen und Beobachtungsdaten eher nicht gewinnen wollen. Experimente als die besonders für Theorie-tests geeignete Methode werden denjenigen, die schriftliche Befragungen oder qualitative Datenerhebungsverfahren bevorzugen, mit einer geringeren Wahrscheinlichkeit in den Blick geraten.

Was ist zu tun? Man kann auf der individuellen Ebene wenig dagegen tun, dass Forscher bestimmte Methoden bevorzugen, weil sie sie besonders gut kennen. Helfen könnten ansatzweise eine breite Methodenausbildung, die eine Vielfalt an Datenanalyse- und -erhebungsmethoden lehrt und die jeweiligen Begrenzungen der Verfahren aufzeigt. In der Ökonometrieausbildung werden z.B. Datenerhebungsverfahren, gruppierende Verfahren der Datenanalyse und Skalenentwicklung zu wenig einbezogen, von qualitativen Verfahren gar nicht zu reden. Eine Ausbildung in quantitativer und qualitativer Sozialforschung wäre sinnvoll. Zum zweiten ist in der Forschung Pluralität und Wettbewerb erforderlich. Besonders nützlich wären Arbeiten, die sich derselben Fragestellung mit unterschiedlichen Methoden nähern. Zum dritten wäre es sinnvoll, wenn Gutachter z.B. auch die Frage stellen, inwieweit empirische Befunde methodenrobust sind.

---

<sup>4</sup> Sicher kann man auch in Regressionsmodellen prinzipiell Interaktionseffekte berücksichtigen, dabei taucht allerdings häufig das Problem auf, dass die Korrelation der Interaktionsvariablen mit den Variablen, auf deren Basis sie konstruiert wurden, recht hoch ist (Multikollinearität).

<sup>5</sup> Ich sehe die Inhaltsanalyse hier nicht nur als Methode der Datenanalyse, sondern auch als Verfahren der Datenerhebung.

### 2.3 Datensatz-Versuchung

Die Versuchung hier bedeutet: Daten locken Fragen an, auch solche, die nicht mit den Daten (direkt) beantwortbar sind.

Auf der einen Seite sollten Datensätze, die nach den üblichen Kriterien als „gut“ gelten und auf deren Erstellung viel Arbeit, Sorgfalt und nicht zuletzt Geld verwendet wurde, möglichst intensiv und extensiv genutzt werden. Auf der anderen Seite entsteht die Versuchung, mit vorhandenen Daten auch solche Fragen zu beantworten, die mit ihnen allenfalls sehr indirekt beantwortet werden können. Diese Tendenz ist ebenfalls erklärbar: Wer viel Mühe in einen Datensatz gesteckt hat, wird ihm tendenziell Qualität zuschreiben, das Antwortpotenzial kann dabei leicht überschätzt werden.

Wenn man mit Hilfe nicht direkt passender Datensätze Fragen zu beantworten versucht, hat dies Folgen. Eine erste, wesentliche Folge besteht darin, dass die Verbindung zwischen den theoretischen und den empirischen Konstrukten „lose“ ist und dass man deswegen in Versuchung gerät, die entstehenden Lücken ad hoc und nicht systematisch und theoretisch geleitet zu überbrücken. Dies wiederum kann nach sich ziehen, dass man eine Präferenz für Theorien entwickelt, die wenig Informationsgehalt aufweisen und sich daher (scheinbar) eignen, die Kluft zu überbrücken.

Beispiele finden sich sehr oft im Zusammenhang mit der Variablen „Betriebsgröße“. Man findet sie als Proxy für den Legitimationsbedarf einer Organisation (je größer die Organisation, desto mehr steht sie in der Öffentlichkeit und desto höher ist der Rechtfertigungsdruck), man interpretiert sie im Sinne von „economies of scale“ oder als Ausdruck von Transaktionskosten für individuelle Aushandlungen von Verträgen im Vergleich zu Kollektivtarifverträgen (je größer der Betrieb, desto größer die Transaktionskosten individueller Verhandlungen, siehe Nienhüser/Hoßfeld 2008). Um nicht missverstanden zu werden: In aller Regel wird man nicht umhinkönnen, zumindest gelegentlich Proxy-Variablen mit gewagten Messhypothesen zu verwenden. Wenn dies aber eine Vielzahl von Variablen in einer statistischen Analyse betrifft und zudem die Geltung der Messhypothesen weder geprüft noch diskutiert wird, dann ist dies problematisch.

Zum zweiten besteht die Schwierigkeit, die Wirkungen von Variablen konsistent inhaltlich zu interpretieren. Je weiter „entfernt“ die einzelnen Variablen von den eigentlich zu messenden Größen sind, umso schwieriger wird es, die Effekte zu interpretieren. Oft verwendet man gerade bei Analysen auf Basis des SOEP oder des IAB-Betriebspanels lange Listen von Variablen, die ad hoc mit Hilfe einer Theorie interpretiert werden. Weit verbreitet ist der Gebrauch der ökonomischen Theorie, die auf jeder Ebene - Individualverhalten und Verhalten von Unternehmen – anwendbar ist und mit deren Hilfe jede Variable einzeln deutbar ist. Aber auch neo-institutionalistische Theorien (z.B. Scott 2001) sind sehr beliebt bei der Nutzung solcher Datensätze. Datensätze ziehen also geradezu die Nutzung von Theorien an, deren Begriffe und/oder Aussagen interpretationsoffen sind. Bei der ökonomischen Theorie fungiert dann der Nutzenbegriff als „leerer Sack“ (Matiaska, Simon: „empty sack“). Bei der neo-institutionalistischen Theorie (stellvertretend: Scott 2001) oder mehr noch bei der Strukturierungstheorie (Giddens 1984) sind jedoch nicht nur die Begriffe inhaltlich zu füllen –

dies ist meist unvermeidlich -, sondern auch die Aussagen sind recht offen. Meine Vermutung ist, dass Theorien auch deswegen Verbreitung finden, weil sie recht flexibel für alle möglichen Deutungen genutzt werden können.

Während ich hier solche Versuchungen als eine Verhaltenstendenz sehe, die leicht zu Beliebigkeit von Interpretationen und letztlich Immunisierung von Aussagen führt, fordert Kelle (2005) sogar, dass man (in der qualitativen Forschung) offene Theorien bzw. Konzepte einsetzen solle, weil man dann die Daten nicht in das Prokrustesbett der Theorie presse.<sup>6</sup>

Was tun? Zu empfehlen ist hier, den geringen Informationsgehalt von Theorien zu kritisieren und als Bewertungskriterium bei Publikationen zu verwenden. Gleiches gilt für Operationalisierungen und die Geltung von Messhypothesen.

#### 2.4 Versuchung der Ignoranz von Effektstärken

Die Versuchung besteht darin, Effektstärken zu ignorieren (Onwuegbuzie/Daniel 2003: 8) und dafür die Signifikanz in den Vordergrund zu rücken (zu den Problemen des Signifikanztests vgl. Martin 1989; Ziliak/McCloskey 2008). Die „Sternchen“ zählen, sonst nichts. Der Versuchung unterliegt man besonders bei solchen statistischen Methoden, deren Koeffizienten inhaltlich nicht ohne weiteres zu interpretieren sind, dies ist z.B. bei der logistischen Regression der Fall. Ob eine Variable tatsächlich einen inhaltlichen Unterschied macht, wird häufig nicht diskutiert (eine Ausnahme ist hier z.B. Dilger 2006).

Dies hat zwei Folgen: Erstens kann man behaupten, eine Hypothese sei bestätigt oder widerlegt, wenn und weil ein Unterschied durch eine Variable bestehe, und zwar unabhängig davon, wie groß dieser Unterschied ist. Dies gilt natürlich vor allem bei großen Fallzahlen, hier sind bekanntermaßen bereits sehr kleine Mittelwertunterschiede oder Prozentsatzdifferenzen, niedrige Korrelationen etc. signifikant (Martin 1989). Zweitens entsteht ohne Berücksichtigung der Effektstärken der Eindruck, dass eine Variable, die einen signifikanten Unterschied macht, auch inhaltlich-praktisch von Bedeutung ist.

Was tun? Bei Regressionen und ähnlichen Verfahren sollten möglichst auch die unstandardisierten bzw. marginalen Effekte dargestellt werden. Auch deskriptive, einfache Mittelwertvergleiche, Kreuztabellen oder Korrelationen sind nützlich, um Effektstärken wenigstens annäherungsweise einschätzen zu können. Es ist darüber hinaus sinnvoll, auch die praktische Bedeutung von Effekten zu diskutieren.<sup>7</sup>

---

<sup>6</sup> “Theoretical concepts with low empirical content, however, can play an extremely useful role if the goal of empirical research is not the testing of predefined hypotheses but the empirically grounded generation of theories, since they do not force data into a Procrustean bed—their lack of empirical content gives them flexibility so that a variety of empirical phenomena can be described with their help. Although such concepts cannot be “tested” empirically, they may be used as heuristic concepts which represent “lenses” through which researcher perceive facts and phenomena in their research field” (Kelle 2005).

<sup>7</sup> So fordern Wilkinson und die “Task Force on Statistical Inference” für die Psychologie: “Always present effect sizes for primary outcomes. If the units of measurement are meaningful on a practical level (e.g., number of cigarettes smoked per day), then we usually prefer an unstandardized measure (regression coefficient or mean difference) to a

Das Problem der Ignoranz von Effektstärken ist keineswegs auf quantitative Studien beschränkt. In qualitativen Studien werden nicht nur Typologien (auf mehr oder weniger nachvollziehbare Art und Weise) gebildet, sondern auch der Einfluss einer oder mehrerer Variablen auf eine oder mehrere andere untersucht bzw. als Resultat dargestellt. Möglicherweise ist dies für die qualitative Forschung schwer, jedenfalls dann, wenn sie ihre Befunde nicht in quantitativen Begriffen formulieren will, wie dies Onwuegbuzie/Daniel (2003: 8) vorschlagen. Dennoch muss die Frage danach, wie stark ein Unterschied ist, den eine Variable macht, gestellt und beantwortet werden.

## 2.5 Versuchung der einseitigen Hypothesendiskussion

Wer empirische Untersuchungen durchführt, wird in aller Regel nicht seine sämtlichen Hypothesen bestätigt finden. Was macht man dann mit den unbestätigten Hypothesen bzw. mit Ergebnissen, die den Hypothesen widersprechen? Üblicherweise diskutiert man mehr oder weniger ausführlich, warum die jeweiligen Hypothesen nicht bestätigt wurden. Man führt Messprobleme an, macht die Stichprobenszusammensetzung dafür verantwortlich etc. Die Versuchung, der viele nachgeben, besteht darin, diese Probleme lediglich für die nicht bestätigten, nicht aber für die bestätigten Hypothesen zu diskutieren. Dabei ist nicht auszuschließen, dass zumindest einige der diskutierten Schwierigkeiten nicht nur dafür verantwortlich sind, dass Hypothesen nicht bestätigt wurden sondern auch dafür, dass einige irrtümlich bestätigt wurden. So mögen vielleicht Messprobleme nur für die nicht bestätigten Hypothesen gelten, wer Stichprobenprobleme anführt, sollte aber prüfen, warum diese für die bestätigten Hypothesen nicht gelten.

Was tun? Zumindest sollte man für den Fall verlangen, dass entsprechende Probleme für die nicht bestätigten Hypothesen diskutiert werden, diese Diskussion auch auf die bestätigten ausgeweitet wird. Dies würde eine Art „Robustheitscheck“ bedeuten, der sich wiederum keineswegs auf quantitative Analysen beschränken sollte.

## 2.6 Versuchung der Verallgemeinerung ohne hinreichende Basis

Anders als bei deduktiven Argumenten geht bei induktiven Argumenten der Gehalt der Konklusionen über den der Prämissen hinaus (Salmon 1983: 163). Die einfachste Form des induktiven Arguments ist die „enumerative Induktion“ (Salmon 1983: 167): Man schließt von Aussagen über Elemente in der untersuchten Stichprobe (einer Teilmenge) auf die (nicht untersuchten) Elemente der gesamten Menge (Grundgesamtheit). Der „Fehlschluss der unzureichenden Statistik“ (oder wie ihn Salmon (1983: 170) auch nennt: der „Fehler des voreiligen Schlusses“) besteht darin, dass man verallgemeinert, ohne genügend Daten gesammelt, also zu wenig Fälle zur Verfügung zu haben. Salmon (1983: 172) nennt zudem den „Fehlschluss der voreingenommenen Statistik“<sup>8</sup>. Hier gibt man der Versuchung nach, auf Grundlage einer nicht-repräsentativen Stichprobe zu verallgemei-

---

standardized measure (r or d). It helps to add brief comments that place these effect sizes in a practical and theoretical context” (Wilkinson/Task Force on Statistical Inference 1999: 12; s.a. Onwuegbuzie/Daniel 2003: 8).

<sup>8</sup> Hier sollte man besser den Ausdruck „Fehlschluss *aufgrund* voreingenommener Statistik“ verwenden.

nern. Diese Fehler sind in der quantitativen Forschung leichter zu erkennen als in der qualitativen Forschung und man wird daher den entsprechenden Versuchungen weniger häufig nachgeben.

Wie sieht es aber in der qualitativen Forschung aus? Man mag sagen, dass Quantifizierungen dort keine Rolle spielen und daher analoge Überlegungen nicht sinnvoll angestellt werden können. Zudem könnte man einwenden, dass in der qualitativen Forschung Verallgemeinerungen nicht – oder zumindest nicht in Form einer enumerativen Induktion vorgenommen würden. Beides ist aber nicht zutreffend. Zum ersten begegnen uns auch in der qualitativen Forschung „Quasi-Statistiken“ (Barton/Lazarsfeld 1984).<sup>9</sup> Man findet häufig (sic!) Aussagen, die auf unbestimmte Häufigkeitsbegriffe (zu diesem Ausdruck Hopf 1985) zurückgreifen, wobei Begriffe wie „häufig“, „meistens“, „die Mehrheit“, „weniger“, „mehr“, „mitunter“, aber auch „typisch“, „in der Regel“ verwendet werden. Gelegentlich findet man auch bestimmte Häufigkeitsbegriffe wie „fünf der neun Befragten“, „in einem von vier Fällen“ oder ähnliches. Dies ist an sich nicht problematisch bzw. sogar begrüßenswert, insb. wenn man fordert, dass Effektstärken auch in der qualitativen Forschung genannt werden sollen. Es besteht nun aber die Versuchung, auf Basis dieser Quantifizierungen zu verallgemeinern.<sup>10</sup>

Das Problem ist besonders gravierend, weil das qualitative Paradigma (bzw. Teile dieses Paradigmas)<sup>11</sup> dadurch charakterisiert ist, dass man keine Theorie vorab haben oder anwenden, sondern aus den empirischen Daten (induktiv?) Hypothesen oder sogar eine Theorie entwickeln soll.

Wie kann auf Basis weniger Fälle verallgemeinert werden? Meiner Meinung nach ist eine rein statistische Verallgemeinerung ohne eine Theorie nicht sinnvoll, man würde ansonsten induktiv von zu wenigen Fällen auf alle Fälle schließen und den „Fehlschluss der unzureichenden Statistik“ (Salmon 1983: 170) vollziehen. Daher wird auch von manchen Vertretern der qualitativen Forschung gefordert, das Prinzip der Falsifikation, das eine Theorie oder wenigstens eine Hypothese voraussetzt, anzuwenden (Kelle 2005).

Was ist zu tun, worauf ist zu achten? Wir neigen alle dazu, zu verallgemeinern. Solange klar ist, dass die Verallgemeinerung lediglich eine Hypothese darstellt, die noch zu testen ist, dürfte dies kein allzu großes Problem sein. Das Problem wäre auch geringer, wenn mehr qualitativ Forschende die Forderung der „Grounded Theory“ ernst nähmen, aus den Daten theoretische Aussagen zu ent-

---

<sup>9</sup> Barton/Lazarsfeld (1984: 70) fassen unter Quasi-Statistiken nicht nur Quasi-Verteilungen, sondern u.a. auch Quasi-Korrelationen. Sie führen als Beispiel für ein Quasi-Korrelation die Aussage an: „Bandenmitglieder betreiben eine Ökonomie des Geldausgebens, während Studenten in einer Ökonomie des Geldsparens leben“ (S. 70). Auch diese Art von Aussagen wird man in der neueren qualitativen Personal- und Organisationsforschung häufig finden.

<sup>10</sup> Eine unzureichende Zahl von Fällen wäre schon dann ein Problem, wenn jemand – was im qualitativen Paradigma jedoch nicht vorkommt – einen strikten Falsifikationismus vertreten würde, der bereits einen widersprechenden Fall als Widerlegung einer Hypothese auffasst. Ein oder wenige Fälle bieten allenfalls Anlass, eine Hypothese zu bezweifeln, nicht aber sie zu verwerfen. Erst recht nicht wird man eine allgemeine Theorie (und nicht nur eine Hypothese) aufgrund weniger abweichender Fälle verwerfen wollen.

<sup>11</sup> Es gibt selbstverständlich nicht *das* qualitative Paradigma. Auch die Richtung der sog. Grounded Theory ist keineswegs homogen. So sind sich Glaser und Strauss, zwei exponierte Vertreter, der qualitativen Forschung, in der Frage nach der Bedeutung und Handhabung vorgängiger Theorien völlig uneinig und sie publizieren nicht mehr gemeinsam (vgl. Kelle 2005).

wickeln (die anschließend von anderen geprüft werden könnten). Für die wissenschaftlichen Zeitschriften bedeutet dies, dass qualitative, kleinzahlige Untersuchungen im Allgemeinen inakzeptabel sein sollten, wenn sich diese zu Beginn ihres Beitrages methodologisch und methodisch auf die „Grounded Theory“ stützen und dann aber eben keine „Theorie“ entwickeln, sondern wie „brave Empiristen“ ihre Befunde berichten und unterlassen, auf die Unmöglichkeit der Verallgemeinerung hinzuweisen.

## 2.7 Versuchung des Schlusses von Sein auf das Sollen

Hier geht es um die Versuchung, aus empirischen Befunden normative Aussagen „abzuleiten“, ohne die für einen korrekten Schluss notwendigen Werte einzuführen und kritisch zu diskutieren oder wenigstens die Diskussion über die Werte durch deren Nennung zu erleichtern ( s.a. detailliert an einem konkreten Fall Martin 1998). Ein Beispiel: Wer herausfindet, dass die Fluktuation in einem Unternehmen überdurchschnittlich hoch ist, kann daraus nicht folgern, dass die Fluktuation gesenkt werden müsse. Warum soll eine hohe Fluktuation schlecht sein? Es mag eine Situation gegeben sein, in der Mitarbeiter freiwillig das Unternehmen verlassen und man daher auf zeit- und kostenträchtige Kündigungen, Abfindungszahlungen etc. verzichten kann. Man müsste also eine normative Aussage wie „Fluktuation ist gut (schlecht)“ einführen. Die Geltung dieser normativen Aussage wäre wiederum zu diskutieren. Es geht mir hier nicht um den Inhalt dieses Beispiels, sondern lediglich um die Logik der Argumentation, um die Logik des Schlusses. Kurz gesagt: Ein Schluss von deskriptiven Aussagen (Sein) auf normative Aussagen (Sollen) ist logisch nicht korrekt. Er wäre aber dann logisch korrekt möglich, wenn man gültige Werte einführt.

Hinzu kommt ein weiteres Problem, wenn man der Versuchung der Entwicklung von Empfehlungen nachgibt. Was ist z.B. mit der Empfehlung gewonnen, man müsse eine bestimmte HRM-Konfiguration einführen, um den Unternehmenserfolg zu steigern? Es wäre zu klären, wie die Konfiguration konkret aussehen soll, ob nicht Nebenwirkungen zu erwarten sind, ob nicht die Kosten der Maßnahme deren Nutzen (über)kompensieren usw. (Martin 1983; 1984; ähnlich Nienhäuser 1988). Dies kann aber zumindest nicht in empirischen Aufsätzen ausgearbeitet werden. Vielmehr erforderte dies eine neue Arbeit mit einer Reihe neuer Fragen. Daher bietet es sich an, entweder auf die Nennung von Gestaltungsempfehlungen zu verzichten oder allenfalls Möglichkeiten anzudeuten, jedoch klarzumachen, dass dies Gegenstand einer neuen Arbeit sein muss. Andernfalls werden Praktiker zu recht mit den Empfehlungen nichts anfangen können.

Eine Empfehlung (sic!) für wissenschaftliche Zeitschriften besteht darin, dass die Entwicklung von Aussagen über Gestaltungsmöglichkeiten im Anschluss an empirische Untersuchungen sinnvoll sein und eine eigene Kategorien von Veröffentlichungen neben Theoriewerken oder empirischen Analysen (und anderen Formen) darstellen können.

### 3 Fazit

Gegen die Versuchungen bzw. gegen ihre negativen Folgen schlage ich zusammenfassend folgendes vor:

1. Gegen die Theoriebestätigungsversuchung: mehr Theorievergleiche und aktive Theoriekonkurrenz; theorievergleichende Arbeiten sollten mehr als bisher als ein interessanter Typ von Publikationen akzeptiert werden.
2. Gegen die Methodenwiederholungsversuchung: eine breite Methodenausbildung, die eine Vielfalt an Datenanalyse- und –erhebungsverfahren lehrt und auch deren jeweiligen Begrenzungen aufzeigt; wissenschaftliche Arbeiten, die sich derselben Fragestellung mit unterschiedlichen Methoden nähern und die Methodenrobustheit empirischer Befunde prüfen.
3. Gegen die Datensatz-Versuchung: Theorien mit geringem Informationsgehalt kritisieren; Operationalisierungen und die Geltung von Messhypothesen problematisieren.
4. Gegen die Versuchung der Ignoranz von Effektstärken: unstandardisierte bzw. marginale Effekte darstellen; ergänzend zu multivariaten Verfahren deskriptive Daten und bivariate Analysen; die praktische Bedeutung von Effekten diskutieren.
5. Gegen die einseitige Hypothesendiskussion: Probleme für die nicht bestätigten Hypothesen auch für die bestätigten diskutieren.
6. Gegen die Versuchung der Verallgemeinerung ohne hinreichende Basis: eine statistische Verallgemeinerung ohne eine Theorie sollte nicht akzeptiert werden; Verallgemeinerungen sollten als zu prüfende Hypothesen dargestellt und aufgefasst werden.
7. Gegen die Versuchung des Schlusses von Sein auf das Sollen: auf Gestaltungsempfehlungen verzichten oder mit Begründungen und Konsequenzen ausarbeiten; dies bedeutet die explizite Einführung, Diskussion und Begründung von Werten.

Die Vorschläge lassen sich in drei Gruppen unterscheiden. Die meisten fallen in eine erste Gruppe, die sich an Gutachter bei der Begutachtung von Publikationen oder Forschungsanträgen wendet. Eine zweite Gruppe von Vorschlägen zielt auf die Ausbildung von (künftigen) Wissenschaftlern bzw. derjenigen, die wissenschaftliche Arbeiten verfassen. Und schließlich gibt es Vorschläge, die am System der Wissenschaft selbst ansetzen, etwa die Forderung nach mehr Wettbewerb und Pluralität. Diese haben an dieser Stelle zugebenermaßen Leerformel-Charakter.

## Literatur

- Barton, A.H.; Lazarsfeld, P.F. (1984): Einige Funktionen von qualitativer Analyse in der Sozialforschung. In: Hopf, C.; Weingarten, E. (Hg.): *Qualitative Sozialforschung*. 2. Aufl., Stuttgart: 41-111.
- Dilger, A. (2006): Kooperation zwischen Betriebsrat und Management. Die Sicht beider Seiten und deren Folgen. In: *Jahrbücher für Nationalökonomie und Statistik*, 226: 562-587.
- Esser, H. (1996): Die Definition der Situation. In: *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 48: 1-34.
- Festinger, L. (1957): *A Theory of Cognitive Dissonance*. Evanston.
- Giddens, A. (1984): *Die Konstitution der Gesellschaft. Grundzüge einer Theorie der Strukturierung*. Frankfurt am Main.
- Habermas, J. (1973): Wahrheitstheorien. In: Fahrenbach, H. (Hg.): *Wirklichkeit und Reflexion*. Pfullingen: 211-266.
- Hopf, C. (1985): Fragen der Erklärung und Prognose in qualitativen Untersuchungen. Dargestellt am Beispiel der Arbeitslosen von Marienthal". In: Lutz, B. (Hg.): *Soziologie und gesellschaftliche Entwicklung. Verhandlungen des 22. Deutschen Soziologentages in Dortmund 1984*. Frankfurt a.M.: 303-316.
- Kelle, U. (2005): "Emergence" vs. "Forcing" of Empirical Data? A Crucial Problem of "Grounded Theory" Reconsidered. In: *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 6(2), Art. 27, <http://nbn-resolving.de/urn:nbn:de:0114-fqs0>.
- Koboldt, C. (1995): *Ökonomik der Versuchung: Drogenverbot und Sozialvertragstheorie*. Tübingen.
- Martin, A. (1983): *Die Entwicklung von Sozialtechnologien. Schwierigkeiten bei der Ableitung von Handlungsempfehlungen. Schriften aus dem Arbeitskreis Betriebswirtschaftliche Verhaltensforschung*. Universität Paderborn.
- Martin, A. (1984): *Theorie - Empirie - Praxis. Ableitungsprobleme am Beispiel des Arbeitszufriedenheitskonzeptes. Schriften aus dem Arbeitskreis Betriebswirtschaftliche Verhaltensforschung*. Universität Paderborn.
- Martin, A. (1989): *Die empirische Forschung in der Betriebswirtschaftslehre*. Stuttgart.
- Martin, A. (1998): Wider die Datenpraxeologie. *Zeitschrift für Personalforschung*, 12(4), 474-487.
- Nienhüser, W. (1988): Probleme der Anwendung von Theorien für personalwirtschaftliche Gestaltungsmaßnahmen. In: *Zeitschrift für Personalforschung*, 2: 3-26.
- Nienhüser, W.; Hoßfeld, H. (2008): *Verbetrieblichung aus der Perspektive betrieblicher Akteure*. Frankfurt am Main.
- Onwuegbuzie, A.J.; Daniel, L.G. (2003): Typology of Analytical and Interpretational Errors in Quantitative and Qualitative Educational Research. In: *Current Issues in Education*, 6: (online: <http://cie.ed.asu.edu/volume6/number2>).
- Salmon, W.C. (1983): *Logik*. Stuttgart.
- Scott, W.R. (2001): *Institutions and Organizations*. Thousand Oaks, London, New Delhi.
- Shaw, M.E. (1981): *Group dynamics: The psychology of small group behavior*. McGraw-Hill College.
- Simon, H.A. (1972): Theories of bounded rationality. In: McGuire, C. B.; Radner, R. (Hg.): *Decision and Organization*. Amsterdam: 161-176.
- Taylor, R.N. (1975): Psychological Determinants of Bounded Rationality: Implications for Decision-making Strategies. In: *Decision Sciences*, 6: 409-429.
- Wilkinson, L.; Task Force on Statistical Inference, (1999): Statistical Methods in Psychology Journals: Guidelines and Explanations. In: *American Psychologist*, 54: 594-604.
- Ziliak, S.T. McCloskey, D.N. (2008): *The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives*. Ann Arbor.