

Forschungsdesign in der Politikwissenschaft: Ein Dialog zwischen Theorie und Daten

Thomas Gschwend und Frank Schimmelfennig

Ob sie als »Quantis« ihre Statistikprogramme auf Tausende von Datenpunkten loslassen oder als »Qualis« einzelne Ereignisse im Detail erzählen – alle Politikwissenschaftlerinnen¹ sind mit prinzipiell den gleichen Herausforderungen bei der Konzipierung ihrer Forschung konfrontiert. Wie sie mit diesen Herausforderungen umgehen, definiert das Forschungsdesign ihrer Projekte. Ein Forschungsdesign ist ein Plan, der festlegt, wie das Forschungsprojekt ausgeführt werden soll, und insbesondere, wie empirische Evidenz dafür verwendet werden soll, um Antworten auf die Forschungsfragen zu erhalten.²

Was ist ein relevantes Forschungsproblem? Wie verbessert man Konzepte und Messungen? Welche und wie viele Variablen und Fälle sollte man auswählen? Wie kann man konkurrierende Erklärungen ausschließen, und welche theoretischen Schlüsse lassen sich aus der Forschung ziehen? Dies sind die zentralen Fragen, die sich stellen, sobald Studierende für eine Masterarbeit oder Dissertation ein eigenes Forschungsprojekt definieren. Sie sind die Adressaten dieses Buches. Es soll dabei helfen, Kernprobleme des Forschungsdesigns zu erkennen und Lösungsansätze zu entwickeln.

Das Buch entstand aus einem Seminar, das die beiden Herausgeber über mehrere Semester hinweg geleitet haben. Gerade weil es von Wissenschaftlerinnen aus dem qualitativen und dem quantitativen »Lager« besucht wurde, waren Missverständnisse und hitzige Debatten unabwendbar. Am Ende führten die Diskussionen jedoch zu einer Reihe von gemeinsamen Überzeugungen.³ Sie bilden die Grundlage für diesen Band:

1 Aus Platzgründen verwenden wir die weiblichen und männlichen Formen in wahlloser Folge.

2 Für ähnliche Definitionen siehe Brady u.a. (2004: 302); De Vaus (2001: S. 9) oder King u.a. (1994: 118).

3 Siehe dazu auch die Diskussion über Trennendes und Verbindendes dieser Lager in Schnapp u.a. (2006: 12–18).

- Der bestehende methodologische Pluralismus ist eine Stärke der Politikwissenschaft und keine Schwäche.
- Die grundlegenden Probleme des Forschungsdesigns sind dieselben für qualitative und quantitative Politikwissenschaft.
- Die methodologische Debatte innerhalb der Disziplin bleibt häufig zu abstrakt und bietet nicht genug praktische Hinweise.
- Die Unterscheidung von qualitativer und quantitativer Forschung ist oftmals unangemessen. Einige Lösungsansätze für Probleme des Forschungsdesigns treffen auf beide zu, andere liegen quer zur qualitativ-quantitativen Kluft.
- Jeder Lösungsansatz hat seine Stärken und Schwächen.

Wir haben uns entschieden, in den Beiträgen zu diesem Band jeweils spezifische Probleme des Forschungsdesigns zu behandeln, anstatt eine breite allgemeine methodologische Diskussion zu führen. Die Beiträge stellen das jeweilige Problem vor, diskutieren verschiedene Lösungsansätze mitsamt ihren Stärken und Schwächen, geben praktische Hinweise und illustrieren diese anhand von Beispielen aus der eigenen Forschung. Im Folgenden bieten wir einen Überblick über die wichtigsten Probleme und die grundlegenden Typen des Forschungsdesigns, mit denen die Beiträge zu diesem Band sich anschließend näher befassen.

Zentrale Probleme des Forschungsdesigns

Grob gesagt kann man wissenschaftliche Forschung als Dialog zwischen Theorie und Daten verstehen. Wissenschaftlerinnen formulieren eine Theorie, analysieren Daten, um die Theorie zu testen, und modifizieren die Theorie anhand der neugewonnen empirischen Befunde. Der gleiche Prozess kann auch bei den Daten beginnen: Wissenschaftler machen Beobachtungen, entwickeln Theorien, um diese Beobachtungen zu erklären, und sammeln dann zusätzliche Daten, um ihre Theorien zu testen. Nicht jedes Forschungsprojekt muss alle Schritte innerhalb dieses Zyklus durchlaufen, denn Forschung ist ein kollektives Unterfangen. Während sich einige Projekte auf das Testen bestehender Hypothesen konzentrieren, erklären andere einzelne Beobachtungen und generieren neue Hypothesen.

Wir behaupten jedoch, dass alle Forschungsprojekte, die Teil des Dialogs zwischen Theorie und Daten sind, die gleichen Kernprobleme des

Forschungsdesigns bearbeiten und lösen müssen: die Definition der Forschungsfrage; die Spezifikation von Konzepten und Theorien; Operationalisierung und Messung; die Auswahl der Fälle und Beobachtungen; die Kontrolle von alternativen Erklärungen; und theoretische Schlussfolgerungen.⁴ In den folgenden Abschnitten gehen wir auf diese Probleme genauer ein.⁵

Definition des Forschungsproblems

Zu allererst stellt sich für jeden Wissenschaftler die Frage: »Woran sollte ich arbeiten?«. Die einfachste Antwort auf diese Frage ist: »An etwas Relevantem«. Aber relevant für wen und inwiefern? Generell können wir theoretische oder wissenschaftliche von gesellschaftlicher Relevanz unterscheiden (vgl. King u.a. 1994: 15). Forschung ist wissenschaftlich relevant, wenn sie den Dialog zwischen Theorie und Daten über den aktuellen Forschungsstand hinaus vorantreibt – zum Beispiel indem sie Theorien testet und verbessert, zusätzliche Daten gewinnt und neue Beobachtungen beschreibt und erklärt. Wissenschaftliche Relevanz zeigt sich darin, dass die Forscherin Rätsel oder Probleme identifiziert und löst, wie zum Beispiel theoretische Kontroversen, ungetestete Theorien, unerklärte Beobachtungen, unzuverlässige oder ungültige Daten und Messungen. Gesellschaftlich relevant ist Forschung dann, wenn sie sich mit gesellschaftlichen Problemen befasst, das Problemverständnis von Bürgern und Entscheidungsträgern verbessert und möglicherweise Lösungsvorschläge bietet. Das Forschungsdesign muss also die gesellschaftliche Relevanz der Forschungsarbeit verdeutlichen und aufzeigen, was sie zum Verständnis und zur Lösung sozialer Probleme beitragen kann (Gerring/Yesnowitz 2006). Natürlich weist der jeweilige Stand der Forschung vielfältige Rätsel und Probleme auf, und an sozialen Problemen ist ebenfalls kein Mangel. Wissenschaftler müssen also immer auch entscheiden (und rechtfertigen), welches dieser Probleme sie vorrangig angehen wollen.

4 Für ähnliche Problemlisten siehe King u.a. (1994) und Collier u.a. (2004a: 36–37).

5 In Wirklichkeit folgen Forschungsdesigns selten dieser strikten Reihenfolge. Anfang und Ende dieses Prozesses sind nicht starr, sondern es handelt sich um einen rekursiven Prozess zwischen den Designproblemen.

Konzeptspezifikation

Ganz gleich ob Theorien formuliert oder getestet werden, ob Beobachtungen beschrieben oder erklärt werden, Politikwissenschaftler benutzen zwangsläufig Konzepte wie »Demokratie«, »Partei«, »Konflikt« oder »Frieden«. Um Forschung relevant zu machen, müssen diese Konzepte nicht nur theoretisch und/oder sozial von Bedeutung sein, sondern auch exakt spezifiziert werden. Es muss klar sein, was genau mit dem Konzept gemeint ist, das heißt was seine definierenden Attribute sind, wie Attribute und Konzepte zueinander stehen und welche empirischen Phänomene sie ein- und ausschließen. Wodurch wird »Demokratie« definiert? Schließt »Frieden« »Konflikt« aus? Wodurch unterscheiden sich »Parteien« von anderen Organisationen? Ohne klare und unmissverständliche Konzepte können überprüfbare Theorien gar nicht erst formuliert werden. Darüber hinaus gehört es zu jeder theoretischen Kontroverse, die Konzepte der rivalisierenden Theorien zu untersuchen – insbesondere wenn sie die gleiche Terminologie benutzen. Auch wenn der Startpunkt der Arbeit bei den Daten liegt, müssen Konzepte genau spezifiziert werden, um etwa Behauptungen wie »Die Mehrheit aller Staaten sind Demokratien« oder »Die Häufigkeit von Kriegen nimmt ab« abstützen zu können.

Theoriespezifikation

Kausale Theorien formulieren Ursache-Wirkungs-Beziehungen zwischen Konzepten. Demnach müssen nicht nur die Konzepte an sich spezifiziert werden, sondern auch deren Beziehungen untereinander. Zunächst legen Theorien die Reihenfolge des kausalen Zusammenhangs zwischen Konzepten fest: Was ist Ursache, was ist Wirkung? Weiterhin können Aussagen zur Form (linear oder nicht-linear) und der Richtung (positive oder negativ) des Zusammenhangs gemacht werden. Auch müssen Theorien die Beziehung zwischen unterschiedlichen vermuteten Ursachen festlegen. Ist die Beziehung additiv, wie im einfachsten Fall des linearen Regressionsmodells angenommen, oder multiplikativ, wenn Interaktionseffekte unterstellt werden? Alternativ können Ursachen als notwendige und/oder hinreichende Bedingungen gekennzeichnet werden. So umschreibt zum Beispiel die Theorie des demokratischen Friedens »gemeinsame Demokratie«, also die Tatsache, dass zwei Länder jeweils Demokratien sind, als hinreichende (aber nicht notwendige) Bedingung für andauernden Frieden zwischen ihnen.

Weiterhin sollten Theorien den kausalen Mechanismus spezifizieren, der Ursache und Wirkung verbindet, und darlegen, durch welchen Prozess die Ursache die kausale Wirkung ausübt. So wurde beispielsweise der demokratische Frieden durch Transparenz und Trägheit in der politischen Entscheidungsbildung von Demokratien erklärt, so dass militärische Eskalation verlangsamt wird und genug Zeit für Verhandlungen und die friedliche Lösung von Konflikten vorhanden ist (Russett 1993: 38–40). Anhänger der Analyse kausaler Mechanismen fordern allgemein, dass sozialwissenschaftliche Theorien die »Mikrofundierungen« kollektiver Handlungsergebnisse präzisieren müssen (Coleman 1990; Hedström/Swedberg 1998). Das heißt, sie müssen darlegen, wie soziale Strukturen und Umwelten individuelle Bedürfnisse und Überzeugungen prägen (makro-mikro), wie die Akteure auf deren Basis Präferenzen ausbilden und handeln (mikro-mikro) und wie das Handeln vieler Individuen kollektive, soziale Folgen hervorbringt (mikro-makro). Hierbei gilt, je genauer eine Theorie spezifiziert ist, desto potenziell vollständiger können Beobachtungen erklärt werden und desto besser kann die Theorie getestet werden.

Messung

Durch die Spezifikation von Konzepten und Theorien entstehen überprüfbare theoretische Aussagen. Um diese Aussagen empirisch zu testen, müssen die Konzepte allerdings operationalisiert und gemessen werden. Natürlich kann man – auch als genau spezifiziertes Konzept – Demokratie nicht direkt beobachten. Dies gilt oftmals auch für ihre Attribute. Alvarez u.a. (1996) definieren Demokratie beispielsweise als ein politisches Regime, in welchem Ämter durch kompetitive Wahlen besetzt werden. In der Folge beschreiben sie die genauen »Operationalisierungsregeln«, um zu spezifizieren, um welche Ämter (die Regierungsspitze und die Legislative) es geht bzw. was genau mit »kompetitiven Wahlen« gemeint ist (insbesondere, dass mehr als eine Partei antritt). Weiterhin sollte die Operationalisierung Indikatoren zur Bestimmung von »Regierungsspitze«, »Legislative« und »Partei« beinhalten. Nach dieser detaillierten Operationalisierung müssen immer noch entsprechende Messinstrumente ausgewählt werden, so zum Beispiel Expertenbeurteilungen oder rechtliche Dokumente. Auf jeden Fall aber muss die Messung sowohl valide (die Daten müssen der Konzeptspezifikation entsprechen) als auch reliabel sein (eine wiederholte Messung desselben Phänomens muss im selben Wert für den Indikator resultieren).

Fallauswahl

Probleme der Fallauswahl und mögliche Verzerrungen (Auswahlbias), die daraus resultieren können, gehören zu den zentralen Themen in Lehrbüchern qualitativer ebenso wie quantitativer Methodik. Genaugenommen muss zwischen Analyseeinheit, Fall und Beobachtung unterschieden werden. Bei der Analyseeinheit handelt es sich um das abstrakte Gebilde, welches analysiert wird (zum Beispiel Staaten, Institutionen, Entscheidungen) und welches häufig durch die Theorie vorgegeben ist. »Fälle« sind die spezifischen Einheiten, die untersucht werden. Ist also die Analyseeinheit »Staat«, dann könnte sich eine Fallstudie mit Schweden oder eine vergleichende Fallstudie mit Schweden und Norwegen befassen. Schließlich kann ein einzelner Fall aus einer einzigen Beobachtung bestehen, wenn lediglich je ein Wert für die abhängige und jede unabhängige Variable erhoben wird. Jedoch kann ein einzelner Fall auch mehrere Beobachtungen enthalten. Eine Forschungsarbeit zum Beitritt der skandinavischen Länder zur Europäischen Union könnte beispielsweise auf einem einzelnen Datensatz pro Land beruhen (mit Werten für unabhängige Variablen wie zum Beispiel BIP pro Kopf, Wachstum und Exportabhängigkeit). Alternativ könnten multiple »Datensatz-Beobachtungen« (data-set observations; Collier u.a. 2004: 252) für einzelne Fälle gemacht werden, etwa durch wiederholte standardisierte Beobachtung der Variablen zu unterschiedlichen Zeitpunkten. Oder wir können auf eine Reihe von »Kausalprozess-Beobachtungen« (causal-process observations) zurückgreifen, um herauszufinden wie strukturelle ökonomische Faktoren zu Entscheidungen über einen EU-Beitritt führten (zum Beispiel Lobbying durch Interessensgruppen und Wahl- oder Referendumsergebnisse).

Um Theorien zu testen, stellt sich die Frage, wie Beobachtungen ausgewählt werden sollen, damit die Ergebnisse der Analyse unverfälscht sind und eine valide Beurteilung der Theorie erlauben. Für die Beschreibung und Erklärung von sozialen Phänomenen ist entscheidend, ob die ausgewählten Beobachtungen die Gruppe der Phänomene angemessen repräsentieren. Ist der 11. September repräsentativ für Terrorismus? Erlaubt eine Studie über Tony Blair allgemeine Schlüsse zur Personalisierung und Medialisierung der Politik in heutigen Demokratien? Bisweilen sind Zufallsstichproben möglich (wie zum Beispiel in Studien zum Wahlverhalten). Doch auch hier ist es möglich, dass die Auswahlprozedur bestimmte Gruppen bevorzugt – beispielsweise Menschen, die viel Zeit zu Hause verbringen und somit eine größere Wahrscheinlichkeit aufweisen, von Umfrage-

instituten angerufen zu werden. Manchmal ist zwar die Grundgesamtheit der Fälle bekannt – wie etwa Demokratien oder post-kommunistische Revolutionen – doch wenn aufgrund eingeschränkter Ressourcen nur wenige Fälle untersucht werden können, würde eine Zufallsstichprobe wahrscheinlich zu verfälschten, das heißt verzerrten Ergebnissen führen. Schließlich kann es vorkommen, dass die Grundgesamtheit nicht bekannt ist. Forscherinnen stehen also vor der Wahl, eine Fallauswahl zu treffen, die entweder unabsichtlich oder absichtlich nicht zufällig ist. Dies muss berücksichtigt werden, um zu validen Generalisierungen und theoretischen Schlüssen zu gelangen.

Kontrolle alternativer Erklärungen

Im Dialog zwischen Theorie und Daten spezifizieren wir Theorien, um die Wirkung der postulierten Ursachen zu testen. Oder wir benutzen (oder entwerfen) eine Theorie, um eine Gruppe von Beobachtungen oder ein spezielles Ereignis zu erklären. Selbst wenn wir eine starke Beziehung zwischen der theoretisierten Ursache und der beobachteten Wirkung finden: Wie können wir sicher sein, dass diese Beziehung keine Scheinbeziehung ist und dass andere kausale Faktoren nicht eine ebenso gute oder gar bessere Erklärungskraft besitzen? Beispielsweise könnte man den demokratischen Frieden auch durch die Hegemonie demokratischer Großmächte oder durch die starke ökonomische Interdependenz zwischen demokratischen Ländern erklären. Kurzum, wir müssen in unserer Forschung stets mögliche alternative Erklärungen berücksichtigen und kontrollieren. Allerdings stellt sich die Frage, welche und wie viele alternative Faktoren oder Variablen einbezogen werden sollten und wie wir entscheiden können, welche der rivalisierenden Theorien oder Ursachen die beste Erklärung bietet.

Theoretische Schlussfolgerungen

Nehmen wir an, eine wohlspezifizierte Theorie wurde mit validen und reliablen Messungen und einer unverfälschten Fallauswahl erfolgreich getestet, und alternative Erklärungen konnten ausgeschlossen werden. In einem solchen Fall kann die Theorie als zunächst bestätigt gelten und muss nicht überarbeitet oder verworfen werden. Leider ist es aber in der Realität oft so, dass Anomalien wie abweichende Fälle oder statistisch unsignifikante

Beziehungen auftreten. Was tun, wenn wir zum Beispiel einen einzelnen Fall finden, in dem Demokratien gegeneinander Krieg führen. Dürften wir ihn einfach ignorieren, oder müssten wir die Theorie des demokratischen Friedens vollständig verwerfen? Oder könnte sie eventuell durch eine verfeinerte Spezifikation oder Einschränkung ihres Geltungsbereiches »gerettet« werden? Jedenfalls lassen sich empirische Ergebnisse nicht automatisch in theoretische Schlussfolgerungen übersetzen. Es handelt sich um einen Abwägungsprozess, bei dem wir unser Wissen weder vorschnell aufgeben noch grundlos schützen dürfen. Zwar scheinen solche Überlegungen auf den ersten Blick nicht zur Thematik des Forschungsdesigns zu gehören, da sie erst aufkommen, nachdem der Forschungsprozess abgeschlossen ist, doch sollte Forschung von Anfang an so gestaltet sein, dass sie es erlaubt, die richtigen Schlüsse aus der Theorie zu ziehen.

Tabelle 1 fasst die wichtigsten Probleme des Forschungsdesigns zusammen. Politikwissenschaftler sollten ein relevantes Forschungsproblem definieren, Konzepte und Theorie klar spezifizieren, Messungen valide und reliabel durchführen, eine Fallauswahl treffen, die valide kausale Schlussfolgerungen und Generalisierungen zulässt, alternative Erklärungen kontrollieren, um Stichhaltigkeit und Überlegenheit der vorgeschlagenen Theorie zu demonstrieren, und durch die richtigen Schlussfolgerungen aus der Analyse zum wissenschaftlichen Fortschritt beitragen. Wie dies erreicht wird oder, bescheidener gefragt, wie man sich diesen Herausforderungen besser stellt, wird Gegenstand der einzelnen Beiträge in diesem Band sein.

<i>Kernprobleme des Forschungsdesigns</i>	<i>Herausforderungen</i>
Forschungsproblem	Relevanz
Konzepte und Theorie	klare Spezifikation
Messung	Validität und Reliabilität
Fallauswahl	valide und generalisierbare Schlussfolgerungen
Kontrolle	Stichhaltigkeit und Überlegenheit von Entscheidungen
Theoretische Schlussfolgerungen	wissenschaftlicher Fortschritt

Tabelle 1: Kernprobleme des Forschungsdesigns und Herausforderungen

Grundtypen des Forschungsdesigns

Im Allgemeinen kann das Forschungsdesign dem konkreten Forschungsproblem individuell angepasst werden. Jedoch gibt es in der Literatur einige grundlegende Typen des Forschungsdesigns, die sich beispielsweise in der Auswahl der Variablen und Fälle, der Verwendung von Daten und Methoden sowie in ihren theoretischen Schlussfolgerungen unterscheiden. In diesem Abschnitt bieten wir einen Überblick über diese Typen, denen die nachfolgenden Beiträge genauer nachgehen werden. Eine zentrale Unterscheidung ist die zwischen X-zentrierten und Y-zentrierten Forschungsdesigns; die andere die zwischen Designs mit großer und mit kleiner Fallzahl.

X- und Y-zentrierte Forschungsdesigns

Die Terminologie von X- und Y-zentrierten Forschungsdesigns geht auf Steffen Ganghof (2005a) zurück, doch sind ähnliche Unterscheidungen auch von anderen Autoren getroffen worden.⁶ Hier werden Forschungsdesigns unterschieden nach der Orientierung der Forschung bzw. nach der Art der kausalen Schlüsse, die eine Forscherin zieht, um ihre Forschungsfrage zu beantworten. Die Unterscheidung folgt der gängigen Bezeichnung von unabhängigen, erklärenden Variablen mit X und der abhängigen, zu erklärenden Variable mit Y. Das Ziel der Forschung kann einerseits sein, Evidenz für spezifische kausale Faktoren oder Mechanismen (X) zu liefern, oder andererseits spezifische Ergebnisse (Y) so genau und vollständig wie möglich zu erklären. Der Wissenschaftler kann beispielsweise besonders daran interessiert sein, wie und unter welchen Bedingungen (X) die Wähler ihre Entscheidungen treffen. Oder aber es geht ihm einfach darum, Wahlergebnisse (Y) vorherzusagen.

Ein Forschungsdesign wird *X-zentriert* genannt, wenn das Hauptinteresse auf der Erklärungskraft kausaler Faktoren liegt. Hier ist es das Ziel, die Richtung, Stärke und Robustheit der kausalen Wirkung einer oder einiger weniger unabhängiger Variablen, X_i ($i = 1, \dots, n$) auf eine abhängige Variable, Y zu bestimmen. Unabhängige Variablen sind hier entweder erklärende oder Testvariablen, die von zentraler Bedeutung für den kausalen Effekt sind, oder aber Kontrollvariablen, die dazu dienen, sich zu versichern, dass der kausale Effekt wirklich der Testvariable entspringt, und

⁶ Siehe George/Bennett (1997; 2005); Gerring (2001: 137); Scharpf (1997: 24–27).

nicht dem Wirken anderer kausaler Faktoren entspricht. Typische Forschungsfragen für X-zentrierte Forschungsdesigns sind: Verursacht X_i Y? Welchen Effekt hat X_i auf Y and wie gross ist dieser? Wenn also das Forschungsinteresse darauf gerichtet ist, wie Parteipräferenzen die Entscheidungen einer Wählerin beeinflussen, sollten alternative Erklärungen von Wahlentscheidungen – wie zum Beispiel Einstellungen zu politischen Sachfragen und Präferenzen für Kandidaten – einbezogen werden, um deren potenziellen Einfluss von dem der Testvariablen unterscheiden zu können.

Hingegen ist ein Forschungsdesign Y-zentriert, wenn das primäre Interesse darin besteht, Ereignisse und Ergebnisse zu erklären. Ziel ist es in diesem Falle, zahlreiche Einflussfaktoren X_i ($i = 1, \dots, n$) in Betracht zu ziehen, die in ihrer Gesamtheit die Varianz der abhängigen Variable, Y, so vollständig wie möglich erklären. Ein Beispiel für eine Y-zentrierte Strategie ist die Suche nach einer umfassenden Erklärung für den ungleichen Erfolg von UN-Friedenssicherungsmissionen oder die uneinheitliche Umsetzung von EG-Richtlinien in den Mitgliedsländern. Y-zentrierte Forschung kann sich aber auch mit einzelnen Ereignissen befassen (die iranische Revolution oder das Ende des kalten Krieges – mit anderen Worten, eine abhängige Variable ohne Varianz). Typische Forschungsfragen dieser Art von Forschung sind: Was verursacht Y? Warum Y? Wenn man also individuelles Wahlverhalten voraussagen möchte, sollte dementsprechend ein Y-zentriertes Forschungsdesign gewählt werden, und es sollten weitere unabhängige Variablen (zum Beispiel Kontext- oder Medieneffekte) miteinbezogen werden, um genauere Vorhersagen treffen zu können (auch wenn diese zur Kontrolle der zentralen Erklärungsvariablen Parteipräferenz, auf die sich ein X-zentriertes Design konzentrieren würde, nicht unbedingt notwendig wären).

Welche Gründe sprechen für die Wahl eines X- oder Y-zentrierten Designs? Nach unserer Auffassung hängt diese Wahl in erster Linie vom Interesse und den Relevanzkriterien der Forscherinnen ab. Wer vor allem daran interessiert ist, wichtige politische Ereignisse (wie Kriege oder Revolutionen) zu erklären oder die Ergebnisse spezifischer politischer Entscheidungen vorherzusagen (wie die Bildung einer Regierungskoalition), wird ein Y-zentriertes Design vorziehen.

Falls die Forschung aber durch ein theoretisches Interesse an kausalen Faktoren (wie zum Beispiel Ressourcen oder Institutionen) oder Mechanismen (wie zum Beispiel politische Sozialisation oder politische Dilemmas) motiviert wird, sind Y-zentrierte Forschungsdesigns in der Regel un-

geeignet. Vielmehr bietet sich ein X-zentriertes Design an. Für ein X-zentriertes Design ist es von zentraler Bedeutung, den Einfluss von Störvariablen zu berücksichtigen und diese Variablen konstant zu halten bzw. sie zu kontrollieren, um den Einfluss des kausalen Faktors oder Mechanismus bestimmen zu können, dem das primäre Interesse gilt. Dies kann mittels verschiedener Strategien erreicht werden, die das kausale Netz entwirren. Kontrollvariablen in Regressionsgleichungen einfügen, Abgleichmethoden oder (Feld-)Experimente sind hier potenzielle Lösungsansätze, die jedoch zahlreiche Beobachtungen voraussetzen. Andererseits sind Strategien möglich, bei denen nur wenige Beobachtungen gebraucht werden. Der Vergleich von sorgfältig ausgewählten Fällen gehört ebenso dazu wie Quasi-Experimente, in denen man denselben Fall vor und nach dem Auftreten eines kausalen Faktors (zum Beispiel eine Veränderung des institutionellen Regimes oder eine Policy-Intervention) vergleicht (George/Bennett 2005: Kapitel 8).

Die Wahl zwischen X- oder Y-zentrierten Forschungsdesigns hängt hingegen nicht unbedingt davon ab, wie weit fortgeschritten die Theorieentwicklung in einem Gebiet ist, auch wenn in »neueren« Feldern oft auf Y-zentrierte Designs zurückgegriffen wird. Diese versuchen das neue Phänomen direkt durch Varianz in der abhängigen Variable zu erschließen. Allerdings sind gute Gründe vorgebracht worden, warum es nicht immer vorteilhaft ist, in frühen Stadien eines Forschungsprojekts theoretische und empirische Erkenntnisse durch eine umfassende Erklärung eines Phänomens anhand von Maximierung der Varianz der abhängigen Variable erlangen zu wollen (zum Beispiel Geddes 2003: Kapitel 2; King u.a. 1994: 169, Fussnote 8). Eine solche Strategie ist oft unausführbar aufgrund von Problemen bei der Datenbeschaffung; sie kann zu ungenauen Schätzungen von kausalen Effekten führen und theoretischen Fortschritt erschweren, der durch die Verwendung bestehender Theorien möglich wäre.

Eine andere Möglichkeit erschließt sich durch ein X-zentriertes Design, bei dem die Forscherin die Erklärung in leichter handhabbare Theoriebausteine unterteilt, für jeden dieser Bausteine die relevanten Variablen identifiziert, und abschließend die Theoriebausteine wieder zusammenfügt. Ob allerdings eine solche »Lego-Strategie« (Pierson/Skocpol 2002: 717) die Akkumulation von Wissen wirklich vereinfacht, kann bezweifelt werden. Kritiker wenden ein, dass die Beantwortung großer und relevanter Fragen zugunsten robuster Antworten auf kleine und potenziell triviale Fragen verdrängt wird (Pierson/Skocpol 2002: 713–718).

Auch wenn der Stand der Theorieentwicklung in einem Forschungsgebiet ein bestimmtes Forschungsdesign nahelegen scheint, ist die Wahl eines X- oder Y-zentrierten Designs alles andere als zwangsläufig. Selbst in theoretisch weit entwickelten Gebieten entscheiden sich Wissenschaftler nicht ausschliesslich für X-zentrierte Designs, auch wenn es leichter sein dürfte, Kausalfaktoren zu isolieren und den Fokus auf Richtung und Stärke des Effekts zu legen. Wenn Wissenschaftler zum Beispiel Ereignisse in der Zukunft wie Wahlen oder den Zusammenbruch von Staaten voraussagen möchten, führt kein Weg an Y-zentrierten Designs vorbei.

Forschungsdesigns mit vielen und wenigen Fällen

Wie gross ist dein »n«? Eine der am meisten verbreiteten Klassifizierungen von Forschungsdesigns bezieht sich auf die Anzahl der Fälle und Beobachtungen, die einer Studie zugrunde liegen. Designs mit kleiner und großer Fallzahl unterscheiden sich darin, wie sie empirische Informationen einsetzen. Studien mit vielen Fällen werden für gewöhnlich mit statistischen Tests von Korrelationen gleichgesetzt, folgen einem probabilistischen Modell von Kausalität und verwenden Datensatz-Beobachtungen, also »Beobachtungen, die als eine Matrix von Werten für einzelne Variablen und für ein bestimmtes Sample von Fällen gesammelt wurden« (Brady u.a. 2004: 12; eigene Übersetzung).

Studien mit kleiner Fallzahl werden hingegen normalerweise mit Fallstudien und Vergleichsstudien (George/Bennett 2005) sowie der Auswertung multipler kausaler Prozess-Beobachtungen für einen einzelnen Fall in Verbindung gebracht. Einzelfallstudien beruhen auf einer Prozessanalyse (process-tracing, Schimmelfennig 2006), um die vermuteten kausalen Mechanismen zu entdecken und zu überprüfen, die die zu erklärenden Zusammenhänge und Phänomene hervorgebracht haben (zum Beispiel George/Bennett 2005: 147–149). Andernfalls beruhen kausale Schlüsse auf dem systematischen, kontrollierten Vergleich einer kleinen Zahl von Fällen (wie etwa George/Bennett 2005: Kapitel 8).

Kurzum, Studien mit großer Fallzahl zielen darauf, die Validität von kausalen Schlüssen durch die Vermehrung von Fällen und Beobachtungen zu erhöhen, während Studien mit kleiner Fallzahl das gleiche Ziel durch einen systematischen Vergleich weniger Fälle und die Vermehrung von Beobachtungen des kausalen Prozesses verfolgen. Designs mit kleiner Fallzahl geben der Tiefe der Analyse gegenüber der Breite den Vorzug, Designs mit

großer Fallzahl setzen auf Breite statt Tiefe. Folglich führt Forschung mit kleiner Fallzahl potenziell zu sehr präzisen und dichten kausalen Analyse eines oder weniger Fälle auf Kosten der Verallgemeinerbarkeit der Ergebnisse. Forschung mit großer Fallzahl stärkt hingegen unser Vertrauen in die Allgemeingültigkeit und die durchschnittliche Stärke kausaler Effekte, wobei allerdings individuelle Fälle und ihr Ursachengefüge weitgehend »unsichtbar« werden (Ragin 2000: 31).

Wie sollte man zwischen Designs mit großer und kleiner Fallzahl auswählen? Ein grundlegendes Prinzip lautet, dass bessere Methoden der Datenerhebung besseren Methoden der Analyse vorzuziehen sind. Demnach ist es die hohe Kunst des Forschungsdesigns, bestehende Informationen geschickt zu verarbeiten oder neue Informationen zu sammeln und gut darüber nachzudenken, wie weitere Datenquellen erschlossen und genutzt werden können. Für gewöhnlich kommen Designs mit vielen Fällen immer dann zum Einsatz, wenn ausreichend quantifizierbare und vergleichbare Informationen erhältlich sind. Aber Vorsicht! Die Zahl der Beobachtungen zu erhöhen hat seinen Preis, auch wenn sie leicht verfügbar sind. Können die neuen Informationen wirklich mit den alten verglichen werden? Muss ich die verwendeten Konzepte zu sehr »dehnen«, um Vergleichbarkeit herzustellen? Passen die Indikatoren auf die neuen Fälle? Kurzum, zusätzliche Beobachtungen können zu Lasten der Qualität der Ergebnisse gehen. Es kann jedoch keineswegs schaden, Kausalprozess-Beobachtungen zur Unterstützung von Analyseergebnissen heranzuziehen, die aus einem bestehenden Datensatz gewonnen wurden. Diese Auffassung wird sogar von eingefleischten »Quantis« geteilt (vgl. zum Beispiel Goldthorpe 2001; Beck 2006).

Die Kluft zwischen Forschung mit vielen und wenigen Fällen ist nicht theoretischer Natur, sondern zeigt sich auch in der Forschungspraxis (siehe Bollen u.a. 1993: 327; Ragin 2000: 25). Auf der einen Seite sehen wir Einzelfallstudien oder Studien mit einer Handvoll Fällen, während sich auf der anderen Seite Studien mit 50 oder mehr Fällen und in der Umfrageforschung sogar Tausenden von Beobachtungen häufen. Knappe Ressourcen oder Zeithorizonte führen dazu, dass Forscher typischerweise entweder auf die detaillierte Analyse weniger Fälle abzielen oder aber auf die Breite und Verallgemeinerbarkeit der Ergebnisse von Studien mit großer Fallzahl setzen. Studien, die zwischen diesen zwei Polen angesiedelt sind, haben zwischen 10 und 50 Beobachtungen. Bei solchen Studien ist es weniger klar, ob der Fokus auf Breite oder Tiefe liegen sollte. Zum einen gibt es

durchaus quantitative Wahlvorhersagemodelle mit weniger als 15 Beobachtungen, die trotzdem die Methode der statistischen Kontrolle benutzen (zum Beispiel Bartels/Zaller 2001; Lewis-Beck/Rice 1992; Gschwend/Norpoth 2005). Zum anderen können qualitativ-vergleichende Methoden wie QCA oder Fuzzy-set-Analyse (Ragin 1987; Ragin 2000; Schneider 2006) auch mit Dutzenden von Beobachtungen arbeiten.

Letztendlich sind alle Wissenschaftlerinnen daran interessiert herauszufinden, warum etwas passiert, um entsprechende Erklärungen zu entwickeln und unser Verständnis von Ursache-Wirkungs-Beziehungen in der sozialen Welt zu verbessern. Jedoch ist es in der Literatur umstritten, wie Kausalität konzeptualisiert werden sollte. Forschung mit wenigen Fällen ist üblicherweise als Analyse von notwendigen und hinreichenden Bedingungen konzipiert. Diese Sicht setzt, zumindest implizit, ein deterministisches (und nicht-lineares) Verständnis von Kausalität voraus. Forschung mit vielen Fällen (wie auch manche Studien mit kleiner Fallzahl) setzt dagegen auf ein probabilistisches Verständnis von Kausalität, in welchem Ursachen als Faktoren gesehen werden, die die Wahrscheinlichkeit des Eintreffens eines Ereignisses vergrößern (Gerring 2001: 129). Natürlich sind deterministische Ursachen besonders hilfreich, wenn davon ausgegangen werden kann, dass die Beziehung zwischen unabhängiger und abhängiger Variable tatsächlich deterministisch ist. In diesem Fall hilft die Unterscheidung von notwendigen und hinreichenden Bedingungen dabei, das kausale Feld zu ordnen und klare Erwartungen zu formulieren, welche Beobachtungen wir machen oder nicht machen sollten, falls die Theorie stimmt.

Aber können wir wirklich sicher sein, dass wir es in der Politikwissenschaft mit deterministischen Ursachen zu tun haben? So kann zum einen der Gegenstand unserer Analyse selbst zu einem gewissen Grad zufällig sein, sodass die Variation in der abhängigen Variable sogar dann niemals komplett erklärt werden kann, wenn alle Konzepte präzise gemessen wurden. Dies würde auch dann gelten, wenn nicht nur alle nur erdenklichen Variablen miteinbezogen wären, sondern auch das potenziell nicht-lineare Modell korrekt spezifiziert wäre. Zum anderen besteht natürlich das Problem, dass keine Messung perfekt ist. Es wird zwangsläufig Messfehler geben. Demnach führt schon der Prozess der Messung eines theoretischen Konzepts, auch wenn wir glauben in einer deterministischen Welt zu leben, zu einem gewissen Grad an Zufälligkeit in der Analyse (Broscheid/Gschwend 2005).

Diese Debatte ist nicht nur aus wissenschaftsphilosophischer Sicht relevant, sie bringt auch wichtige Implikationen für das Forschungsdesign und Interpretation von Ergebnissen mit sich. Wie soll zum Beispiel mit einzelnen Fällen oder Beobachtungen umgegangen werden, die beträchtlich von einem ansonsten klaren kausalen Muster abweichen? Falls man von einer deterministischen kausalen Welt mit perfekten Messungen und korrekt spezifizierten Modellen von Beziehungen überzeugt ist, stellt dies ein ernsthaftes Problem dar. Anhänger probabilistischer Kausalbeziehungen werden abweichende Fälle hingegen schlicht als Ausreißer betrachten. Doch auch bei einem deterministischen Verständnis von Kausalität und perfekt spezifizierten theoretischen Erklärungen können Ausreißer vorkommen, weil womöglich einzelne Bausteine der Theorie ungenau konzeptualisiert wurden oder sich beim Messen zufällige Fehler einschleichen. Solche Ausreißer machen deswegen die Hypothese nicht automatisch ungültig.

Unabhängig von der Zahl der Beobachtungen kann in verschiedenster Weise mit Ausreißern umgegangen werden. Eine Möglichkeit ist zu argumentieren, dass das Modell korrekt spezifiziert ist und Beobachtungen nur wegen Messfehlern vom generellen Muster abweichen. Eine andere Möglichkeit ist es, Ausreißer dazu benutzen, die Theorie zu überdenken. In diesem Falle sind Ausreißer von substanziellem Interesse und werden nicht als Folge von ungenauer Messung verstanden. Beim Überdenken einer Theorie liegt es nahe, die Reichweite bzw. die Bedingungen der Gültigkeit der Theorie festzulegen (Ragin 2000: 61–62; Walker/Cohen 1985) Vielleicht gilt die Theorie nur für eine Teilstichprobe – angenommen die Beobachtungen sind prinzipiell vergleichbar (Achen 2002: 446–447). Eine weitere Möglichkeit besteht darin, alle Beobachtungen beizubehalten und die erwartete universelle kausale Beziehung so zu modifizieren, dass Interaktionseffekte oder nicht-lineare Transformationen von unabhängigen Variablen miteinbezogen werden. Hiermit würde man konditionale oder nicht-lineare Effekte von unabhängigen Variablen auf die abhängige Variable nachweisen. Weiterhin kann das Hinzufügen unabhängiger Variablen nützlich sein, um abweichende Beobachtungen besser zu erklären. Wahrscheinlich aufgrund der Anzahl der Fälle – weil ein einzelner Ausreißer mehr Gewicht hat, wenn die Zahl der Beobachtungen bei fünf anstelle von 5.000 liegt – tendieren Forscherinnen, die sich mit wenigen Fällen beschäftigen, dazu sich intensiv mit Ausreißern auseinanderzusetzen, während Wissenschaftler, die hohe Fallzahlen bearbeiten, eher dazu neigen, dass Prob-

lem statistisch zu lösen, wenn es denn überhaupt als ein solches gesehen wird (vgl. Western 1995).

Dennoch sind viele wichtige Theorien als Konstellationen von notwendigen und hinreichenden Bedingungen formuliert (siehe zum Beispiel Dion 1998; Goertz/Starr 2003; Seawright 2002). Um die Kluft zwischen deterministischen und probabilistischen Weltanschauungen zu überbrücken, haben neuere methodische Ansätze Tests entwickelt und schätzen Modelle, die auf notwendigen und hinreichenden Bedingungen basieren – zum Teil mit Bayesianischen Methoden, um nicht bereits technische Probleme mit kleinen Fallzahlen zu bekommen (Braumoeller/Goertz 2000; Braumoeller 2003; Clark u.a. 2006; Seawright 2002). Also ist die Wahl zwischen Forschung mit kleiner und großer Fallzahl zumindest teilweise unabhängig davon, ob statistische Tests benutzt werden, Datensatz- oder Kausalprozess-Beobachtungen zum Tragen kommen, oder man ein probabilistisches oder deterministisches Verständnis von Kausalität besitzt.

Es lässt sich folglich eine zweidimensionale Konzeptualisierung von Forschungsdesigns entwerfen, die über die altbekannte Dichotomie von qualitativer und quantitativer Forschung hinausgeht, wie sie durch die Hauptdiagonale in Tabelle 2 repräsentiert wird: X-zentrierte Forschungsdesigns benutzen die Methode der statistischen Kontrolle oder (Feld-) Experimente, um einen bestimmten kausalen Faktor zu bestimmen, wenn viele Fälle zur Verfügung stehenden. Y-zentrierte Forschung liefert eine tiefgehende, detaillierte Untersuchung von möglichen Faktoren und kausalen Prozessen innerhalb einer Fallstudie, um einzelne Ereignisse so genau und umfassend wie möglich zu erklären. Doch sind die Felder jenseits der Hauptdiagonalen keineswegs leer. Einerseits können fokussierte Kreuzvergleiche oder Quasi-Experimente in X-zentrierten Forschungsdesigns auch dann benutzt werden, wenn nur wenige Beobachtungen zur Verfügung stehen. Andererseits gibt es Y-zentrierte Forschungsdesigns mit hoher Beobachtungszahl, die die Möglichkeit bieten, anhand von statistischen und qualitativ-vergleichenden Methoden Phänomene zu beschreiben und vorherzusagen.

		<i>Typ kausaler Schlüsse</i>	
		<i>X-zentriert</i>	<i>Y-zentriert</i>
<i>Fallzahl</i>	<i>Groß</i>	Statistische Kontrolle, (Feld-) Experimente	Prognose, qualitative-vergleichende Methoden
	<i>Klein</i>	Kreuzvergleiche, Quasi-Experimente	Fallstudien

Tabelle 2: Grundtypen von Forschungsdesigns

Kapitelübersicht

In den vorhergehenden Abschnitten haben wir verschiedene Kernprobleme und Grundtypen von Forschungsdesigns vorgestellt. Um aber unser Werkzeug zur Entwicklung von Forschungsprojekten bestmöglich nutzen zu können, müssen wir wissen, ob die Kernprobleme für alle Grundtypen von Forschungsdesign die gleichen sind. Die Literatur in diesem Zusammenhang ist zweigeteilt. Die Einen argumentieren, dass qualitative Forschung die Grundsätze der quantitativen Forschung befolgen sollte, um gute Forschung zu praktizieren (King u.a. 1994). Dies ist »quantitativer Imperialismus«. Andere vertreten dagegen die Meinung, quantitative und qualitative Forschung unterlägen grundsätzlich verschiedenen Forschungslogiken (siehe zum Beispiel McKeown 1999; Thomas 2005). Demnach könnten quantitative und qualitative Forschungsdesigns – seien sie X- oder Y-zentriert, mit großer oder kleiner Fallzahl – weder mit einander kommunizieren noch von einander lernen. Dies ist »qualitativer Separatismus«. (Natürlich wäre auch qualitativer Imperialismus oder ein quantitativer Separatismus denkbar, allerdings scheinen solche Positionen in der aktuellen Politikwissenschaft selten zu sein.)

Dieses Buch geht von der Annahme aus, dass Forschung grundsätzlich aus einem Dialog zwischen Theorie und Daten besteht, und dass alle Arten von Forschungsdesign sich mit denselben Problemen und Herausforderungen auseinandersetzen müssen. Ob sich für alle Arten von Forschungsdesign auch die gleichen Lösungen anbieten, ist aber zunächst eine offene Frage, die in den folgenden Kapiteln näher betrachtet wird. Alle Kapitel

folgen dem gleichen Muster. Sie beginnen mit einem spezifischen Problem des Forschungsdesigns in der Politikwissenschaft. Im zweiten Teil beschreiben sie das Problem und diskutieren verschiedene Lösungen mitsamt ihren Stärken und Schwächen. Drittens werden praktische Hinweise für den Umgang mit dem Design-Problem in der konkreten Forschung vorgestellt. Im vierten Teil werden diese Hinweise dann anhand eines Beispiels aus der Forschung der Autorinnen und Autoren illustriert.

In ihrem Kapitel zur Relevanz von Forschungsfragen definieren und unterscheiden *Matthias Lehnert*, *Bernhard Miller* und *Arndt Wonka* theoretische und gesellschaftliche Relevanz. Das Kapitel konzentriert sich dann auf die weithin vernachlässigte gesellschaftliche Relevanz von Forschungsdesigns. Lehnert, Miller und Wonka bestreiten, dass sich gesellschaftliche und theoretische Relevanz wechselseitig ausschließen oder beeinträchtigen und zeigen, wie die Relevanz jedes Forschungsprojektes durch die Beantwortung von drei Fragen gesteigert werden kann. Wer ist wovon betroffen? Wie können die Effekte bewertet werden? Welche Ratschläge können wir geben?

»Weißt Du wirklich, wovon Du sprichst?«. Um diese Frage geht es in *Arndt Wonkas* Kapitel zu Konzeptspezifikation – einem zentralen Thema von Forschungsdesigns in den Politikwissenschaften. Auch wenn nicht selten in der Literatur unterschiedliche definierende Attribute für ein und dasselbe Konzept benutzt werden, behauptet Wonka, dass mehrdeutige Konzepte Forschungsergebnisse entwerten. Wonka gibt praktische Hinweise, wie solche Mehrdeutigkeit vermieden werden kann, und wendet sie auf das Konzept der »Supranationalität« aus der EU-Forschung an.

Im Kapitel zu Typologien in der sozialwissenschaftlichen Analyse befasst sich *Matthias Lehnert* mit einem Spezialfall der Konzeptspezifikation. Er durchleuchtet kritisch, ob und wie Typologien für die Beschreibung und Erklärung von sozialen Phänomenen benutzt werden können. Mittels dreier Kriterien, mit denen »Typen von Typologien« unterschieden werden können, grenzt er die Nutzbarkeit von Typologien für spezielle Zwecke ein. Typologien bieten vereinfachte Beschreibungen komplexer Phänomene und können dabei helfen, die Homogenität der Untersuchungseinheiten in X- und Y-zentrierten Forschungsdesigns herzustellen. Neben allgemeinen praktischen Hinweisen zur Verwendung von Typologien in der Politikwissenschaft zeigt Lehnert anhand der Literatur und eigener Forschung, wie Typologien in der Analyse politischer Institutionen und deren Folgen sinnvoll eingesetzt werden können.

Messung ist das Thema von *Bernhard Millers* Kapitel und der logische nächste Schritt nach der Konzeptspezifikation. Miller spricht die zentralen Herausforderungen der Reliabilität und Validität von Messungen an und diskutiert Möglichkeiten, mit diesen Herausforderungen umzugehen. Sein Augenmerk liegt dabei auf der Nutzung von Indizes als zusammengesetzten Maßen. Das Kapitel betont die zentrale Rolle von Theorie und Konzepten in allen Messvorgängen, unabhängig vom gewählten Forschungsdesign. Miller erörtert die Abwägungen im Messprozess, die man in der Forschungspraxis unweigerlich treffen muss, und gibt klare Hinweise, wie neue Maße konzipiert werden können. Dies illustriert er anhand seiner Forschung zu Koalitionsausschüssen.

Im Kapitel von *Julia Rathke* geht es um Probleme der Vergleichbarkeit und Äquivalenz von Messungen, denen sekundäre Datenquellen zugrunde liegen. Nach ihrer Auffassung ist eine vergrößerte Zahl von Beobachtungen nicht die Lösung *per se*, sondern benötigt zumindest konzeptuell äquivalente Messungen. Rathke unterscheidet dabei zwischen zwei unterschiedlichen Strategien: Erhöhung des Abstraktionsniveaus und funktionale Äquivalenz. Nach praktischen Hinweisen dazu, wie Daten und Indikatoren vergleichbar gemacht werden können, demonstriert Rathke diese anhand ihrer Forschung zu den Wirkungen von Sozialkapital auf politische Orientierungen in Deutschland.

Mehrere Kapitel befassen sich mit Problemen der Fallauswahl. In ihrer Diskussion zum Auswahlbias in Forschungsdesigns mit großer Fallzahl befasst sich *Janina Thiem* mit einer Herausforderung, der sich die quantitative Forschung oft gegenüberstellt: das Universum der Fälle – sprich die zu analysierende Population – ist sehr gross und theoretisch gut definiert, aber nur teilweise beobachtbar. Wenn nun die nicht-beobachtbaren Fälle nicht zufallsverteilt sind, ist jede kausale Inferenz aus der Stichprobe verfälscht. Es existieren zwar altbekannte statistische Lösungen für solche Situationen, die mehr oder weniger hilfreich für das jeweilige Forschungsproblem sein können, doch Thiem argumentiert, dass das Problem in erster Linie theoretischer Natur ist. Sie gibt praktische Hinweise, wie solche Probleme der Fallauswahl identifiziert und theoretisch wie statistisch angegangen werden können, und wendet sie schliesslich auf den Auswahlbias bei der Analyse von namentlichen Abstimmungen im Europäischen Parlament an.

Dirk Leuffens Kapitel zu Fallauswahl und Auswahlbias in Forschungsdesigns mit kleinen Fallzahlen behandelt eine für qualitative Forscher wohlbekannte Situation: Das Universum der Fälle ist gross, aber wenig bekannt

oder schlecht definiert. Nach einer Diskussion von Mills klassischen Differenz- und Konkordanzmethoden und den gleichermassen bekannten Forschungsdesigns mit möglichst ähnlichen und möglichst verschiedenen Systemen (*most-similar systems* und *most-dissimilar systems*), stellt *Leuffen* theoriegeleitete Typologien als Strategie der Fallauswahl vor. *Leuffen* argumentiert, dass es von großem Nutzen sein kann, den Untersuchungsbereich einzugrenzen, wenige Zellen der Typologie herauszugreifen und sich auf die »harten Fälle« zu konzentrieren. Illustriert wird die Strategie anhand eines Beispiels seiner Forschung über die *cohabitation* in Frankreich.

Christoph Hönniges Kapitel setzt die Diskussion über möglichst ähnliche beziehungsweise möglichst verschiedene Systeme als typische quasi-experimentelle Forschungsdesigns fort und hebt sie als einen Königsweg zwischen Forschungsdesigns mit kleiner und großer Fallzahl heraus. Allerdings stellt die bestehende Literatur wenige Hilfestellungen bereit, welche Fälle denn nun eigentlich ausgewählt werden sollen. Immerhin birgt die intentionale Auswahl von Fällen in Forschungsdesigns mit kleinen Fallzahlen das erhebliche Risiko verfälschte Rückschlüsse zu ziehen. Das Kapitel zeigt Strategien der gezielten Fallauswahl auf, die möglichst geringe Verzerrungen beinhalten. *Hönnige* erläutert diese Strategien anhand der Fallauswahl zu einem eigenen komparativen Forschungsprojekt zum Klageverhalten der Opposition vor dem jeweiligen Verfassungsgericht.

Das Kapitel von *Ulrich Sieberer* befasst sich mit dem Thema der Kontrolle und diskutiert einige grundlegende theoretische und methodische Möglichkeiten bei der Auswahl von unabhängigen Variablen samt ihren Vor- und Nachteilen. *Sieberer* argumentiert, dass der Status von unabhängigen Variablen als Kontrollvariablen stark davon abhängt, ob es sich um X- oder Y-zentrierte Forschung handelt, aber unabhängig davon ist, ob die Forschung mit vielen oder wenigen Fällen operiert. *Sieberer* illustriert seine praktischen Hinweise zur Kontrolle anhand seiner Arbeiten zur Erklärung innerfraktioneller Geschlossenheit in parlamentarischen Abstimmungen.

Andreas Dür nimmt sich der Herausforderung rivalisierender Erklärungen in Y-zentrierter qualitativer Forschung an. Hierbei unterscheidet er drei Probleme – Verfälschung durch unberücksichtigte Variablen, die Überdeterminierung der Erklärung, sowie Unbestimmtheit – und stellt verschiedene Strategien zum Umgang mit diesen Problemen vor: die Entdeckung von logischen Inkonsistenzen in alternativen Erklärungen, die Untersuchung von kausalen Mechanismen durch Prozessanalyse und die Auswahl von zusätzlichen »höchst wahrscheinlichen« oder »höchst unwahr-

scheinlichen« Fällen. Dür diskutiert die Stärken und Schwächen dieser Strategien und illustriert sie anhand seiner eigenen Forschung zur Liberalisierung des internationalen Handels.

Der Beitrag von *Dirk De Bièvre* befasst sich mit der Schlussphase des Forschungsprozesses: Wie soll mit einer Theorie umgegangen werden, die empirisch getestet und für unzulänglich befunden wurde? De Bièvre vertritt ein theoretisches Verständnis von Falsifizierung, welches darauf basiert, fehlerhafte Hypothesen durch neuere, mutmaßlich Bessere zu ersetzen, und entwirft Vorschläge, um Hypothesen so zu formulieren und den Forschungsprozess so zu organisieren, dass die Forscherin einen größtmöglichen Nutzen aus der Falsifizierung schlagen kann. De Bièvre rekurriert dabei auf sein eigenes Forschungsprojekt über die Wirkungen der Verrechtlichung in der Welthandelsorganisation (WTO).

Das Schlusskapitel fasst zusammen, was wir aus den Beiträgen lernen können, um den Dialog zwischen Theorie und Daten zu verbessern. Wir stellen fest, dass zwar alle Forschung vor die gleichen Herausforderungen und Probleme gestellt ist, es aber kein Universalrezept zu der Lösung gibt. Stattdessen bieten und erfordern unterschiedliche Forschungsdesigns unterschiedliche Lösungsstrategien, wobei jede Strategie gewisse Vor- und Nachteile mit sich bringt. Die Wahl des Forschungsdesign sollte sich folglich daraus ergeben, wie diese Vor- und Nachteile für das konkrete Forschungsvorhaben bewertet werden.

Literatur

- Achen, Christopher H. (2002), »Toward a new political methodology: micro foundations and ART«, *Annual Review of Political Science*, Jg. 5, S. 423–450.
- Alvarez, Mike/Cheibub José A./Limongi, Fernando/Przeworski, Adam (1996), »Classifying political regimes«, *Studies in Comparative International Development*, Jg. 31, H. 2, S. 3–36.
- Bartels, Larry M./Zaller John (2001), »Presidential vote models: A recount«, *PS: Political Science and Politics*, Jg. 34, H. 1, S. 9–20.
- Beck, Nathaniel (2006), *Causal Process »Observation«: Oxymoron or Old Wine*, Manuscript, New York, NY. Erhältlich unter:
<http://www.nyu.edu/gsas/dept/politics/faculty/beck/cponew.pdf>
- Bollen, Kenneth/Entwisle, Barbara/Alderson, Arthur S. (1993), »Macrocomparative research methods«, *Annual Review of Sociology*, Jg. 19, S. 321–351.

- Brady, Henry/Collier, David/Seawright, Jason (2004), »Refocusing the discussion of methodology«, in Henry Brady/David Collier (Hg.), *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, Boulder, S. 3–20.
- Braumoeller, Bear F. (2003), »Causal complexity and the study of politics«, *Political Analysis*, Jg. 11, H. 3, S. 209–233.
- Braumoeller, Bear F./Goertz, Gary (2000), »The methodology of necessary conditions«, *American Journal of Political Science*, Jg. 44, H. 4, S. 844–858.
- Broscheid, Andreas/Gschwend, Thomas (2005), »Zur statistischen Analyse von Vollerhebungen«, *Politische Vierteljahresschrift*, Jg. 46, H. 1, 16–26.
- Clark, William R./Gilligan, Michael J./Golder, Matt (2006), »A simple multivariate test for asymmetric hypotheses«, *Political Analysis*, Jg. 14, H. 3, S. 311–331.
- Coleman, James S. (1990), *Foundations of Social Theory*, Cambridge.
- Collier, David/Brady Henry E./Seawright, Jason (2004), »Sources of leverage in causal inference: toward an alternative view of methodology?«, in: Henry E. Brady/David Collier (Hg.), *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, Lanham, S. 229–266.
- De Vaus, David A. (2001), *Research design in social research*, London.
- Dion, Douglas (1998), »Evidence and inference in the comparative case study«, *Comparative Politics*, Jg. 30, H. 2, S. 127–145.
- Ganghof, Steffen (2005a), »Kausale Perspektiven in der vergleichenden Politikwissenschaft: X-zentrierte und Y-zentrierte Forschungsdesigns«, in: Sabine Kropp/Michael Minkenberg (Hg.), *Vergleichen in der Politikwissenschaft*, Wiesbaden, S. 76–93.
- Geddes, Barbara (2003), *Paradigms and Sand Castles: Theory Building and Research Design in Comparative Politics*, Ann Arbor, MI.
- George, Alexander L./Bennett, Andrew (1997), *Process Tracing in Case Study Research*.
Erhältlich unter:
<http://web.mit.edu/17.878/www/Bennett/Process%20Tracing%20in%20CS%20Research.htm>
- (2005), *Case Studies and Theory Development in the Social Sciences*, Cambridge, MA.
- Gerring, John (2001), *Social Science Methodology. A Critical Framework*, Cambridge.
- Gerring, John/Yesnowitz, Joshua (2006), »A normative turn in political science?«, *Polity*, Jg. 38, H. 1, S. 101–133.
- Goertz, Gary/Starr, Harvey (Hg.) (2003), *Necessary conditions: Theory, Methodology, and Applications*, Lanham, MD.
- Goldthorpe, John H. (2001), »Causation, statistics, and sociology«, *European Sociological Review*, Jg. 17, H. 1, S. 1–20.
- Gschwend, Thomas/Norpoth, Helmut (2005), »Prognosemodell auf dem Prüfstand: Die Bundestagswahl 2005«, *Politische Vierteljahresschrift*, Jg. 46, H. 4, S. 682–688.
- Hedström, Peter Swedberg, Richard (Hg.) (1998), *Social Mechanisms. An Analytical Approach to Social Theory*, Cambridge.

- King, Gary/Keohane, Robert O./Verba, Sidney (1994), *Designing Social Inquiry – Scientific Inference in Qualitative Research*, Princeton, NJ.
- Lewis-Beck, Michael S./Tom W. Rice (1992), *Forecasting Elections*, Washington, D.C.
- McKeown, Timothy J. (1999), »Case studies and the statistical worldview: Review of King, Keohane, and Verba's designing social inquiry: Scientific inference in qualitative research«, *International Organization*, Jg. 53, H. 1, S. 161–190.
- Pierson, Paul/Skocpol, Theda (2002), »Historical institutionalism in contemporary political science«, in: Ira Katznelson/Helen Milner (Hg.), *Political Science: The State of the Discipline*, New York, NY, S. 693–721.
- Ragin, Charles C. (1987), *The Comparative Method: Moving beyond Qualitative and Quantitative Strategies*, Berkeley, CA.
- (2000), *Fuzzy-Set Social Science*, Chicago, IL.
- Russett, Bruce (1993), *Grasping the Democratic Peace. Principles for a Post-Cold War World*, Princeton, NJ.
- Scharpf, Fritz W. (1997), *Games Real Actors Play. Actor-Centered Institutionalism in Policy Research*, Boulder, CO.
- Schimmelfennig, Frank (2006). »Prozessanalyse«, in: *Methoden der Politikwissenschaft: Neuere qualitative und quantitative Analyseverfahren*, Joachim Behnke/Thomas Gschwend/Delia Schindler/Kai-Uwe Schnapp (Hg.), Baden-Baden, S. 273–285
- Schnapp, Kai-Uwe/Schindler, Delia/Gschwend, Thomas/Behnke, Joachim (2006). »Qualitative und Quantitative Zugänge: Eine integrative Perspektive«, in: *Methoden der Politikwissenschaft: Neuere qualitative und quantitative Analyseverfahren*, Joachim Behnke/Thomas Gschwend/Delia Schindler/Kai-Uwe Schnapp (Hg.), Baden-Baden, S. 11–26.
- Schneider, Carsten Q. (2006), »Qualitative comparative analysis und fuzzy sets«, in: Joachim Behnke/Thomas Gschwend/Delia Schindler/Kai-Uwe Schnapp (Hg.), *Methoden der Politikwissenschaft: Neuere qualitative und quantitative Analyseverfahren*, Baden-Baden, S. 263–271.
- Seawright, Jason (2002), »Testing for necessary and/or sufficient causation: which cases are relevant?«, *Political Analysis*, Jg. 10, H. 2, S. 178–193.
- Thomas, George (2005), »The Qualitative foundations of political science methodology«, *Perspectives on Politics*, Jg. 3, H. 4, S. 855–866.
- Walker, Henry A./Cohen, Bernard P. (1985), »Scope statements: Imperatives for evaluating Theory«, *American Sociological Review*, Jg. 50, H. 3, S. 288–301.
- Western, Bruce (1995), »Concepts and suggestions for robust regression analysis«, *American Journal of Political Science*, Jg. 39, H. 3, S. 786–817.