

Hans-Uwe Otto
Andreas Polutta
Holger Ziegler (Hrsg.)

What Works – Welches Wissen braucht die Soziale Arbeit?

Zum Konzept evidenzbasierter Praxis

Verlag Barbara Budrich



What Works –

Welches Wissen braucht die Soziale Arbeit?

Hans-Uwe Otto
Andreas Polutta
Holger Ziegler (Hrsg.)

What Works –
Welches Wissen braucht
die Soziale Arbeit?

Zum Konzept evidenzbasierter Praxis

Verlag Barbara Budrich
Opladen & Farmington Hills, MI 2010

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen
Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über
<http://dnb.d-nb.de> abrufbar.

Alle Rechte vorbehalten.

© 2010 Verlag Barbara Budrich, Opladen & Farmington Hills, MI
www.budrich-verlag.de

eISBN 978-3-86649-660-6

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlages unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Umschlaggestaltung: disegno visuelle kommunikation, Wuppertal – www.disenjo.de
Druck: paper & tinta, Warschau
Printed in Europe

Inhalt

Hans-Uwe Otto, Andreas Polutta & Holger Ziegler
Zum Diskurs um evidenzbasierte Soziale Arbeit 7

Teil I Wissen und Professionalität

Jennifer L. Bellamy, Sarah E. Bledsoe & Edward J. Mullen
Evidenzbasierte Sozialarbeitspraxis – Konzepte und Probleme der
Implementation 29

Fabian Kessl & Alexandra Klein
Das Subjekt in der Wirkungs- und Nutzerforschung..... 63

Daniel Gredig & Peter Sommerfeld
Neue Entwürfe zur Erzeugung und Nutzung
lösungsorientierten Wissens 83

Gert Biesta
Evidenz und Werte in Erziehung und Bildung: Drei weitere Defizite
evidenzbasierter Praxis 99

Teil II Evidenzbasierte Praxis und Wirkungsorientierte Forschung

Matthias Hüttemann
Woher kommt und wohin geht die Entwicklung
evidenzbasierter Praxis?..... 119

David L. Albright & Bruce A. Thyer
Die Anwendung des evidenzbasierten Praxismodells auf
die Soziale Arbeit..... 137

Heinz-Günter Micheel
Die Leistungsfähigkeit empirischer Sozialforschung im Kontext
sozialpädagogischer Praxis 151

Christoph M. Schmidt

| | |
|---|-----|
| Grundlagen evidenzbasierter Politikgestaltung bei sozialen Dienstleistungen..... | 165 |
|---|-----|

Teil III

Eine neue Soziale Arbeit?

Stephen A. Webb

| | |
|--|-----|
| Zur Validität von evidenzbasierter Praxis in der Sozialen Arbeit – Einige Überlegungen..... | 187 |
|--|-----|

Heinz-Jürgen Dahme & Norbert Wohlfahrt

| | |
|---|-----|
| Evidenzbasierte Soziale Arbeit und wettbewerblich gesteuerte Sozialwirtschaft..... | 203 |
|---|-----|

Enola Proctor & Aaron Rosen

| | |
|--|-----|
| Herausforderungen beim Gebrauch empirischer Evidenz in der Praxis | 217 |
|--|-----|

Stefanie Albus, Heinz-Günter Micheel & Andreas Polutta

| | |
|--|-----|
| Empirie der Wirkungsorientierung – Perspektiven einer evidenzbasierten Professionalisierung | 231 |
|--|-----|

Jean-Michel Bonvin & Emilie Rosenstein

| | |
|--|-----|
| Jenseits evidenzbasierter Steuerungsmodelle: Kognitive Rahmen und ihre normativen Implikationen in „Steuerungsmodellen zur sozialen Integration“ | 245 |
|--|-----|

| | |
|--------------------------|-----|
| Autorenverzeichnis | 267 |
|--------------------------|-----|

Hans-Uwe Otto, Andreas Polutta & Holger Ziegler

Zum Diskurs um evidenzbasierte Soziale Arbeit

Die Debatte um eine wirkungsorientierte Steuerung und um die damit verbundene Idee einer ‚evidenzbasierten Sozialen Arbeit‘ ist im vollen Gange und politisch wie fachlich hoch umstritten.

Der vorliegende Sammelband bringt, zum ersten Mal in deutscher Sprache, die Perspektiven internationaler ProtagonistInnen aber auch SkeptikerInnen in einem gemeinsamen Sammelband zusammen. Eine auf breiter Expertise fundierte, fachlich-kritische Debatte um die Entwicklung der Sozialen Arbeit hin zu einer evidenzbasierten Praxis ist überfällig. Denn die Debatte um Wirkungsorientierung in der Sozialen Arbeit beinhaltet nicht nur eine pragmatische Diskussion darüber, wie die Soziale Arbeit auf Basis eines wissenschaftlichen Fundaments ihre Zielerreichung dokumentieren und effizienter gestalten kann, sondern auch eine Debatte darüber, was der Auftrag und das Ziel Sozialer Arbeit sein soll und wie Soziale Arbeit durchzuführen sei. Pointiert formuliert ist die Debatte um Wirkungsorientierung nicht zuletzt eine politische Auseinandersetzung darüber, was gute Soziale Arbeit ist und sein soll und in der Tat besteht wenig Zweifel daran, dass eine administrative Steuerung und praktische Gestaltung Sozialer Arbeit nach den Prämissen der Wirkungsorientierung den Charakter der Sozialen Arbeit nachhaltig verändern wird.

Die sich derzeit international abzeichnenden Versuche einer ‚wirkungsorientierten Steuerung‘ passen häufig allzu nahtlos in einen neo-liberalen Kontext der Reform der personenbezogenen Wohlfahrtsproduktion und schließen sich Steuerungsformen an, die betriebswirtschaftlich orientierten Mustern des New Public Management folgen (vgl. Gray/MacDonald 2006). Im Verbund mit Benchmarks, Verfahren der Budgetierung, des Kontraktmanagements, der dezentralisierten Ergebnisverantwortung (Accountability) und anderen Instrumenten fiskalischer und bürokratischer Fremdkontrolle professioneller Wohlfahrtsproduktion verspricht die Wirkungsorientierung eine kardinal messbare Metrik zur effizienzorientierten Sicherstellung von „Value for Money“ zu liefern, das die entscheidende Größe im wettbewerbsorientierten „Market Game“ im öffentlichen Bereich darstellt (vgl. Otto/Schnurr 2000). In diesem ideologischen Kontext ist nicht von der Hand zu weisen, dass die pragmatische und vermeintlich nicht-ideologische Idee einer evidenzbasierten Sozialen Arbeit schnell in eine Verlängerung der bislang primär organisationsbezogenen Prämissen der neuen Steuerung in den sozial-

arbeiterischen Kernbereich der Gestaltung personenbezogener Wohlfahrtsproduktion münden kann.

Während es in einer fahrlässigen Weise naiv erscheint diesen Kontext nicht zu berücksichtigen, ist die *wissenschaftliche Debatte* über die Bestimmung von Wirksamkeit vergleichsweise unabhängig von der *politischen Frage* des Stellenwerts einer wirkungsorientierten Neuausrichtung im Kontext des New Public Managements. Sofern es der wissenschaftlichen Debatte gelingt, sich von einer naiven Anwendungsorientierung in bestehenden Settings zu lösen, können sich aus einer gleichsam professionstheoretisch wie methodologisch aufgeklärten wissenschaftlichen Debatte Einsichten über die fachliche Leistungsfähigkeit Sozialer Arbeit ergeben, die nicht in eine naive Legitimation managerieller Steuerungen münden müssen, die ein Primat funktionaler Rationalität in der Erbringung Sozialer Arbeit als fraglos voraussetzt und technologisch durchzusetzen trachtet. Gleichwohl scheint die Debatte um eine evidenzbasierte Soziale Arbeit in einem Spannungserhältnis zu professions- und professionalisierungstheoretischen Prämissen zu stehen.

Unabhängig von dem Postulat, dass Soziale Arbeit ihre Leistungen in einer wirksamen Weise gestalten soll, ist die Frage welches Wissen eine gute und angemessene Form Sozialer Arbeit brauche, nicht nur ein Gegenstand politischer, sondern auch analytischer, methodologischer, ethischer und nicht zuletzt professions- und wissenschaftstheoretischer Kontroversen (vgl. Farley et al. 2009).

Die Positionen dieser Debatte werden in diesem Sammelband auf höchstem internationalem Niveau vorgetragen. Die Herausgeber dieses Bandes sind keinesfalls neutral in dieser Debatte. Daher haben wir uns entschieden, in dieser Einleitung darauf zu verzichten einen bloßen zusammenfassenden Überblick über die Debatte dieses Bandes zu geben, sondern einen eigenständigen Beitrag zum Einstieg in die Thematik zu verfassen. Dieser verweist auf einige zentrale methodische Prämissen der Wirkungsforschung und fragt nach den Konsequenzen, die eine direkte Verkopplung der wissenschaftlichen Agenda einer validen Wirksamkeitsprüfung mit der praktischen Agenda einer angemessenen Leistungserbringung impliziert. Denn bereits hier unterscheidet sich die zunehmend an Bedeutung gewinnende Idee einer wirkungsorientierten Steuerung und Erbringung Sozialer Dienste von dem bislang maßgeblichen Entwicklungs- und Legitimationsmodell einer Professionalisierung der Sozialen Arbeit.

Kritik an der professionellen Steuerung Sozialer Arbeit

Im Professionalisierungsmodell der Erbringung und Steuerung Sozialer Dienste findet sich die Vorstellung, dass wissenschaftliches Wissen in einer

besonderen Relationierung zu anderen relevanten Wissenstypen, vor allem vermittelt über die Figur des wissenschaftlich ausgebildeten Professionellen in die Praxis der Sozialen Arbeit kommt. Professionelle SozialarbeiterInnen, so etwa die professionstheoretische Argumentation von Bernd Dewe und Hans-Uwe Otto, greifen für die stellvertretende Bearbeitung der sich in der Besonderheit des Einzelfalls manifestierenden lebenspraktischen Probleme zwar konstitutiv auf wissenschaftliches Wissen in Gestalt von Erklärungs-, Deutungs- und Problemlösungswissen zurück, im Zentrum professionellen Handelns stehe dabei jedoch „die Fähigkeit der diskursiven Auslegung und Deutung von lebensweltlichen Schwierigkeiten und Einzelfällen mit dem Ziel der Perspektiveneröffnung bzw. einer Entscheidungsbegründung unter Ungewissheitsbedingungen“ (Dewe/Otto 2005: 179) wobei das entscheidende Kriterium für das Richtige des professionellen Handelns weniger die Wahrheitsdifferenz als solche, sondern das der Angemessenheit sei (vgl. Dewe/Otto 2005: 194).

In der gegenwärtigen Debatte um eine Wirkungsorientierung geht es weniger um die professionstypische Relationierung von wissenschaftlichem Wissen, sondern um eine veränderte Figuration, die vor allem in den instruktiven medizinsoziologischen Arbeiten von Steve Harrison als *wissenschaftlich-bürokratisches Organisationsmodell* beschrieben worden ist. Wissenschaftlich sei dieses Modell insofern, „that it draws on the accumulated evidence of large-scale research, and ‚bureaucratic‘ in the sense that it translates the output of such research into a particular species of bureaucratic rule, the ‚clinical guideline‘, for application in [... social welfare¹] organisations“ (Harrison/Moran/Wood 2002: 7; vgl. White/Stacombe 2003).

Wie wir im Folgenden argumentieren, läuft die Idee einer wirkungsorientierten evidenzbasierten Sozialen Arbeit der Logik nach auf die Überführung eines „reflexive practice model“ in ein „scientific bureaucratic model“ hinaus (vgl. Harrison/Moran/Wood 2002; Harrison/McDonald 2003).

Vordergründig hängt die Konjunktur der Wirkungsorientierung damit zusammen, dass Soziale Arbeit verstärkt mit der Forderung konfrontiert ist, die Effektivität und Effizienz ihrer Leistungen unter Beweis zu stellen. Diese Forderung impliziert jedoch per se nicht viel grundlegend Neues. Da Soziale Arbeit eine öffentliche Institution darstellt, die systemische Eingriffe in lebensweltliche Konstellationen und die Lebensführung ihrer KlientInnen durchführt, gibt es keine relevante fachliche Position, die bestreiten würde, dass sich die Soziale Arbeit unter anderem über ihre Wirksamkeit zu legitimieren hat. Zwar verweisen insbesondere berufsethische Positionen darauf, dass nicht jede Maßnahme, die gewünschte Wirkungen entfaltet, auch eine angemessene Maßnahme darstellt, umgekehrt gilt jedoch allemal, dass Maß-

1 Bei Harrison et al. ist von „medical care organisations“ die Rede.

nahmen, die keine Nutzwerte im Sinne positiver Effekte auf die Zustände und Lebensführungen ihrer KlientInnen zeitigen, nicht zu rechtfertigen sind.

Sofern man dies als Konsens voraussetzen kann, erscheint es allemal überraschend, dass die Wirkungsorientierung nicht nur und nicht primär das Ergebnis einer fachlichen Debatte ist. Vielmehr ist sie mit dem verwaltungstechnischen, politischen und moralischen Programm einer neuen wettbewerbsorientierten und an Kosten-Nutzen-Kriterien orientierten Steuerung öffentlicher Leistungen verknüpft. Eine solche Steuerung verlangt zum einen Zurechenbarkeit und Sicherheit und ist im Falle von formalisierten Arbeitsroutinen allemal leichter durchzusetzen als eine professionelle Praxis, die Handlungsautonomie für sich reklamiert und mit Blick auf die praktische Arbeit am und mit dem einzelnen Fall die Dimensionen von „Kontingenz, Kommunikation, Verstehen und Empathie“ (Dewe 2009: 96) betont, die einer Standardisierung und externen Regulation enge Grenzen setzen.

War die Steuerung Sozialer Dienste traditionell primär mit der Rationalisierung und Optimierung von Verwaltungsabläufen beschäftigt, widmet sie sich gegenwärtig – auch unter dem Eindruck einer seit den 1990er Jahren „im Kern technokratische[n], vornehmlich effizienz- und leistungsorientierte[n] Debatte um ‚Qualität‘“ (vgl. Dewe/Otto 2005: 187; vgl. Beckmann et al. 2004; Klatetzki 2005) – auch auf die pädagogischen Prozesse der unmittelbaren Erbringung Sozialer Dienste, die bislang weitgehend im Ermessen und der Entscheidungskompetenz von Wohlfahrtsprofessionellen lagen.

Die dabei sichtbar werden Versuche einer managerialistischen Steuerung sind nicht zuletzt durch einen Zweifel an der Effektivität und Effizienz professioneller Handlungs- und Entscheidungsautonomie gekennzeichnet. Denn eine Steuerung, die die Produktion personenbezogener Wohlfahrtsleistungen in das Ermessen und die Entscheidungskompetenz von Wohlfahrtsprofessionellen wird – zusammen mit dem Vorwurf eines generellen „Mangel[s] an Effektivitäts- und Effizienzinteresse bei der verbandlichen Wohlfahrtsproduktion“ (Wilken 1998: 226) – Sorge für Dysfunktionen im System der Wohlfahrtsproduktion, da das Interesse an der Hilfeerbringung selbst und nicht das Interesse an deren erfolgreiche Beendigung belohnt werde (vgl. Struzyna 2006).

Zugleich wird die professionstheoretische Betonung des Ermessens und die Entscheidungskompetenz der Professionellen massiv kritisiert. Die („unstrukturierten“ und konsensusbasierten) Ermessensurteile von Professionellen seien in der Regel nicht sonderlich zuverlässig und valide (Baird/Wagner 2000) und der Spielraum von Professionellen stelle keine gegenstandsangemessene Notwendigkeit dar, sondern eine Strategie der Kritikimmünisierung und Privilegiensicherung einer anmaßenden, im Wesentlichen „authority-based profession“ (Gambrill 2001): „Professions such [...] social work“, argumentiert etwa Eileen Gambrill (2001: 170) „have been successful in bamboozling the public and those who fund service programs into belie-

ving that professionals offer unique services that require specialized training and experience. The armentarium of propaganda stratagems developed to maintain this view is fascinating in its variety and impressive in its success. It is its own little orchestra, with the violins saying ‘trust us’, the brass saying ‘we rigorously appraise claims and teach critical thinking’, the drums and oboes saying ‘be careful what you will lose if you limit us’ and the conductor, sociological propaganda and its reflection in professional organizations and academic social work, influencing all [...]. This bamboozlement serves important functions for the profession (e.g., maintaining and expanding turf), for professionals (e.g., freedom to do whatever they want without regard for evidentiary criteria), and for clients (e.g., removing responsibility for personal conduct, muting the uncertainties of life [...]).”

Abhilfe mit Blick auf die Probleme der professionalistischen Ideologie wird von den Ergebnissen valider empirischer Wirkungsevaluationen erwartet. Diese sollen auf wissenschaftlicher Evidenz basierte Ergebnisse bereit stellen, die es erlauben, drängende praktische Probleme zu lösen und eine effektive Praxis anzuleiten. Unausgesprochen scheint dabei eine manualisierte Erbringung von Leistungen, wie sie etwa in der evidenzbasierten Humanmedizin an Bedeutung gewinnt, als Vorbild zu dienen. Nicht der einzelne Professionelle reflektiert und relationiert wissenschaftliche Erkenntnisse mit Blick auf seine Deutung des Einzelfalls, sondern die Organisation hält Programme bereit, die mit Blick auf trennscharf diagnostizierbare Problemkonstellationen und definierte Leistungsziele die höchste Wirkungswahrscheinlichkeit versprechen. Dies setzt voraus, dass nicht die je einzelnen Professionellen entscheiden, welche Vorgehensweise mit Blick auf die Besonderheiten des Falls angemessen erscheinen, sondern dass sie – gerade um idiosynkratische Prozesse zwischen Professionellen und Klienten zu vermeiden – in einer formalisierten Weise die je einzelnen Schritte eines spezifischen Programms bzw. einer Behandlungsleitlinie durchführen, die mit Blick auf eine Problempopulation das höchste Maß an Zielereichung versprechen (vgl. Auckenthaler 2000). Der Fokus auf Problem- bzw. Risikopopulationen anstelle des professionskonstitutiven Einzelfallbezugs wird in der internationalen Debatte auch als „actuarialism“ oder „actuarial risk assessment“ (versicherungsmathematische Risikobestimmung) verhandelt und meint „the formal statistical calculation of risk based upon aggregated data“ (vgl. Kemshall 2010; Webb 2009).

Wie die Beiträge in diesem Band verdeutlichen, ist mit Blick auf die Wirkungsorientierung gerade der „actuarialism“ sowie die Frage ob und inwiefern eine Manualisierung der Praxis der Sozialen Arbeit fachlich angemessen und zielführend ist, hoch umstritten. Methodologisch ist die Frage der Manualisierungsmöglichkeiten sozialarbeiterischer Praxis eng mit der Debatte um (randomisierte) kontrollierte Experimente verbunden, die häufig als Goldstandard der Wirkungsforschung verstanden werden.

Dabei findet sich kaum ein Wissenschaftler, der die Aussagekraft und Zuverlässigkeit methodisch anspruchsvoller Experimentalforschungen per se bezweifeln würde. Strittig ist jedoch, ob die Wirkungsbeschreibungen der Experimentalforschung eine notwendige und vor allem eine hinreichende Grundlage sind, um die sozialarbeiterische Praxis ‚anzuleiten‘.

Um diese Debatte zu rahmen, soll zunächst der Frage nachgegangen werden, was durch (experimentelle) Wirkungsmessung denn genau gemessen wird.

Was misst (experimentelle) Wirkungsmessung?

An Untersuchungen, die behaupten, Wirkungen der Sozialen Arbeit zu messen, mangelt es nicht. Allerdings finden sich bisher nur wenige (deutschsprachige) Studien, die dieses Versprechen in einer methodisch haltbaren Weise einlösen können. Zu den Studiendesigns, die es erlauben, belastbare Aussagen über die Wirkung von Maßnahmen zu formulieren, gehören ohne Zweifel die (randomisierten) kontrollierten Experimentalstudien (RCTs).²

RCTs basieren darauf, Ergebnisse von AdressatInnengruppen, die eine bestimmte Maßnahme erhalten haben, mit den Ergebnissen von Gruppen zu vergleichen, die diese Maßnahme nicht erhalten. Sofern methodisch sicher gestellt ist, dass die Differenz der Ergebnisse dieser Gruppen nicht aufgrund anderer Einflüsse als durch die Maßnahme zustande gekommen sein können, gilt diese Differenz als Wirkung der Maßnahme. Gemessen wird dabei zunächst die Wahrscheinlichkeit mit auf ein Ereignis ‚A‘ (z.B. eine sozialarbeiterische Maßnahme) unter definierten Bedingungen auf ein Ereignis ‚B‘ (= die potenzielle Wirkung der Maßnahme) folgt. Eine notwendige Bedingung, um von Wirkung (zunächst im Sinne einer Granger-Kausalität) zu sprechen, besteht darin, dass das ‚ursächliche‘ Ereignis A dem Ereignis B zeitlich vorausgeht³ und A dabei einen messbaren Einfluss auf B in der Form hat, dass die konditionale Wahrscheinlichkeit von B in Anbetracht von A höher ist als die unbedingte Wahrscheinlichkeit von B. Bei Evaluationen, die sich auf

-
- 2 Ob RCTs deswegen tatsächlich den „Goldstandard“ der Wirkungsforschung darstellen, ist umstritten: Es finden sich überzeugende methodische und methodologische Argumente dafür, dass es neben RCTs eine Reihe weiterer Verfahren gibt, die in der Lage sind, in einer validen Weise Kausalinferenzen zu beschreiben (vgl. Cartwright 2007). So weist auch der Sachverständigenrat für die Konzentrierte Aktion im Gesundheitswesen (2003: 87) darauf hin, dass „[r]andomisierte, kontrollierte, klinische Studien, die lediglich die Effektivität (und gegebenenfalls Effizienz) einer Maßnahme unter artifizielle Studienbedingungen (efficacy) beschreiben, [...] in vielen Fällen überbewertet [werden]“.
 - 3 RCTs basieren auf mindestens zwei Messzeitpunkten: Eine Messung vor der Maßnahme und eine nach der Maßnahme. Kann man mehr als zwei Messzeitpunkte heranziehen, ist es unter Umständen auch möglich, Wirkungsdynamiken und Verläufe sichtbar zu machen.

Überprüfung dieser Zusammenhangskonstellationen richten, hat man es zunächst im Wesentlichen mit Outcome-Analysen zu tun. Solche Outcome-Analysen lassen sich von einer validen Wirkungsanalyse im engeren Sinne deshalb unterscheiden, weil sie nicht ausschließen können, dass eine Reihe von weiteren bekannten oder unbekanntem Faktoren unabhängig von A die Wahrscheinlichkeit von B beeinflussen. Hier besteht die besondere Leistungsfähigkeit von RCTs: Sie können sicherstellen, dass sich keine anderen Einflüsse (sog. Störvariablen) finden, die den Zusammenhang von A und B statistisch irrelevant machen. RCTs stellen ein Studiendesign dar, das es erlaubt, den Einfluss von Störvariablen durch Verfahren der Randomisierung zu eliminieren. Randomisierung meint die zufällige Verteilung der Probanden in eine Experimentalgruppe, die die zu untersuchende Maßnahme und in eine Kontrollgruppe, die keine oder eine alternative Maßnahme erhält. Sofern die Stichprobe in jeder Gruppe groß genug ist, um gemäß dem zentralen Grenzwertsatz⁴ davon auszugehen, dass die Wahrscheinlichkeiten der t-Verteilung kontinuierlicher Zufallsvariablen approximativ mit der Normalverteilung berechnet werden kann (als Faustregel, ab $n > 30$), stellt eine solche Zufallsaufteilung sicher⁵, dass sich die bekannten und unbekanntem Störvariablen in der Experimental- und Kontrollgruppe ebenfalls approximativ gleich verteilen oder zumindest nicht systematisch in einer der Gruppen ballen. Eine Gleichverteilung der Störvariablen sorgt dafür, dass sich auch deren systematischer Einfluss in der Experimental- und Kontrollgruppe gleich verteilt. Bezogen auf den Gruppenvergleich ist dieser Einfluss damit im statistischen Sinne eliminiert. Deshalb kann etwa David Farrington (2003: 53) zu Recht argumentieren: „Randomization is the only method of assignment that controls for unknown and unmeasured confounders as well as those that are known and measured.“ Die statistische Eliminierung des Einflusses der Störvariablen gilt jedoch nur für die Aggregationsebene der Gruppe. Sie gilt nicht für die jeweils einzelnen Individuen dieser Gruppe. Die Wirkungsaussagen von RCTs beziehen sich insofern auf Gruppenmittelwerte und treffen nicht auf einer Individualebene zu.⁶

Was impliziert der Einsatz einer solchen experimentellen Wirkungsfor-schung nun für ein Praxisfeld wie das der Sozialen Arbeit? Aufschlussreich ist ein Blick in das Standardwerk zur Experimentalforschung in der Sozial-wissenschaft: In *Experimental and Quasi-Experimental Designs for Generalized Causal Inference* zeigen William Shadish, Thomas Cook und Donald

4 Der zentrale Grenzwertsatz lautet kurz gesagt, dass die Summe von stochastisch unabhängigen Zufallsvariablen annähernd normal verteilt ist.

5 Dabei ist fraglich, ob eine solche Zufallsverteilung bei Forschung mit Menschen in Bezug auf sozial oder rechtlich relevanten Interventionen in einer unverzerrten Weise gelingen kann (vgl. Graebisch 2007) ohne ethisch fragwürdige Konsequenzen zu implizieren.

6 Aus aggregierten Maßen bezüglich einer zentralen Tendenz lassen sich keine validen Aussagen über die Eignung der Maßnahme für einen bestimmten Klienten ziehen (dazu: Westen/Novotny 2004).

Campbell (2002) einzigartige Stärken von RCTs auf: das Beschreiben der Folgen, die einer bewussten Variation einer Behandlung zugeschrieben werden können. Shadish et al. nennen die Ergebnisse dieses Typs der Wirkungsmessung treffend *Kausalbeschreibung*. Das Problem besteht aber darin, dass RCTs nichts über die Bedingungen und Mechanismen aufklären können, durch die die beschriebenen Kausalbeziehungen zu Stande kommen. In der Sprache von Shadish et al. gelingt es ihnen nicht *Kausalerklärungen* zu liefern.

Wie wir im Folgenden zeigen, ist der Unterschied von Kausalbeschreibungen und Kausalerklärungen dafür verantwortlich, dass die praktische Implementation von Maßnahmen, die sich in Kausalanalysen der Experimentalforschung als wirksam erwiesen haben, letztlich nur dann möglich, wenn die Maßnahmen im Falle ihrer Evaluation und ihrer Replikation an anderer Stelle auf einem „actuarial risk assessment“ in der diagnostischen Klassifikation der Fälle aufbauen und daran anschließend in einer möglichst genau manualisierten Form erfolgen.⁷

Das heißt nicht, dass die Ergebnisse von RCTs jenseits einer manualisierten Praxis irrelevant wären. Im Gegenteil: Die praktische Relevanz von Kausalbeschreibungen kann enorm sein, wenn es ‚lediglich‘ darum geht herauszufinden, ob ein bestimmtes Behandlungsprogramm eine bestimmte, erwartete Wirkung erreicht oder nicht erreicht hat. Es hat nichts damit zu tun, die Leistungsfähigkeit der Experimentalforschung zu bestreiten, wenn man darauf verweist, dass der Nutzen von RCTs davon abhängig ist, wozu die Ergebnisse der Kausalbeschreibungen dienen sollen. Im Falle der evidenzbasierten Sozialen Arbeit scheint das wesentliche Erkenntnisinteresse offensichtlich weniger darin zu bestehen, festzustellen, in welchem Ausmaß eine bestimmte Maßnahme erfolgreich oder nicht erfolgreich *war*, sondern vielmehr darin, wirk-

7 Dies lässt sich gut am Beispiel des prominenten Elterntrainingsprogramms Tripple P (Positive Parental Program) zeigen. Eine Reihe von Studien mit mäßiger Qualität hat große Hoffnungen bezüglich der Leistungsfähigkeit dieses sehr erfolgreich vermarkteten und verbreiteten Programms geschürt. Die Kausalbeschreibungen einer hochwertigen randomisiert kontrollierten Experimentalstudie, die Manuel Eisner und Denis Ribeaud durchgeführt haben, zeichnen ein anderes Bild: Das Tripple P Programm, so weisen die Autoren nach, hat „keinerlei positive Wirkungen im Verhalten des Kindes ausgelöst, weder auf externalisierenden noch auf internalisierenden Verhaltensdimensionen“. Im Sozialverhalten der Kinder zeigte sich vielmehr „ein unerwünschter Effekt von Tripple P. Die Lehrpersonen beobachteten bei Kindern von Eltern, die an einem Kurs teilgenommen hatten, tendenziell eine Zunahme von nicht-aggressivem Problemverhalten, während bei den Kindern der Kontrollgruppe ein Rückgang wahrgenommen wurde“ (Eisner/Ribeaud 2008: 5). Die Kausalbeschreibung dieser Effekte erscheint allemal ausreichend, um Tripple P mit Skepsis zu begegnen. Wenn die Aufgabe einer Wirkungsevaluation in der bloßen Feststellung besteht, ob ein bestimmtes Programm – bemessen an einem klar definierten End – wirksam oder unwirksam *war*, und man darauf verzichten kann, die Zusammenhänge zu eruieren, die diesen Erfolg oder Misserfolg bedingen, fällt es schwer, sich eine zuverlässigere Methode vorzustellen als das randomisierte kontrollierte Experiment (vgl. Boruch 1997; LaLonde 1984; Pignotti/Thyer 2009).

same Praktiken zu erkunden, um sie *künftig* in Form einer evidenzbasierten Sozialen Arbeit durchzuführen.

Eine evidenzbasierte Arbeit, die Maßnahmen wiederholen möchte, deren Wirksamkeit an anderer Stelle nachgewiesen wurde, ist nur um den Preis einer Manualisierung zu haben. Der Grund dafür liegt in einer offensichtlichen methodischen Konsequenz von RCTs, deren Kausalbeschreibungen notorisch mit dem Problem ‚externer Validität‘ konfrontiert sind (vgl. Campbell/Stamley 1963). „The main threats to external”, so erläutert etwa David Farrington (2003: 55), „consist of interactions of causal relationships (effect sizes) with types of persons, settings, interventions, and outcomes”. Die Frage der ‚externen Validität‘ verweist zunächst weniger auf ein Studien-, sondern auf ein Situationskriterium und bezieht sich auf die Verallgemeinerungsfähigkeit der Studienergebnisse. Wie ausgeführt, besteht der Sinn von RCTs darin, sowohl den Einfluss der Kontexte, in denen die Experimente stattfinden, zu minimieren als auch den Einfluss der vermittelnden Faktoren, die zwischen Interventionen und Effekten intervenieren. Dies lässt sich auch als Versuch der Sicherstellung einer möglichst hohen internen Validität bzw. einer Validität der Ceteris-paribus Bedingungen beschreiben (vgl. Farrington 2003). Das Problem besteht nun darin, dass sich im Falle von RCTs ein „trade-off between internal and external validity” findet (Cartwright 2007: 11; vgl. auch Lambert et al. 2004): „RCTs have high internal validity but the formal methodology puts severe constraints on the assumptions a target population must meet to justify exporting a conclusion from the test population to the target” (Cartwright 2007: 11). Da RCTs nun nicht in der Lage sind die Bedingungen und generischen Mechanismen zu erfassen, die Wirkung hervorbringen (vgl. Pawson/Tilley 2009; Shadish et al. 2002), sind alle Replikationen der Kausalbeschreibungen von RCTs mit dem Problem konfrontiert, dass diese Wirkungsaussagen zunächst ausschließlich für die tatsächlich geprüften Kontexte und Programmgestaltungen gültig sind. RCTs klären über die relative Wahrscheinlichkeit auf, mit der definierte Maßnahmen, die gegenüber einer definierten Zielgruppe in einer ganz bestimmten Form und in einem ganz bestimmten Setting durchgeführt werden, im Vergleich zu anderen Maßnahmen oder einer Non-Intervention zu einem klar definierten End führen. Eine Übertragung der Ergebnisse von RCTs auf andere Kontexte und Programmgestaltungen (oder eine Ableitung inwiefern die geprüften Maßnahmen „ähnliche“ Programmziele (Ends) erreichen), ist nicht möglich ohne genau das zu gefährden, was RCTs mühevoll sicherzustellen versuchen, nämlich die interne Validität der Wirkungsaussagen. Anders formuliert: In Struktur und Ausprägung veränderte ‚Rahmenbedingungen‘ oder Modifizierungen der Interventionsschritte machen die Aussagen der Kausalbeschreibung als ‚objektive‘ Entscheidungsgrundlage empirisch wertlos.

Nicht von ungefähr fordert der Wissenschaftliche Beirat Psychotherapie (WBP) in seinen im Juni 2009 verabschiedeten *Verfahrensregeln zur Beurteilung der wissenschaftlichen Anerkennung von Methoden und Verfahren der Psychotherapie* sowohl standardisierte Diagnosen als auch ein „Therapiemanual, bei dem die Interventionen so beschrieben sind, dass das therapeutische Vorgehen vergleichbar und replizierbar ist“, eine „Festlegung und umfassende Beschreibung der Kontrollbedingung“ sowie ein Belegen der „Manual-treue/Treatmentintegrität durch externe Beobachter“ oder zumindest „durch Fragebögen“ (WBP 2009: 31). In assessment protocols, wie sie etwa durch die Campbell Collaboration oder in Form der Maryland Scientific Methods Scale (vgl. Farrington et al. 2002) vorgeschlagen werden, werden Ergebnisse von Studien, die diese Kriterien nicht erfüllen, als unzuverlässig eingeschätzt und gar nicht in Review-Studien einbezogen.

Diese Forderung an Evaluationsstudien, diese Kriterien zu erfüllen, ist deshalb folgerichtig, weil bei der Implementation evidenzbasierter Programme bereits leichte Modifikationen der Programme, deren Kontexte oder deren AdressatInnen die Validität der kausaldeskriptiven Aussausagen massiv in Mitleidenschaft ziehen können. Interventionen, so lautet daher die Forderung, „should be rigorously managed and delivered as designed“ (Chapman/Hough 1998: 27) und selbst energische VertreterInnen einer „evidence-based practice“ weisen darauf hin, dass das Verhältnis der „fidelity to the evidence base versus innovation“ eine der wesentlichen ungeklärten „tensions inherent in evidence-based programmes“ darstelle (Nutley/Homel 2006: 5). Da bislang noch kein Weg gefunden wurde, dieses Spannungsverhältnis zu lösen⁸, besteht der methodisch überzeugendste Weg der Implementation evidenzbasierter Programme darin, ihre Integrität und Replizierbarkeit über detaillierte Manuale sicherzustellen, d.h. durch schriftlich fixierte Prozessanweisungen, die PraktikerInnen vorschreiben, in einer definierten Situation X die Handlung Y zu vollziehen, was im Wesentlichen – unabhängig von einer Verwissenschaftlichung der Techniken der Leistungserbringung – einer Ent-Professionalisierung zugunsten einer erhöhten Bürokratisierung und Regulation der Praxis gleichkommt (dazu: Berg 1997). Genau in diesem Sinne sprechen Harrison et al. von einem wissenschaftlich-bürokratischen Organisationsmodell, das das reflektiert, was Robert Castel bereits in den frühen 1980er Jah-

8 Eine Reihe von VertreterInnen sprechen sich inzwischen für eine Kombination der Ergebnisse von experimentellen und „naturalistischen“ Wirkungsanalysen aus. So etwa David Sackett et al. (1997), die im medizinischen Feld argumentieren, der evidenzbasierten Praxis gehe es um die Verknüpfung der besten verfügbaren externen Evidenz mit individueller klinischer Expertise. Zwar hört es sich (wie immer) gut an, von einer „Entweder-Oder-“ zu einer „Sowohl-Als-Auch-Perspektive“ zu kommen, das Problem besteht nur darin, dass sich RCTs mit Studien im naturalistischen Feld nicht ohne Weiteres – und möglicherweise überhaupt nicht – vergleichen lassen (vgl. Tschuschke 2005) zumal sie unterschiedliche Gegenstände untersuchen.

ren – damals vor allem auf den versicherungsmathematischen Risikodiskurs bezogen – als „tiefgreifende Neuerungen in Bezug auf die Tradition der Mentalmedizin und der Sozialarbeit“ thematisierte, die ein „krasses Missverhältnis zwischen dem Standpunkt des Fachpersonals [...] und dem der Verwaltungskräfte, die diese [Sozial- und] Gesundheitspolitiken definieren und verwirklichen [impliziere [: Die ...] Verwaltung kann sich ganz und gar verselbstständigen [und damit] der Kontrolle der vor Ort Arbeitenden, die zu bloßen Befehlsempfängern werden, völlig entgleiten“ (Castel 1983: 51).

Vor dem Hintergrund der methodischen Prämissen und Gültigkeitskriterien der Experimentalforschung scheint eine wirkungsorientierte Steuerung auf der Basis der entsprechenden Kausalbeschreibungen nicht nur (transparente) Verfahrensstandards der Organisation, sondern Standardisierungen der sozialarbeiterischen Praxis nahe zu legen, die allemal in einem Spannungsverhältnis stehen zu der Betonung des fallinterpretativ-hermeneutischen Ermessens bei Handlungsentscheidungen, das bislang als ein wichtiges Gütekriterium professioneller Sozialer Arbeit galt.

Warum wird eine evidenzbasierte Manualisierung der Praxis kritisiert?

Das Spannungsverhältnis zwischen einem professionalistischen und einem wissenschaftlich-bürokratischen Organisationsmodell spricht per se nicht gegen die dominante Debatte um eine evidenzbasierte Neuausrichtung der Sozialen Arbeit. Eine Reihe von Studien (vgl. Baird/Wagner 2000; Dawes et al. 1989; Grove et al. 2000) legen nahe, dass das subsumtionslogisch-statistische „actuarial risk assessment“ bzw. die „mechanical prediction“, die ohne Subjektbezug und Verfahren des Sinnverstehens auskommt, zu insgesamt valideren und zielgenaueren Fallzuordnungen kommt, als das rekonstruktionslogische professionelle „clinical judgement“, in dem Fallverstehen als eine gleichermaßen konstitutive wie nicht-standardisierbare Dimension des professionellen Handelns erscheint (zur Kritik: Schrödter 2005): „On average“, so argumentieren William Grove et al. (2000: 19), „mechanical-prediction techniques were about 10% more accurate than clinical predictions. Depending on the specific analysis, mechanical prediction substantially outperformed clinical prediction in 33%-47% of studies examined [...] in only a few studies (6%-16%) were [clinical predictions] substantially more accurate“.

Andere fordern explizit die auf Autonomie und Selbstbestimmung pochende und sich nur selbst bestätigende Ideologie des Professionalismus zu überwinden und die „authority-based profession“, durch eine „evidence-based practice“ zu ersetzen (vgl. Gambrill 1999; 2001): „Rendering the [... Social

Work] practice more ‘scientific’ will outweigh the potential caveats of explicating and formalising work routines” (Berg 1997: 1086). In der Tat ist das, was als Professionalität verhandelt wird kein Selbstzweck. Wenn der Nutzwert der Sozialen Arbeit für ihre KlientInnen durch eine manualisierte, evidenzbasierte Soziale Arbeit erhöht wird, haben die professionstheoretischen Prämissen, die dieser Neuformulierung widersprechen, ausgedient.

Allerdings formulieren KritikerInnen dieser Neuausrichtung mindestens drei gewichtige Einsprüche: Der erste Einwand hinterfragt die Konstruktion der Problemlagen, die die Soziale Arbeit bearbeitet.⁹ Implizit scheint die EBP-Idee davon auszugehen, dass Soziale Arbeit „quasi-technische“ Problemlagen bearbeitet, die mehr oder weniger eindeutig wissenschaftlich-empirisch auflösbar sind. Demgegenüber argumentieren KritikerInnen, dass sich die Soziale Arbeit auf menschliche Praktiken richtet, d.h. auf Gegenstände, die nicht nur – wie Naturereignisse – ‚Ursachen‘ haben, sondern – als Handlungen – auf ‚Gründen‘ basieren, die nur unter Berücksichtigung von Motiven und Gründen angemessen zu erfassen sind. Dabei habe es Soziale Arbeit in der Regel mit den individualbiografischen Manifestierungen lebensführungspraktischer Problemlagen, und d.h. in der Regel mit politisch und moralisch umkämpften Problemkonstellationen zu tun, die in einem hohen Maße durch Ambiguitäten gekennzeichnet seien. Soziale Arbeit habe – sowohl in ihren Diagnosen als auch in der praktischen Erbringung ihrer Leistungen – einer (häufig konflikthaften) Pluralität von Haltungen, Auffassungen und Lebensentwürfen Rechnung zu tragen. Bei der Bearbeitung solcher Ambiguitäten erscheint neben dem Kriterium statistisch messbarer Wirksamkeit vor allem das nicht manualisierbare Kriterium der fall- und situationsbezogenen Angemessenheit wesentlich. Bei einer technologischen Anwendung wirksamer Programme rücke dieser Aspekt in den Hintergrund.

Ein zweiter Einwand bezieht sich auf den Fokus der experimentellen Wirkungsforschung, in deren Mittelpunkt in aller Regel die Wirkung einer spezifischen Methode bzw. eines bestimmten Verfahrens steht. Dies gilt, obwohl spezifische manualisierbare Verfahren gerade im Bereich personenbezogener Humandienstleistungen nur einen bescheidenen Teil der Wirkungsvarianz erklären: Metaanalysen der Psychotherapieforschung sprechen für vergleichsweise geringe Effekte spezifischer Interventionen und eine hohe Effektstärke genereller Wirkfaktoren. Setzt man Bona-Fide-Therapien voraus und kontrolliert die ‚Affiliation‘ der Behandelnden und der KlientInnen mit der untersuchten Methode, so klären spezifische Therapieverfahren in der

9 In ihrem Plädoyer für eine evidence-based practice und einen „consistent use of manuals and greater structure“ argumentieren etwa Mark Chaffin und Bill Friedrich, dass „EBP models are well suited to programs delivering a defined and consistent service that ideally does not vary across individual practitioners [... and that are] targeted at a well-defined problem [...and] a well-defined and relatively homogeneous client population“ (Chaffin/Friedrich 2004: 1105).

Regel zwischen 15 % und 1 % erklärter Varianz der Ergebnisse der Therapie auf (vgl. Lambert/Ogles 2004; Wampold 2001). Aufgrund der vergleichsweise starken Einbettung in intersubjektive und lebensweltliche Bedingungskonstellationen und ihrer hohen Abhängigkeit von Fallverstehen und kontextadäquaten Vorgehensweisen ist kaum zu erwarten, dass der Effekt spezifischer Verfahren in der Sozialen Arbeit stärker sein wird als in der Psychotherapie.

Aus der bislang umfassendsten quasi-experimentellen Wirkungsstudie im Bereich der erzieherischen Hilfen im Kontext des Modellprogramms „Wirkungsorientierte Jugendhilfe“ (vgl. Albus et al. 2009; Albus et al. et al. in diesem Band) ist demgegenüber zu entnehmen, dass gut 30 % der Wirkungsvarianz durch partizipative Kontexte, Beziehungsgestaltungen und Arbeitsbündnisse zu erklären ist. Diese ‚unspezifischen‘ Wirkdimensionen haben eine Reihe (‚steuerbarer‘) institutioneller Voraussetzungen und sind professionell gestaltbar. Allerdings ist die Gestaltung dieser Dimensionen fallspezifisch, von intersubjektiven Prozessen und Konstellationen abhängig und insofern kaum sinnvoll manualisierbar.

Ein dritter Kritikstrang macht sich an der Tatsache fest, dass die experimentelle Wirkungsforschung nur standardisierbare Maßnahmen in kontrollierten Kontexten überprüfen kann. Wie ausgeführt, basiert die Experimentalforschung auf der Sicherstellung, dass keine maßnahmeexternen Ereignisse und Kontexte die Wirkungsaussagen konfundieren. Die experimentelle Wirkungsmessung setzt hierzu methodische Verfahren ein, die letztlich dem Versuch gleichkommen, einen bestimmten Wirkzusammenhang von breiteren sozialen und institutionellen Kontexten zu isolieren und damit freilich auch zu dekontextualisieren. Das Bestreben, die Wirkungen von Kontextprozessen und anderen ‚Störvariablen‘ zu eliminieren, kommt in der Sozialen Arbeit einem ‚anti-sozialen Bias‘ nahe (dazu: Hope 2005). Denn die Kontrolle dieser Einflüsse ist letztlich unmöglich, wenn die Interventionen überwiegend fallspezifisch unterschiedlich gestaltete soziale Prozesse sind, die in vielfältigen Kontexten stattfinden, die von den Planungs- und Administrationsstellen nicht vollständig zu kontrollieren sind (vgl. Hammersley 2009).

Während z.B. kognitiv-behaviourale Trainingsmaßnahmen vergleichsweise gut manualisierbar sind, und auch bei einer Eliminierung von Kontexten Wirkung entfalten können, sind typische Vorgehensweisen der Kinder- und Jugendhilfe auf kaum standardisierbare soziale Prozesse, Kontexte und Konstellationen verwiesen. Werden diese als (vermeintlich) programmexterne Bedingungen durch Randomisierungsverfahren eliminiert und die Vorgehensweisen in manualisierter Form standardisiert, dann können solche Maßnahmen gar nicht mehr funktionieren. In diesem Sinne ließe sich der anti-soziale Bias auch als methodisch induzierte fehlerhafte Setzung der Nullhypothese im Sinne eines „Type II error with regard to desirability of social interventions“ (Hope 2005: 275) verstehen.

Die experimentelle Anstrengung „to cancel out difference“ und die Wirkung von ihren Kontexten zu entkoppeln, wird von einigen sozialwissenschaftlichen ForscherInnen daher als absurde Anstrengung beschrieben „to write out what is essential to a program – social conditions favourable to it“ (Pawson/Tilley 1997: 52). Auf die Relevanz sozialer Bedingungskonstellationen haben insbesondere jene Formen der „hard-to-do science“ einzugehen, zu denen im Anschluss an David C. Berliner (2002) auch die Forschungen in der Sozialen Arbeit gerechnet werden können. Die Schwierigkeit dieser Forschungen resultiert daraus, dass sich die Soziale Arbeit im hohen Maße auf partikulare, individual-biografische Problemlagen richtet, die sich innerhalb von lokalen lebensweltlichen Bedingungskonstellationen ereignen und in ihrer Sinnhaftigkeit erschließen. Genau dies limitiert ihre Generalisierung (vgl. Boettcher et al. 2009) und ihre Anschlussfähigkeit an „covering-law“ Theorien, die den erkenntnistheoretischen Referenzrahmen jener experimentellen Erforschungen von Wirksamkeit darstellen, die Kausalität im Anschluss an den Empiristen David Hume als „constant conjunction“ von Ereignissen modellieren.

Diese Einwände legen in keiner Weise nahe, dass Wirkung und Wirkungsforschung kein Gegenstand für die Soziale Arbeit sein soll. Die Herausforderung scheint darin zu bestehen, Wirkungsforschungen zu etablieren, die nicht nur relative Erfolgswahrscheinlichkeit spezifischer Maßnahmen hinsichtlich spezifisch definierter Ziele feststellen, sondern in der Lage sind, empirisch fundiertes Erklärungs- und Reflexionswissen auf Basis einer Analyse von „Kontext-Mechanismus-Ergebnis-Konfigurationen“ herauszuarbeiten (vgl. Pawson/Tilley 1997) bzw. empirisch belastbare Theorien (mittlerer Reichweite) über Wirkungszusammenhänge zu formulieren, die es erlauben, professionelle Entscheidungen wirkungsorientiert zu fundieren. Aufgabe einer solchen Wirkungsforschung wäre es, über Wirkmechanismen aufzuklären, um eine Grundlage für eine empirisch informierte, professionelle Prüfung der Angemessenheit einer Intervention bereit zu stellen. Hierfür ist eine mechanistische Wirkungsforschung erforderlich. Der – inzwischen etablierte – Neologismus ‚mechanistisch‘ kann schnell in die Irre führen: Es geht nicht um ein mechanistisches bzw. mechanisches Denken. Ein Mechanismus¹⁰ ist vielmehr das, was auch als ‚causal power‘ (vgl. Elder-Vass 2010, Sayer 1992, Ziegler 2008) beschrieben wird, d.h. das, was in einer materiellen, (logisch) notwendigen und prüfaren Weise in der Lage ist, ein Geschehen hervorzubringen bzw. zwischen ‚Inputs‘ und ‚Outputs‘ zu vermitteln.

Die Etablierung einer solchen mechanistischen Forschung setzt daher voraus, über den seriellen Wirkungsbegriff einer ‚successionist causality‘ hin-

10 Mechanismen sind „complexes of interacting individuals, usually classified into specific social categories, which generate causal relationships between aggregate-level variables“ (Steel 2004: 59).

ausgehen (dazu: Harré/Madden 1975), der alleine auf der Prüfung phänomenologischer Zusammenhangshypothesen basiert. Das Problem eines Begriffs von Wirkung, der das regelmäßige Zusammentreffen von unmittelbar beobachtbaren Ereignissen fokussiert, besteht darin, dass sich *soziale* Wirklichkeit nicht auf beobachtbare Eigenschaften von Objekten und Ereignissen reduzieren lässt. Forschungen auf Basis einer seriellen Modellierung von Wirkung können letztlich lediglich ‚Black-box-Erklärungen‘ bereitstellen, „aber keine erklärungsdienlichen Mechanismen angeben, welche die beobachteten Zusammenhänge hervorbringen“ (Schmid 2006: 103). Neuere sozialwissenschaftliche Konzeptionen von Wirkung gehen demgegenüber von einem mechanistischen oder ‚generativen‘ Wirkungsbegriff aus. Die Prämisse lautet „X is a cause of Y if and only if there is a mechanism from X to Y“ (Steel 2004: 60). Damit wird impliziert, dass Wirkungen eine „Folge des Wirkens der zugrunde liegenden Mechanismen und nicht schlicht statistische Zusammenhänge [sind]“ (Mackert 2006: 136). Da ein Wirkungsmechanismus unterschiedliche Effekte erzeugen kann und umgekehrt einem bestimmten Effekt verschiedene generative Ursachen zugrunde liegen können, erschließt sich der Zusammenhang von Wirkungsmechanismen und Effekten darüber hinaus erst im Zusammenspiel mit spezifischen sozialen Kontextbedingungen. Wie es Andrew Sayer (1992: 107) formuliert, ist die Beziehung zwischen Mechanismen bzw. „causal powers“ und ihren Effekten „not fixed but contingent; indeed causal powers exist independently from their effects [...] It is in the view of the fact that causal powers contingently related to their conditions that when we activate a mechanism for our own purpose we take care to ensure that the conditions under which it operates are those which will produce the desired effect“. Es reicht demnach „nicht aus Mechanismen [zu] benennen die einem empirischen Ereignis zugrunde liegen [. ... Vielmehr gilt es] auch den Kontext [zu] spezifizieren [...], innerhalb dessen sie entstanden [sind]“ und wirksam werden können (Mackert 2006: 136).

Vor diesem Hintergrund scheint für eine angemessene Form der Evidenzbasierung Sozialer Arbeit die Frage entscheidend zu sein, wie bestimmte soziale Prozesse mit Ressourcenausstattungen, Gelegenheitsstrukturen, Handlungsbefähigungen und Handlungsentscheidungen der AkteurInnen sowie spezifischen sozialen Kontexten (z.B. institutionalisierten Regeln, kulturellen Sinndeutungen, Machtungleichgewichten etc.) interagieren, vor deren Hintergrund sich die Effekte sozial wirksamer generativer Mechanismen realisieren. Dies läuft auf eine Form der Wirkungsforschung hinaus, die darauf zielt, Verursachungsprozesse und -mechanismen in ihrem Kontext zu analysieren und erklärende Deutungsangebote über Wirkungszusammenhänge zu machen. Das Ergebnis dieser Forschung wäre ein empirisch fundiertes, nichtsdestoweniger *theoretisches* Wissen über Prozess-Mechanismus-Wirkungs-Konfigurationen, das einen Hintergrund für professionelle (Ermessens-)Entscheidungen darstellen kann. Es geht also um nicht anderes als die wissenschaftstheoretische und

empirische Fundierung des wirklichkeitswissenschaftlichen Aspekts des Professionswissens (Dewe/Otto 2010) und genau nicht um eine technologische Anleitung professioneller Praxis.

Ohne Zweifel können insbesondere quasi-experimentelle Modellierungen dabei eine wichtige Grundlage für Prüfungen von Kontext-Mechanismus-Ergebnis-Konfigurationen liefern. Sofern sich Wirkungsforschungen aber lediglich auf die Prüfung von bestimmten standardisierbaren Verfahren reduzieren und versuchen, die Einflüsse von Kontexten und Mechanismen nicht systematisch zu erfassen, sondern methodisch zu eliminieren, besteht erstens die Gefahr, eine Reihe gestaltbarer Prozesse systematisch auszuklammern, die für die Wirksamkeit der Praxis (mit-)verantwortlich sind. Zweitens besteht die Gefahr, eine managerielle Standardisierung und Technologisierung der Praxis voranzutreiben, die hinter dem tatsächlichen Potenzial der Sozialen Arbeit zurückfällt, fallspezifisch angemessene und lebensweltlich nachhaltige Leistungen erbringen zu können.

Literatur

- Albus, S., Greschke, H., Klingler, B., Messmer, H., Micheel, H.-G., Otto, H.-U. & Polutta, A. (2009): Abschlussbericht der Evaluation des Bundesmodellprogramms „Wirkungsorientierte Jugendhilfe“. Entwurfsfassung. Universität Bielefeld, Bielefeld.
- Auckenthaler, A. (2000): Die Manualisierung der Psychotherapie: Ziele und Implikationen. In: Hermer, M. (Hg.): Psychotherapeutische Perspektiven am Beginn des 21. Jahrhundert. Tübingen.
- Baird C. & Wagner D. (2000): The Relative Validity of Actuarial- and Consensus-Based Risk Assessment Systems. In: Children and Youth Services Review 22, 839-871.
- Beckmann, C., Otto, H.-U., Richter, M. & Schrödter, M. (Hrsg.) (2004): Qualität in der sozialen Arbeit. Zwischen Nutzerinteresse und Kostenkontrolle. Wiesbaden.
- Berg, M. (1997): Problems and promises of the protocol. In: Social Science and Medicine, 44, 8, 1081–1088.
- Berliner, D. (2002): Comment: Educational Research: The Hardest Science of All In: Educational Researcher 31, 18-20.
- Boruch, R. (1997): Randomized Experiments for Planning and Evaluation: A Practical Guide. Thousand Oaks.
- Böttcher, W., Dicke, J.N. & Ziegler, H. (2009): Erziehungswissenschaft, Bildungspolitik und Bildungspraxis. Anmerkungen zu einem schwierigen Verhältnis. In: Böttcher, W., Dicke, J.N. & Ziegler, H. (Hrsg.): Evidenzbasierte Bildung. Wirkungsevaluation in Bildungspolitik und pädagogischer Praxis. Münster et al.
- Campbell, D. & Stanley, J. (1963): Experimental and Quasi-Experimental Designs for Research. Chicago.
- Cartwright, N. (2007): Are RCTs the Gold Standard? In: BioSocieties, 2, 11-20.

- Castel, R. (1983): Von der Gefährlichkeit zum Risiko. In: Wambach, M. (Hg.): *Der Mensch als Risiko*. Frankfurt a.M.
- Chaffin, M. & Friedrich, B. (2004): Evidence-based treatments in child abuse and neglect. *Children and Youth Services Review*, 26, 1097-1113.
- Chapman, T. & Hough, M. (1998): *Evidence Based Practice: A Guide to Effective Practice*. London.
- Dawes R., Faust, D. & Meehl, P. (1989): Clinical Versus Actuarial Judgment. In: *Science*, 243, 1668-1674.
- Dewe, B. (2009): Reflexive Sozialarbeit im Spannungsfeld von evidenzbasierter Praxis und demokratischer Rationalität – Plädoyer für die handlungslogische Entfaltung reflexiver Professionalität. In: Becker-Lenz, R., Ehlert, G. & Müller, S. (Hrsg.): *Professionalität in der sozialen Arbeit. Standpunkte, Kontroversen, Perspektiven*. Wiesbaden.
- Dewe, B. & Otto, H.-U. (2010): Reflexive Sozialpädagogik. Grundstrukturen eines neuen Typs dienstleistungsorientierten Professionshandelns. In: Thole, W. (Hg.): *Grundriss Soziale Arbeit. Ein einführendes Handbuch*. 3. Auflage. Wiesbaden.
- Eisner, M. & Ribeaud, D. (2008): *Das Zürcher Interventions- und Präventionsprogramm an Schulen 'zipp's' - Eine Zusammenfassung der zentralen Ergebnisse*. Bern.
- Elder-Vass, D. (2010): *The Causal Power of Social Structures*. Cambridge
- Farley, A. et al. 2009: The Challenges of Implementing Evidence Based Practice: Ethical Considerations in Practice, Education, Policy, and Research. In: *Social Work & Society*, 7, 2, 246-259
- Farrington, D. (2003): Methodological quality standards for evaluation research. In: *The Annals of the American Academy of Political and Social Science*, 587, 49-68.
- Farrington, D., Gottfredson, D., Sherman, L. & Welsh, B. (2002): The Maryland Scientific Methods Scale. In: Sherman, L., Farrington, D., Welsh, B. & MacKenzie, D. (Hrsg.): *Evidence-Based Crime Prevention*. London.
- Gambrill, E. (1999): Evidence-based practice: an alternative to authority-based practice. In: *Families in Society: The Journal of Contemporary Human Services*, 80, 4, 341-350.
- Gambrill, E. (2001): Social Work: An Authority-Based Profession. In: *Research on Social Work Practice*, 11, 2, 166-175.
- Graebisch, C. (2007): Kompatibilitätsprobleme zwischen experimenteller Erkenntnisgewinnung und rechtlichem Entscheidungsprogramm. In: Lösel, F., Bender, D. & Jehle, J.-M. (Hrsg.): *Kriminologie und wissenschaftsbasierte Kriminalpolitik. Entwicklungs- und Evaluationsforschung*. Godesberg.
- Gray, M. and McDonald, C. (2006): Pursuing Good Practice: The Limits of Evidence-Based Practice. In: *Journal of Social Work*, 6, 7-20.
- Grove, W., Zald, D., Lebow, B., Snitz, B. & Nelson, C. (2000): Clinical versus mechanical prediction: A meta-analysis. In: *Psychological Assessment*, 12, 1, 19-30.
- Hammersley, M. (2009): What is Evidence for Evidence-Based Practice? In: Otto, H.-U., Polutta, A. & Ziegler, H. (Hrsg.): *Evidence-based Practice - Modernising the Knowledge Base of Social Work?* Opladen, Farmington Hills.
- Harré R. & Madden, E. (1975): *Causal powers*. Oxford.

- Harrison, S. & McDonald, R. (2003): Science, Consumerism and Bureaucracy: New Legitimations of Medical Professionalism'. *International Journal of Public Sector Management*, 16, 2, 110-121.
- Harrison, S., Moran, M. & Wood, B. (2002): Policy emergence and policy convergence: The case of 'scientific-bureaucratic medicine' in the United States and United Kingdom. In: *British Journal of Politics and International Relations*, 1, 1-24.
- Hope, T. (2005): Pretend It Doesn't Work: The 'Anti-Social' Bias in the Maryland Scientific Methods Scale. *European Journal on Criminal Policy and Research* 11, 275-96.
- Kamshall, H. (2010): Risk Rationalities in Contemporary Social Work Policy and Practice. In: *British Journal of Social Work (BJSW Advance Access)* (<http://bjsw.oxfordjournals.org/cgi/reprint/bcp157v1>).
- Klatetzki, T. (2005): Professionelle Arbeit und kollegiale Organisation. In: Klatetzki, T. & Tacke, V. (Hrsg.): *Organisation und Profession*. Wiesbaden.
- LaLonde, R. (1986): Evaluation the econometric evaluations of training programs with experimental data. In: *American Economic Review*, 76, 604-620.
- Lambert, M. & Ogles, B. (2004): The efficacy and effectiveness of psychotherapy. In Lambert, M. (Hg.): *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change*. 5th ed. New York.
- Mackert, J. (2006): *Ohnmächtiger Staat? Über die sozialen Mechanismen staatlichen Handelns*. Wiesbaden.
- Nutley, S. & Homel, P. (2006): Delivering evidence-based policy and practice: lessons from the implementation of the UK Crime Reduction Programme. In: *Evidence and Policy*, 2, 5-26.
- Pawson, R. & Tilley, N. (2009): Otto, H.-U./ Polutta, A./Ziegler, H. (Eds.): *Evidence-based Practice - Modernising the Knowledge Base of Social Work?* Opladen/Farmington Hills
- Pawson, R. & Tilley, N. (1997): *Realistic Evaluation*. London.
- Pignotti, M. & Thyer, B. (2009): Why randomized clinical trials are important and necessary to social work practice. In: Otto, H.-U., Polutta, A. & Ziegler, H. (Hrsg.): *Evidence-based Practice - Modernising the Knowledge Base of Social Work?* Opladen, Farmington Hills.
- Sachverständigenrat für die Konzertierte Aktion im Gesundheitswesen (2003): *Finanzierung, Nutzerorientierung und Qualität. Band I: Finanzierung und Nutzerorientierung. Band II: Qualität und Versorgungsstrukturen. Gutachten 2003. Kurzfassung*. Bonn
- Sackett D., Rosenberg W., Muir Gray J., Haynes R. & Richardson W. (1997): Was ist evidenzbasierte Medizin und was nicht? In: *Münchener Medizinische Wochenschrift*, 139, 644-645.
- Schmid, M. (2006): *Die Logik mechanistischer Erklärungen*. Wiesbaden.
- Otto, H.-U / Schnurr, S. (Hg.) 2000: *Privatisierung und Wettbewerb in der Jugendhilfe – Marktorientierte Modernisierungsstrategien in internationaler Perspektive*. Neuwied
- Sayer, A. (1992): *Method in Social Science: A Realist Approach*. London
- Schrödter, M. (2005): *Wer macht bessere Diagnosen: Der Mensch oder die Maschine? Zu den interventionspraktischen Möglichkeiten und Gefahren diag-*

- nostischer Klassifikationssysteme. Manuskript. Universität Bielefeld. (<http://www.uni-bielefeld.de/paedagogik/agn/ag8/mschroedter.html>).
- Shadish, W., Cook, T. & Campbell, D. (2002): *Experimental and quasi-experimental designs for generalized causal inference*. Boston.
- Steel, D. (2004): Social mechanisms and causal inference. In: *Philosophy of the Social Sciences* 34, 55-78.
- Struzyna, K.-H. (2006): Wirkungsorientierung in den Hilfen zur Erziehung - warum und wofür? Zu Hintergründen und Bedeutung des strategisch-methodischen Ansatzes. In: *Jugendhilfe*, 44, 6, 289-293.
- Tschuschke, V. (2005): Die Psychotherapie in Zeiten evidenzbasierter Medizin. In: *Psychotherapeutenjournal*, 2, 106-115.
- Wampold, B. (2001): The Great Psychotherapy Debate. Models, Methods, and Findings. Mahwah.
- WBP (2009): Methodenpapier des Wissenschaftlichen Beirats Psychotherapie nach § 11 PsychThG. Verfahrensregeln zur Beurteilung der wissenschaftlichen Anerkennung von Methoden und Verfahren der Psychotherapie. Version 2.7 (www.wbpsychotherapie.de/downloads/Methodenpapier2720090709.pdf).
- Webb, S. (2009): Risk, Governmentality and Insurance. The Actuarial Re-Casting of Social Work. In: Otto, H.-U., Polutta, A. & Ziegler, H. (Hrsg.): *Evidence-based Practice. Modernising the Knowledge Base of Social Work?* Opladen, Farmington Hills.
- Westen, D. & Novotny, C.M. (2004): The empirical status of empirically supported psychotherapies: Assumptions, findings, and reporting in controlled clinical trials. In: *Psychological Bulletin*, 130, 631-663.
- White, S. & Stancombe, J. (2003): *Clinical Judgement in the Health and Welfare Professions: Extending the Evidence Base*. Maidenhead.
- Wilken, U. (2002): Die Rückgewinnung einer sozialen Gerechtigkeitsperspektive angesichts von Individualisierung und Ökonomisierung des Sozialen. In: Greving, H. & Gröschke, D. (Hrsg.): *Geistige Behinderung – Reflexionen zu einem Phantom. Ein interdisziplinärer Diskurs um einen Problembegriff*. Bad Heilbrunn.
- Ziegler, H. (2008): Kleine Verteidigung ontologischer Theorien in der Sozialen Arbeit. In: *Widersprüche Heft 108*, 43-52

