

SWP-Studie

Stiftung Wissenschaft und Politik
Deutsches Institut für Internationale
Politik und Sicherheit

Peter Rudolf / Sascha Lohmann

Außenpolitikevaluation im Aktionsfeld Krisenprävention und Friedensaufbau

S 20
Oktober 2013
Berlin

Alle Rechte vorbehalten.

Abdruck oder vergleichbare Verwendung von Arbeiten der Stiftung Wissenschaft und Politik ist auch in Auszügen nur mit vorheriger schriftlicher Genehmigung gestattet.

SWP-Studien unterliegen einem Begutachtungsverfahren durch Fachkolleginnen und -kollegen und durch die Institutsleitung (*peer review*). Sie geben ausschließlich die persönliche Auffassung der Autoren und Autorinnen wieder.

© Stiftung Wissenschaft und Politik, Berlin, 2013

SWP

Stiftung Wissenschaft und Politik
Deutsches Institut für
Internationale Politik und
Sicherheit

Ludwigkirchplatz 3-4
10719 Berlin
Telefon +49 30 880 07-0
Fax +49 30 880 07-100
www.swp-berlin.org
swp@swp-berlin.org

ISSN 1611-6372

Inhalt

- 5 **Problemstellung und Schlussfolgerungen**
- 7 **Krisenprävention und Evaluation**
- 7 Krisen- und Konfliktprävention:
Ansätze und Annahmen
- 10 Der Ruf nach mehr Evaluation
- 13 **Evaluation als angewandte Wissenschaft**
- 13 Evaluationstypen
- 15 Theoretische und methodologische Probleme
- 19 **Evaluation im Bereich Gewaltprävention
und Friedensaufbau:**
- 19 **Eine schwierige Herausforderung**
- 19 Spezifische Probleme
- 22 Begrenzte Erträge
- 25 **Grenzen des Möglichen:**
- 25 **Was von Evaluation erwartet werden kann**

*Dr. Peter Rudolf ist Senior Fellow der Forschungsgruppe
Amerika*

*Sascha Lohmann war wissenschaftlicher Mitarbeiter im
Projekt »Möglichkeiten und Grenzen der Außenpolitik-
evaluation im Aktionsfeld Zivile Krisenprävention,
Konfliktlösung und Friedenskonsolidierung« und ist
Stipendiat in der Forschungsgruppe Amerika*

Außenpolitikevaluation im Aktionsfeld Krisenprävention und Friedensaufbau

Die Prävention von Krisen und gewalthaltigen Konflikten sowie der Aufbau stabiler Friedensstrukturen sind Ziele deutscher Außenpolitik, die sich konzeptionell im Aktionsplan der Bundesregierung »Zivile Krisenprävention, Konfliktlösung und Friedenskonsolidierung« von 2004 niedergeschlagen haben. Wer Krisen vorbeugen, die Ursachen gewalthaltiger Konflikte beseitigen und die Voraussetzungen für einen dauerhaften friedlichen Konfliktaustrag schaffen will, muss wissen, ob Eingriffe von außen diesem Ziel förderlich sind, sei es in Gestalt umfassenderer Programme oder einzelner Projekte. Was funktioniert unter welchen Bedingungen, was nicht? Wie lässt sich das eigene Handeln verbessern, wie der Ressourceneinsatz optimieren? Das sind die Fragen, die das Interesse an Evaluation speisen, wie es seit einiger Zeit vor allem im Deutschen Bundestag geäußert wird.

Dies ist nur allzu verständlich, denn Maßnahmen im Aktionsfeld Krisenprävention sind einerseits an überaus ambitionierten Zielen ausgerichtet. Andererseits beruhen solche Eingriffe zur Steuerung sozialer und politischer Prozesse auf einer schmalen und alles andere als sicheren Wissensgrundlage, wie und unter welchen Bedingungen bestimmte Projekte und Programme tatsächlich wirken.

Verbreitet ist die Erwartung, Evaluation werde zu Lerneffekten und damit zu einer besseren Politik beitragen. Doch kann Evaluation das überhaupt leisten? Es ist schon eine große methodische Herausforderung, die unmittelbare Wirkung einzelner entwicklungspolitischer Projekte rückblickend oder prozessbegleitend einzuschätzen. Noch weit schwieriger wird es, wenn es um die Bewertung von Strategien, Programmen und Projekten geht, mit denen gewaltsame Konflikte verhindert und stabile Friedensstrukturen aufgebaut werden sollen. Ist dies überhaupt möglich und, falls ja, wie?

Vor diesem Hintergrund liefert die Studie eine Orientierungsbasis dafür, wie Möglichkeiten und Grenzen von Außenpolitikevaluation gerade im Aktionsfeld »Zivile Krisenprävention, Konfliktlösung und Friedensaufbau« bewertet werden können. Dabei baut sie auf dem Diskussionsstand in der Evaluationsforschung und den bisherigen Erträgen einschlägiger Evaluationen auf. Es geht im Folgenden nicht um eher

handwerkliche, die Anwendung und Durchführung von Evaluation betreffende Aspekte. Stattdessen wird ausgelotet, was von herkömmlicher Evaluation bei nüchterner Betrachtung grundsätzlich erwartet werden kann und welche Alternativen zur Verfügung stehen.

Methodisch reflektierte Evaluationen sind eine Form angewandter Sozialwissenschaft, die politische Interventionen und deren Programme oder Projekte aus- und bewertet – nicht zwangsläufig, aber idealerweise mit Blick auf ihre Wirksamkeit. Eine Evaluation kann vor Beginn einer Intervention stattfinden, sie kann diese begleiten, um steuernd einzugreifen, und sie kann nach deren Abschluss erfolgen. Außenpolitik-evaluationen gab es lange fast nur in der Entwicklungszusammenarbeit, doch seit gut einem Jahrzehnt finden sie sich auch zu Konfliktprävention und Friedensaufbau. Sie beruhen meist auf einem neopositivistischen Wissenschaftsverständnis: Danach bewertet Evaluation als objektiv angesehene, in quantifizierter Form aufbereitete Fakten und versucht auf dieser Grundlage, als linear begriffene Ursache-Wirkungs-Zusammenhänge zu identifizieren.

Ein solcher Ansatz gerät schnell an seine Grenzen, wenn Interventionen analysiert werden, mit denen komplexe soziale Prozesse beeinflusst werden sollen. Es ist daher kein Zufall, dass Qualität, Aussagekraft und Ertrag selbst finanziell, personell und zeitlich aufwendiger Evaluationen im weiten Feld der Krisenprävention, der Konfliktlösung und des Friedensaufbaus zu wünschen übriglassen, wenn sie denn überhaupt die zugrundeliegenden Wirkungsannahmen und die tatsächlichen Wirkungen zu überprüfen versuchen. Methodisch ist es äußerst schwierig, verlässliche Wirkungsnachweise zu führen, zumal solche, die über unmittelbare, lokal begrenzte Wirkungen auf der Mikroebene hinausgehen und für die übergeordneten Ziele Konfliktvermeidung und Friedensaufbau auf der Makroebene relevant sind. Zwar lassen sich falsche Wirkungsannahmen feststellen, richtige jedoch kaum erfassen. Selbst wenn gewisse kausale Zusammenhänge ansatzweise plausibel gemacht werden können, bleibt die Frage offen: Lässt sich ein Programm, das in einem bestimmten Kontext wirksam war, auf andere Kontexte übertragen, in denen vielleicht die ursprünglich notwendigen unterstützenden Faktoren fehlen? Die Kontextgebundenheit von Interventionen erlaubt keine einfache Verallgemeinerung darüber, ob gewünschte Wirkungen auch an anderen Orten zu erwarten sind. Evaluationen der üblichen Art sagen zudem wenig darüber aus,

ob, wie und warum bestimmte Maßnahmen wirken. Der Beitrag methodisch mehr oder minder anspruchsvoller Projekt- und Programmevaluationen, die handlungsrelevantes Orientierungswissen hervorbringen sollen, bleibt in den allermeisten Fällen gering.

Wenn es bei der Evaluation zuallererst um strategisches Lernen und nicht eher um nachträgliche Rechenschaftslegung gehen soll, bieten sich zwei Alternativen zu herkömmlichen Formen der Evaluation an. Die erste wäre auf der Programm- und Projektebene angesiedelt und als prospektiv-begleitende Programmqualitätseinschätzung angelegt. Sie könnte zwar nicht auf Wirkungseinschätzungen verzichten, würde diese aber nicht als lineare Ursache-Wirkungsketten verstehen, sondern danach fragen, wie Interventionen in einem komplexen Konfliktsystem die treibenden Kräfte beeinflussen. Evaluation in diesem Sinne ist ein anhaltender Prozess, in dem Interventionen den Veränderungen in einem Konfliktsystem situativ angepasst werden.

Die zweite Alternative läge auf Ebene der Strategie und bestände in einer Politikevaluation, die Inhalte und Prozesse mit Blick auf strategische Kohärenz untersucht. Sie würde die erklärten Ziele, die eingesetzten Mittel und die oft unausgesprochenen Annahmen im Aktionsfeld »Zivile Krisenprävention, Konfliktlösung und Friedensaufbau« unter die Lupe nehmen: Sind die Ziele realistisch, die verwendeten Ressourcen angemessen und die Maßnahmen so gebündelt, dass über lokal auftretende Wirkungen hinaus ein relevanter Beitrag zu Krisenprävention und Friedensaufbau erhofft werden kann?

Krisenprävention und Evaluation

Interventionen im Aktionsfeld Krisen- und Konfliktprävention sowie Friedensaufbau folgen meist ehrgeizigen Zielen, verfügen aber oft nur über eine dürftige Wissensgrundlage für politisches Handeln. Dies macht Strategieformulierung, -implementierung und -bewertung schwierig. Aus diesem Grund ist die Erwartung gewachsen, durch mehr Evaluation von Interventionen ließe sich Wissen darüber schaffen, was funktioniert und was nicht, um Strategien nach diesen Erkenntnissen auszurichten.

Krisen- und Konfliktprävention: Ansätze und Annahmen

Zivile Krisenprävention umfasst die Verhütung und Bewältigung von Konflikten sowie die Sicherung von Frieden und bildet seit dem Koalitionsvertrag von 1998 eine ausdrückliche Orientierung deutscher Außenpolitik. Hintergrund waren vor allem die Krisen und Konflikte auf dem Balkan, aber auch der Völkermord in Ruanda. Diese Orientierung spiegelt zudem die wachsende Bedeutung wider, die Konfliktprävention und Friedensaufbau seit den frühen 1990er Jahren im VN-System gewonnen haben.¹ Die konzeptionelle Verankerung in Gestalt des Aktionsplans »Zivile Krisenprävention, Konfliktlösung und Friedenskonsolidierung« von 2004 lässt sich auch als Erfolg der Friedensforschung lesen, aus deren Reihen immer wieder die Forderung zu hören war, der Gewaltprävention einen größeren Stellenwert einzuräumen, ja die zivile Krisenprävention zu einem Markenzeichen deutscher Außenpolitik zu machen.² Im Mittelpunkt stand an-

fänglich die institutionelle Verankerung dieser außenpolitischen Orientierung. Sie entsprach dem in Öffentlichkeit und Eliten dominierenden Selbstverständnis Deutschlands als »Zivilmacht«, die vorrangig mit nichtmilitärischen Mitteln internationale Verantwortung übernimmt.³ Vorherrschend ist ein zyklisches Verständnis von »Prävention, Bearbeitung, Nachsorge«,⁴ bei dem eine Reihe von Instrumenten auf konfliktentschärfende Faktoren zugeschnitten wird. Zu solchen Instrumenten zählen Programme und Projekte in den Bereichen Friedensunterricht, Gewaltpräventionstraining, Wiedereingliederung ehemaliger Kämpfer, Vergangenheitsbewältigung, Presse- und Demokratieförderung, Justiz- und Polizeireform, Berufstraining, Armutsbekämpfung, Bildung sowie regionale Wirtschaftsentwicklung.

Zivile Krisenprävention ist eher eine allgemeine Grundausrichtung als ein konkretes Politikfeld,⁵ eher eine ressortübergreifende Querschnittsaufgabe, ein Aspekt oder Ziel dessen, was mittlerweile vernetzte Sicherheit heißt.⁶ Zu diffus ist das Konzept, zu sehr mangelt es an einer klaren Prioritätensetzung, wo, wann und wie lange Krisenprävention betrieben

den Frieden. Idee – Erfahrungen – Ziele, Opladen 2000, S. 35–47.

³ Wie sehr das Thema moralisch aufgeladen war, ist nachzulesen bei Christian Egbering, *Tagesordnungspunkt »Zivile Konfliktbearbeitung«. Eine Diskursanalyse anhand von Reden im Deutschen Bundestag*, Neumünster 2012.

⁴ Siehe *Antwort der Bundesregierung auf die Kleine Anfrage der Abgeordneten Kerstin Müller (Köln), Omid Nouripour, Volker Beck (Köln), weiterer Abgeordneter und der Fraktion Bündnis 90/Die Grünen*, Deutscher Bundestag, 17. Wahlperiode, Drucksache 17/9318, 16.4.2012.

⁵ »Die Umsetzung der krisenpräventiven Politik der Bundesregierung erfolgt als stetiger Prozess. Dabei ist Krisenprävention kein in sich geschlossener Politikbereich, sondern eine politische Grundausrichtung, die in zahlreiche Politik- und Handlungsfelder der Bundesregierung hineinwirkt und sich in ihnen reflektiert.« *3. Bericht der Bundesregierung über die Umsetzung des Aktionsplans »Zivile Krisenprävention, Konfliktlösung und Friedenskonsolidierung«*, Berichtszeitraum: Mai 2008 bis April 2010, Berlin 2010, S. 6.

⁶ Siehe Andreas Wittkowsky/Jens Philip Meierjohann, *Das Konzept der Vernetzten Sicherheit: Dimensionen, Herausforderungen, Grenzen*, Berlin: Zentrum für Internationale Friedenseinsätze (ZIF), 2011 (AG Vernetzte Sicherheit, Policy Briefing).

¹ Als Überblick und zum Stand der Diskussion zur Jahrtausendwende siehe Volker Matthies, *Krisenprävention. Vorbeugen ist besser als Heilen*, Opladen: Leske + Budrich, 2000.
² Siehe etwa Tobias Debiel/Martina Fischer/Volker Matthies/Norbert Ropers, *Effektive Krisenprävention. Herausforderungen für die deutsche Außen- und Entwicklungspolitik*, Bonn: Stiftung Entwicklung und Frieden (SEF), Juni 1999 (Policy Paper Nr. 12). Zu den Ursprüngen in der Friedensbewegung siehe Hannelore Köhler, *Zivile Konfliktbearbeitung als Friedensstrategie. Chancen und Grenzen*, Oldenburg 2005 (Schriftenreihe des Fritz-Küster-Archivs). Siehe auch Cornelia Brinkmann, *Zivile Konfliktbearbeitung – Friedensfachdienst – Ziviler Friedensdienst. Begriffe und Bezüge*, in: Tilman Evers (Hg.), *Ziviler Friedensdienst – Fachleute für*

werden soll.⁷ Schon der Begriff ist eher ambitioniert als präzise. Verhindert werden sollen demnach nicht nur gewaltsam ausgetragene Konflikte, sondern schon die krisenhafte Zuspitzung jedweder Auseinandersetzungen, die das Risiko eines Gewaltausbruchs bergen. So heißt es im Aktionsplan: »Krisenprävention ist nicht erst für die Phase der unmittelbar bevorstehenden Gewalteskalation wichtig, sondern muss bereits in früheren Phasen einsetzen. Eindämmung und Beilegung gewaltsamer Konflikte wie Friedenskonsolidierung nach dem Ende des Krieges gehören ebenso zur Prävention zukünftiger Kriege.«⁸

Während in Deutschland vor allem von Krisenprävention die Rede ist, wird im internationalen Sprachgebrauch überwiegend der Ausdruck »conflict prevention« verwendet. Unterschieden wird dabei meist zwischen einerseits »direkter« oder »operativer« und andererseits »struktureller« Prävention. Direkte Prävention soll die gewaltsame Eskalation in einer Krisensituation verhindern, strukturelle Prävention soll Bedingungen schaffen, dass gewaltträchtige

Konflikte erst gar nicht entstehen.⁹ Ob nun von Krisen- oder Konfliktprävention gesprochen wird: Beide Begriffe meinen eigentlich dasselbe – die Verhütung gewalttätiger Konflikte in unterschiedlichen Phasen.¹⁰ Dass dies bei entsprechendem Willen möglich und daher geboten sei, ist die selten hinterfragte Prämisse, die dem Präventionsansatz zugrunde liegt. Zweifel an dieser Annahme waren zwar früh zu hören: Zu optimistisch sei es zu glauben, es gebe wissenschaftlich gesichertes Wissen darüber, wann gewaltsame Konflikte ausbrechen; noch optimistischer sei es anzunehmen, man wisse, was getan werden könne, um dies zu verhindern.¹¹ Doch derartige Zweifel änderten nichts daran, dass Konfliktprävention international zu einem normativ unumstrittenen Konzept avancierte.¹²

Prävention als Methode moderner Gesellschaften im Umgang mit Risiken wurde nach Ende des Ost-West-Konflikts zu einem Aktionsfeld der Außenpolitik erhoben. Sie beruht auf den Annahmen, künftige Entwicklungen ließen sich anhand erkennbarer Indikatoren mit hoher Wahrscheinlichkeit prognostizieren, ohne einen Eingriff trete eine solche befürchtete

7 Siehe Daniel Dückers/Andreas Mehler, »Deutsche Außenpolitik und Krisenprävention«, in: Thomas Jäger/Alexander Höse/Kai Oppermann (Hg.), *Deutsche Außenpolitik. Sicherheit, Wohlfahrt, Institutionen und Normen*, 2. Auflage, Wiesbaden: VS Verlag für Sozialwissenschaften, 2011, S. 247–272; als bewertenden Überblick siehe auch Frank A. Stengel/Christoph Weller, »Action Plan or Faction Plan? Germany's Eclectic Approach to Conflict Prevention«, in: *International Peacekeeping*, 17 (Februar 2010) 1, S. 93–107; ferner Christoph Weller, »Zivile Krisenprävention und Konfliktbearbeitung: Politische Herausforderungen und der Aktionsplan der Bundesregierung«, in: Renate Grasse/Wilhelm Nolte/Peter Schlotter (Hg.), *Berliner Friedenspolitik? Militärische Transformation – zivile Impulse – europäische Einbindung*, Baden-Baden: Nomos, 2008 (AFK-Friedensschriften, Bd. 34), S. 109–136.

8 Die Bundesregierung, *Aktionsplan »Zivile Krisenprävention, Konfliktlösung und Friedenskonsolidierung«*, Berlin, 12.5.2004, S. 6. Oder in anderer Formulierung: »Ziel einer vorausschauenden und verantwortlichen Politik muss es jedoch sein, krisenhafte Entwicklungen frühzeitig zu erkennen, ihnen effizient entgegenzuwirken und somit ihre Zuspitzung, ganz besonders aber den Ausbruch gewaltsamer Konflikte wo immer möglich zu verhindern.« Und weiter: »Gewaltsamen Konflikten effizient vorzubeugen kann aber nur nachhaltig gelingen, wenn die Ursachen des Konflikts an ihrer Wurzel beseitigt werden. Eine vorausschauende, präventive Politik muss daher potentielle Konflikte in einem möglichst frühen Stadium erkennen, ihre Ursachen analysieren und ihnen mit geeigneten Maßnahmen entgegenwirken. Hierbei sind in erster Linie zivile Mittel gefordert.« Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung, *Krisenprävention mit zivilen Mitteln. Ein Beitrag zu Frieden und Sicherheit*, Bonn/Berlin 2010, S. 14f.

9 Erfolg und Fehlschlag sind in der direkten und der strukturellen Prävention unterschiedlich zu bestimmen. Bei direkter Prävention lässt sich von Erfolg sprechen, wenn ein bewaffneter Konflikt verhindert werden kann, von Misserfolg, wenn er ausbricht. Strukturelle Prävention ist erfolgreich, wenn sie friedlichen Wandel ermöglicht, und fehlgeschlagen, wenn eine gewaltträchtige Situation entsteht. So Peter Wallensteen/Frida Möller, *Conflict Prevention: Methodology for Knowing the Unknown*, Uppsala: Uppsala University, Department of Peace and Conflict Research, 2003 (Uppsala Peace Research Papers Nr. 7). Die Unterscheidung zwischen strukturellen und direkten Ansätzen lässt sich taxonomisch verfeinern, wenn man zwischen »a priori measures«, die für eine bestimmte Gruppe von Staaten gelten, etwa bestimmten Normen und Regimen, und »ad hoc measures« differenziert, also Initiativen und Programmen für bestimmte ausgewählte Länder. Siehe die Taxonomie von Michael S. Lund, »Conflict Prevention: Theory in Pursuit of Policy and Practice«, in: Jacob Bercovitch/Victor Kremenyuk/I. William Zartman (Hg.), *The Sage Handbook of Conflict Resolution*, Thousand Oaks u.a.: Sage Publications, 2009, S. 287–321 (292).

10 Zur Begrifflichkeit siehe auch Matthies, *Krisenprävention* [wie Fn. 1], S. 29–32.

11 So die Kritik von Stephen John Stedman, »Alchemy for a New World Order: Overselling »Preventive Diplomacy«, in: *Foreign Affairs*, (Mai/Juni 1995).

12 Zur Entwicklung siehe Alice Ackermann, »The Idea and Practice of Conflict Prevention«, in: *Journal of Peace Research*, 40 (2003) 3, S. 339–347; Robert Muggah/Natasha White, *Is There a Preventive Action Renaissance? The Policy and Practice of Preventive Diplomacy and Conflict Prevention*, Oslo: Norwegian Peacebuilding Resource Center, Februar 2013.

Entwicklung ein und mit einem frühzeitigen Eingriff ließen sich diese Risiken verringern. Wirksame Prävention setzt also ein hohes Maß an Wissen voraus, nämlich über die Ursachen bestimmter risikobehafteter Entwicklungen und über die Wahrscheinlichkeit ihrer künftigen Zuspitzung.¹³

Oft erwecken Befürworter einer entschiedenen Präventionspolitik den Eindruck, der Ausbruch gewalthaltiger Konflikte ließe sich früh erkennen; es liege vor allem am politischen Willen, wenn diesen nicht rechtzeitig entgegengewirkt werde. Im Nachhinein lässt sich leicht sagen, die Politik habe trotz vieler Warnungen vor einem Gewaltausbruch nicht gehandelt. Doch Vorhersagen künftiger Entwicklungen sind zutiefst problematisch, gleich ob sie sich auf Modelle stützen, die zahlreiche Indikatoren heranziehen, oder ob sie eher auf Einschätzungen von Experten beruhen.¹⁴ Es gibt zwar etliche Modelle, die den Anspruch erheben, mit beträchtlicher Wahrscheinlichkeit Konflikte voraussagen zu können. Aber politische Entscheidungsträger begegnen solchen Prognosen zu Recht mit Skepsis, denn Kenntnisse über Mechanismen, die Gewaltkonflikte auslösen, stehen auf schwachen Fundamenten.¹⁵ Nicht nur lässt sich über die Gewichtung einzelner Indikatoren streiten. Und nicht nur ist die Datenlage, auf der solche Voraussagen basieren, oft unzulänglich. Modellbasierte Vorhersagen suggerieren auch oftmals mehr, als auf wissenschaftlich redliche Weise überhaupt geleistet werden kann.¹⁶

13 Siehe Ulrich Bröckling, »Vorbeugen ist besser ... Zur Soziologie der Prävention«, in: *Behemoth. A Journal of Civilisation*, 1 (2008) 1, S. 38–48 (38f).

14 Zur Problematik siehe John Brante/Chiara de Franco/Christoph Meyer/Florian Otto, »Worse, Not Better?« *Reinvigorating Early Warning for Conflict Prevention in the Post-Lisbon European Union*, Brüssel: Egmont – The Royal Institute for International Relations, Juni 2011.

15 Es gibt keine gesicherte Basis an Erkenntnissen darüber, wie und unter welchen Bedingungen Konflikte eskalieren und zum Ausbruch von Bürgerkriegen führen. Die Forschung ist sich nur darüber weitgehend einig, dass Bürgerkriege aus Eskalationsprozessen entstehen, einer Konfliktdynamik, an deren Anfang die konfliktbeladene Interaktion zwischen Dissidenten und Staatsapparat steht. Siehe Christian Davenport/David A. Armstrong II/Mark I. Lichbach, *Conflict Escalation and the Origins of Civil War*, College Park: University of Maryland, o.J. [2005].

16 Siehe Frederick Barton/Karin von Hippel/Sabina Sequeira/Mark Irvine, *Early Warning? A Review of Conflict Prediction Models and Systems*, Washington, D.C.: Center for Strategic and International Studies, Februar 2008, S. 11f.

Erfolgreiche Konfliktprävention ist deshalb nicht nur eine Frage der politischen Bereitschaft, bei Frühwarnungen schnell zu handeln. Selbst wenn Gewaltkonflikte einigermaßen verlässlich prognostiziert werden könnten, fehlt es an solidem Wissen, welche Eingriffe in einem konkreten Fall tatsächlich konfliktverhindernd wirken.¹⁷ Der Werkzeugkasten ist zwar bekannt,¹⁸ doch darüber, welche Instrumente in welcher Form und Zusammensetzung wann geeignet sind, gibt es nur wenig Erkenntnisse, an denen sich Politiker und Diplomaten orientieren können.¹⁹ Auch hier kann die Wissenschaft der Politik nur wenig gesichertes Wissen anbieten.²⁰ Gleichzeitig setzt Prävention voraus, dass komplexe soziale Prozesse sich entscheidend beeinflussen lassen. Die implizite Annahme, gelungene Konfliktprävention hänge maßgeblich von den eingesetzten Instrumenten ab, fußt auf dem Glauben, die Welt ließe sich steuern. Intervention

17 Siehe Christoph O. Meyer, »Normative, theoretische und praxeologische Defizite der Friedensforschung am Beispiel der Konfliktprävention«, in: *Zeitschrift für Internationale Beziehungen*, 19 (2012) 1, S. 185–197 (192ff).

18 Als guten Überblick siehe Claudia Major/Tobias Pietz/Elisabeth Schöndorf/Wanda Hummel, *Toolbox Krisenmanagement. Von der zivilen Krisenprävention bis zum Peacebuilding: Prinzipien, Akteure, Instrumente*, Berlin: Stiftung Wissenschaft und Politik/Zentrum für Internationale Friedenseinsätze, 2011.

19 Zum Stand siehe Lawrence Woosler, *Preventing Violent Conflict: Assessing Progress, Meeting Challenges*, Washington, D.C.: United States Institute of Peace, September 2009, S. 11f; Robert Muggah, *Conflict Prevention and Preventive Diplomacy: What Works and What Doesn't?*, New York: International Peace Institute, März 2012; Lund, »Conflict Prevention: Theory in Pursuit of Policy and Practice« [wie Fn. 9]; Oliver Ramsbotham/Tom Woodhouse/Hugh Miall, *Contemporary Conflict Resolution. The Prevention, Management and Transformation of Deadly Conflicts*, 3. Auflage, Cambridge: Polity Press, 2011, S. 138. Wenig hat sich an dem Befund geändert, den Michael Lund vor einiger Zeit so zusammenfasste: »So far, very few existing analyses of specific conflict situations try to, or are able to, back up what they recommend to policy-makers with a sound or at least plausible analysis of whether the recommended actions are likely to work and why.« Michael Lund, »Improving Conflict Prevention by Learning from Experience: Issues, Approaches and Results«, in: Michael Lund/Guenola Rasamoelina (Hg.), *The Impact of Conflict Prevention Policy: Cases, Measures, Assessments*, SWP – Conflict Prevention Network (SWP-CPN), Yearbook 1999/2000, Baden-Baden: Nomos, 2000, S. 63–88 (65). Siehe ferner Erik Melander/Claire Pigache, »Conflict Prevention: Concepts and Challenges«, in: Walter Feichtinger/Predrag Jureković (Hg.), *Konfliktprävention zwischen Anspruch und Wirklichkeit*, Wien: Landesverteidigungsakademie, 2007, S. 9–17.

20 Zu dieser zutreffenden Einschätzung allgemein siehe Hanns W. Maull, »Wissenschaftliche Außenpolitik-Evaluation: Ein Oxymoron? Eine Replik auf Peter Rudolf«, in: *Zeitschrift für Internationale Beziehungen*, 15 (2008) 1, S. 113–123 (116).

nen zum Zweck der Krisenprävention sind jedoch keineswegs bloß technisch-administrative Hilfestellungen zur Friedenssicherung, sondern politische Eingriffe in einen Prozess, in dem zahlreiche innerstaatliche Akteure um Einfluss auf die zukünftige Ordnung ringen.²¹

Die Auffassung, politische Prozesse ließen sich von außen lenken, ist besonders für die »strukturelle« Prävention wichtig, also wenn angestrebt wird, die »Grundursachen« (root causes) gewaltträchtiger Konflikte zu beseitigen und die Voraussetzungen für einen dauerhaften friedlichen Konfliktaustrag zu schaffen. Die Herausbildung von »good governance«, Rechtsreformen und der Aufbau demokratischer Institutionen erfordern tiefe Eingriffe in politische Systeme, ja unter Umständen eine »radikale Rekonstruktion« anderer Gesellschaften.²² Ökonomische Entwicklung und der Aufbau demokratischer Strukturen gelten zwar weithin als Kern struktureller Konfliktprävention. Doch gerade während Demokratisierungsbemühungen kann sich das Potential gewaltsamer Konflikte erhöhen.²³ Der Rückgriff auf Gewalt ist auch eine strategische Entscheidung von Regierenden oder Aufständischen in bestimmten Situationen.²⁴

Der Ruf nach mehr Evaluation

Die Prävention von Krisen und gewalthaltigen Konflikten in anderen Gesellschaften ist ein ehrgeiziges Unterfangen, das auf einer unsicheren Wissensbasis operieren muss. Es ist daher nicht verwunderlich, dass sich mit der Zunahme von Programmen und Projekten zu Krisenprävention und Friedensaufbau im Laufe des letzten Jahrzehnts international eine Debatte da-

rüber entwickelt hat, ob und wie sich Aktivitäten in diesen Bereichen bewerten lassen. Durch den Aufstieg des Paradigmas der evidenzbasierten Politik,²⁵ zuerst bei nichtstaatlichen Geberorganisationen,²⁶ kam neuer Optimismus auf, mit Hilfe von Evaluation lasse sich gesichertes Orientierungswissen erzeugen.²⁷ Evidenzbasierte Politik fußt auf dem Ideal eines wissenschaftlich begründeten rationalen Mitteleinsatzes. Gute Politik ist aus dieser Sicht dadurch gekennzeichnet, dass sie diesem Ideal entspricht und nicht das Ergebnis eines von Pragmatismus und Kompromissen geprägten politischen Prozesses ist.²⁸

Evaluation ist seit geraumer Zeit in vielen Politikbereichen fest etabliert, insbesondere in der Sozial-, Bildungs- und Umweltpolitik. Außenpolitik dagegen blieb lange international ein weißer Fleck auf der Landkarte der Politikevaluation.²⁹ Institutionalisierte

25 Der Ansatz der evidenzbasierten Politik – der Begriff »evidenzbasiert« wurde ursprünglich vor allem in der Medizin verwendet – spielt vor allem in Großbritannien seit Mitte der 1990er Jahre eine große Rolle. Er hat sich aber auch innerhalb der EU niedergeschlagen, in der prospektive Folgenabschätzungen und restrospektive Evaluationen in unterschiedlichen Politikfeldern institutionalisiert sind. Zur Herkunft des Ansatzes und zur Situation in Deutschland siehe Uwe Jun/Karsten Grabow, *Mehr Expertise in der deutschen Politik? Zur Übertragbarkeit des »Evidence-based policy approach«*, Gütersloh: Bertelsmann Stiftung, 2008.

26 Die Wirkungsorientierung ist in der Arbeit gemeinnütziger Nichtregierungsorganisationen länger verankert als in der staatlichen Außenpolitik und gab dieser wichtige Impulse. Siehe das Interview mit dem Chef der US-amerikanischen Entwicklungshilfeorganisation USAID, der fast zehn Jahre bei der Bill-und-Melinda-Gates-Stiftung gearbeitet hat. Rahim Kanani, »USAID Chief Rajiv Shah Talks Philanthropy, Innovation, Transparency and More«, in: *Forbes* (online), 19.6.2012, <www.forbes.com/sites/rahimkanani/2012/06/19/usaaid-chief-rajiv-shah-talks-philanthropy-innovation-transparency-and-more/>.

27 Siehe *When Will We Ever Learn? Improving Lives through Impact Evaluation*, Washington, D.C.: Center for Global Development, Mai 2006 (Report of the Evaluation Gap Working Group); zur Forderung nach mehr Evidenzbasierung in der Entwicklungspolitik siehe OECD, *Evaluating Development Activities. 12 Lessons from the OECD DAC*, Paris 2013.

28 Zu dieser Gegenposition siehe grundsätzlich Charles E. Lindblom, »The Science of »Muddling Through««, in: *Public Administration Review*, 19 (1959) 2, S. 79–88.

29 Siehe Peter Rudolf, »Außenpolitikevaluation. Konzeptuelle Überlegungen zu einem vernachlässigten Bereich der Außenpolitikanalyse«, in: *Zeitschrift für Internationale Beziehungen*, 14 (2007) 2, S. 319–330. Für eine Auflistung möglicher Gründe siehe Thomas Widmer, »Evaluation in der Außenpolitik: Gründe für eine Evaluationslücke«, in: *Zeitschrift für Internationale Beziehungen*, 15 (2008) 1, S. 125–137 (130–134). Widmer argumentiert weiterhin, die »Evaluationslücke«

21 Siehe Jonathan Goodhand/Oliver Walton, »The Limits of Liberal Peacebuilding? International Engagement in the Sri Lankan Peace Process«, in: *Journal of Intervention and Statebuilding*, 3 (2009) 3, S. 303–323 (beonders 314).

22 Siehe Eli Stamnes, *The Responsibility to Protect and »Root Cause« Prevention*, Oslo: Norwegian Institute of International Affairs, 2010, S. 23.

23 Zu dieser Problematik der Prävention siehe Sven Chojnacki/Wolf-Dieter Eberwein, *Die Kultur der Prävention. Ein Ansatz zur Zivilisierung internationaler Politik?*, Berlin: Wissenschaftszentrum Berlin für Sozialforschung, März 2000, S. 5ff; zu den destabilisierenden Folgen politischer und wirtschaftlicher Liberalisierung siehe Roland Paris, *At War's End. Building Peace After Civil Conflict*, Cambridge: Cambridge University Press, 2004.

24 Siehe Richard Gowan/Bruce D. Jones/Sara Batmanglich/Andrew Hart, *Back to Basics: The UN and Crisis Diplomacy in an Age of Strategic Uncertainty*, New York: New York University, Center on International Cooperation, Juli 2010, S. 11ff.

Evaluation in der Außenpolitik beschränkte sich bis Mitte des ersten Jahrzehnts des 21. Jahrhunderts vielerorts weitgehend auf die Entwicklungspolitik.³⁰ So sind Evaluationen in der Entwicklungszusammenarbeit westlicher Geberländer seit langem selbstverständlich. Doch auch in der Entwicklungspolitik mangelt es an soliden Analysen, die der Frage nachgehen, ob Veränderungen tatsächlich auf die untersuchten Interventionen zurückzuführen sind.³¹

Anfänglich lehnte sich die Evaluation von Krisenpräventionspolitik stark an den Ansatz in der Entwicklungszusammenarbeit an, der sich vor allem auf quantitative Methoden stützt.³² Inzwischen hat sich dieser Ansatz verbreitert und immer häufiger werden auch qualitative Methoden verwendet. In den letzten Jahren wurde auf supranationaler Ebene einiges unternommen, um im weiten Feld der Krisenprävention, Konfliktlösung und Friedenskonsolidierung einen einheitlichen Standard für die Evaluationsmethodik zu setzen.³³

Auch in Deutschland ist der Ruf nach verstärkter Evaluierung der Instrumente zur Krisenprävention lauter geworden, sowohl in der Zivilgesellschaft als auch im Bundestag.³⁴ So wurde in einem Kommentar

beziehe sich auf die Analyse von Wirkungen und nicht auf den Evaluationsansatz per se, da Politikformulierung und -implementation bereits Gegenstand von Bewertung seien.

30 Siehe etwa Reinhard Stockmann, *Evaluation staatlicher Entwicklungspolitik*, in: Reinhard Stockmann (Hg.), *Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder*, 2., überarbeitete Auflage, Opladen: Leske + Budrich, 2004, S. 375–410.

31 Siehe *When Will We Ever Learn?* [wie Fn. 27]; Asian Development Bank, *Impact Evaluation. Methodological and Operational Issues*, Manila 2006; skeptisch zur Qualität mancher Evaluationen Paul Clements/Thomaz Chianca/Ryoh Sasaki, »Reducing World Poverty by Improving Evaluation of Development Aid«, in: *American Journal of Evaluation*, 29 (Juni 2008) 2, S. 195–214.

32 Siehe Claus C. Rebien, »Development Assistance Evaluation and the Foundations of Program Evaluation«, in: *Evaluation Review*, 21 (1997) 4, S. 438–460.

33 Siehe OECD, *Evaluating Peacebuilding Activities in Settings of Conflict and Fragility. Improving Learning for Results*, Paris 2012 (DAC Guidelines and Reference Series).

34 Erste Diskussionen fanden ihren Niederschlag in Jörg Calließ (Hg.), *Evaluation in der zivilen Konfliktbearbeitung? Tun wir das, was wir tun, richtig? Tun wir das Richtige?*, Rehburg-Loccum 2006 (Loccumer Protokolle 14/05). Im Aktionsplan selbst heißt es zum Thema Evaluation: »Der Beirat für zivile Krisenprävention kann Evaluierungsempfehlungen in Auftrag geben. Deren Ergebnisse werden bei der weiteren Umsetzung des Aktionsplanes berücksichtigt.« Bundesregierung, *Aktionsplan* [wie Fn. 8], S. 67. Evaluation des Politikfeldes oder Formulierung eines realistischen Strategieansatzes – das sind die beiden Alternativen, die gegenwärtig in der politischen und zivilgesellschaftlichen Diskussion zu vernehmen sind. Siehe

der Gemeinsamen Konferenz Kirche und Entwicklung zum 2. Bericht der Bundesregierung über die Umsetzung des Aktionsplans »Zivile Krisenprävention, Konfliktlösung und Friedenskonsolidierung« bemängelt, es fehlten »Evaluationsverfahren und Wirkungsanalysen«. ³⁵ Auch im wissenschaftlichen Umfeld heißt es immer wieder, Erfolge und Misserfolge der Krisenprävention und des Friedensaufbaus müssten evaluiert werden, um einigermaßen gesichertes Wissen darüber zu erlangen, welche Instrumente unter welchen Bedingungen wirken. ³⁶

Ähnliche Stimmen sind aus dem Bundestag zu vernehmen. Die Fraktion Bündnis 90/Die Grünen kritisiert in einem Antrag vom 25. Mai 2011, es gebe keine »umfassende und systematische Erfassung und Auswertung der Aktivitäten im Bereich der Zivilen Krisenprävention. Das Monitoring, die Wirkungs- und Fortschrittsanalyse von Maßnahmen, ist mangelhaft«. Die Fraktion fordert, »mit der Evaluation der Infrastruktur zur Krisenprävention zu beginnen und ein einheitliches Konzept zur Überprüfung und Überwachung (Monitoring) ziviler Krisenprävention, ihrer Instrumente und Strukturen zu entwickeln«. ³⁷ In einem Antrag der SPD-Fraktion vom 25. Januar 2011, in dem sie eine kohärente Strategie für die zivile Krisenprävention anmahnt, ist die Rede von der Notwendigkeit eines »Evaluierungskonzeptes, um Fortschritte und Erfolge zu messen und dadurch Erkenntnisse für die Weiterentwicklung der zivilen Krisenprävention zu gewinnen und ihre politische und gesellschaftliche Akzeptanz zu erhöhen«. ³⁸ Im »3. Bericht der Bundesregierung über die Umsetzung des Aktionsplans »Zivile Krisenprävention, Konfliktlösung und

den Bericht über eine Tagung zum Thema »Evaluierung der zivilen Krisenprävention«: Jan-Thilo Klimisch, »Maßstäbe für den Erfolg. Tagung der Evangelischen Akademie Loccum«, 7.–9. Dezember 2012, in: *Wissenschaft und Frieden*, (2013) 1, S. 61f.

35 Gemeinsame Konferenz Kirche und Entwicklung (Hg.), *Vertrauen auf die Kraft des Zivilen. Kommentar zum 2. Bericht der Bundesregierung über die Umsetzung des Aktionsplans »Zivile Krisenprävention, Konfliktlösung und Friedenskonsolidierung«*, Bonn/Berlin, 25.9.2008, S. 26.

36 So Volker Matthies, *Wie Kriege verhindert werden können – Bestandsaufnahme der Debatte um Konfliktprävention*, Vortrag an der Universität Hamburg, 25.11.2010.

37 Antrag der Abgeordneten Kerstin Müller u.a. und der Fraktion Bündnis 90/Die Grünen: *Zivile Krisenprävention ins Zentrum deutscher Außenpolitik rücken*, Deutscher Bundestag, 17. Wahlperiode, Drucksache 17/5910, 25.5.2011, Zitate S. 3, 6.

38 Antrag der Abgeordneten Edelgard Bulmahn u.a. und der Fraktion der SPD: *Deutschland braucht dringend eine kohärente Strategie für die zivile Krisenprävention*, Deutscher Bundestag, 17. Wahlperiode, Drucksache 17/4532, 25.1.2011, S. 3.

Friedenskonsolidierung« heißt es, erste Erfahrungen aus der Evaluierung solcher Initiativen sollten stärker als bisher für die weitere Ausgestaltung krisenpräventiver Aktivitäten der Bundesregierung ausgewertet werden.³⁹ Der Unterausschuss Zivile Krisenprävention und vernetzte Sicherheit hat gegen Ende der 17. Legislaturperiode seine Empfehlung erneuert, »dass eine systematische Auswertung der Aktivitäten im Bereich der zivilen Krisenprävention und Konfliktbearbeitung durchgeführt wird, einschließlich einer wissenschaftlichen Evaluierung. Dies betrifft v.a. das Monitoring sowie die Wirkungs- und Fortschrittsanalyse von Maßnahmen. Dafür bedarf es transparenter und objektiver Kriterien, um Fehlentwicklungen rechtzeitig aufdecken und die notwendigen politischen Kurskorrekturen vornehmen zu können.«⁴⁰

Allgemein wird offenbar erwartet, Evaluation führe automatisch zu Lerneffekten und damit zu einer besseren Politik in einem Aktionsfeld, das auf einer sehr dünnen Wissensgrundlage ruht. Dieser Befund, nämlich das geringe Maß an handlungsrelevantem Wissen, wird auf politischer Ebene in Deutschland übrigens selten so offen reflektiert wie in Großbritannien, wo das Thema evidenzbasierte Politik und damit auch Evaluation eine weit größere Rolle spielen als hierzulande.⁴¹

39 3. Bericht der Bundesregierung über die Umsetzung des Aktionsplans »Zivile Krisenprävention, Konfliktlösung und Friedenskonsolidierung« [wie Fn. 5], S. 11.

40 Abschlussdokument des Unterausschusses Zivile Krisenprävention und vernetzte Sicherheit, S-17(3)73_UA ZivKri, Berlin, 10.6.2013, <www.bundestag.de/bundestag/ausschuesse17/a03/a03_ga/Abschluss.pdf>.

41 So heißt es in einem zentralen Strategiedokument: »... the overall evidence base and conceptual foundations for engagement in fragile states remain patchy, underdeveloped and, in some areas, contested.« Department for International Development/Foreign and Commonwealth Office/Ministry of Defence, *Building Stability Overseas Strategy*, London 2011, S. 34. Mit diesem strategischen Konzept sollen vor allem gewaltsame Konflikte verhindert und stabile legitime politische Strukturen aufgebaut werden.

Evaluation als angewandte Wissenschaft

Systematische, methodisch reflektierte Evaluation ist angewandte Sozialwissenschaft.⁴² Es handelt sich dabei um eine mit wissenschaftlichen Methoden durchgeführte Aus- und Bewertung zielgerichteter politischer Eingriffe (Interventionen) eines staatlichen oder nichtstaatlichen Akteurs.⁴³ Evaluation ist am übergeordneten Ziel ausgerichtet, problemlösungsorientierte Politiken zu verbessern und Handlungsalternativen zu eröffnen.⁴⁴ Idealtypisch zugespitzt: Nicht um Erklärungswissen geht es, sondern um Orientierungswissen für operatives Handeln.⁴⁵

Evaluation benötigt klare Kriterien. Dabei kann es sich um ein normatives Kriterienraster handeln, aber auch um praktische Ziele wie Relevanz, Angemessenheit, Effektivität, Effizienz oder Nachhaltigkeit, wie sie das Development Assistance Committee (DAC) der Organisation for Economic Co-Operation and Development (OECD) ursprünglich für die Evaluation von

Entwicklungspolitik formuliert hat und die Eingang in den einschlägigen Leitfaden für die Evaluation friedensschaffender Aktivitäten gefunden haben.⁴⁶ Entscheidend ist, dass sich die Forschung auf die Wirkungen statt auf die Formulierung von Politik konzentriert.⁴⁷ Diese kann jedoch insofern von Bedeutung sein, als sich erst durch die Analyse des Prozesses die Ziele einer Intervention klar erkennen lassen.⁴⁸

Evaluationstypen

Evaluation kann an verschiedene Phasen im Politikzyklus anknüpfen.⁴⁹ Hier lassen sich drei Typen von Evaluationen identifizieren, die wiederum je nach Erkenntnisinteresse unterschiedliche inhaltliche Ausrichtungen aufweisen können.⁵⁰

42 Der Begriff Evaluation oder Evaluierung, wie es im Deutschen häufig (etwa bei der GIZ) heißt, bezieht sich auf drei Referenzebenen: Auf der ersten Ebene steht Evaluation für ein Denkmodell im Sinne einer nachvollziehbaren und nachprüfbar bewerteten, auf der zweiten bezeichnet sie die konkrete Durchführung und Anwendung und auf der dritten das Urteil von Evaluationsvorgängen. Siehe Helmut Kromrey, *Evaluation – ein Überblick*, in: Heidrun Schöch (Hg.), *Was ist Qualität? Die Entzauberung eines Mythos*, Berlin: Wissenschaftlicher Verlag, 2005 (Schriftenreihe Wandel und Kontinuität in Organisationen, Band 6), S. 31–85 (31).

43 Für eine Unterscheidung politischer Eingriffe und Aktivitäten im Allgemeinen siehe Widmer, »Evaluation in der Außenpolitik« [wie Fn. 29], S. 125.

44 Siehe Ulrich Klöti, »Inhaltliche und methodische Anforderungen an wissenschaftliche Politikevaluationen«, in: Werner Bussmann/Ulrich Klöti/Peter Knoepfel (Hg.), *Einführung in die Politikevaluation*, Basel/Frankfurt am Main: Helbing & Lichtenhahn Verlag, 1997, S. 39–57; umfassend Alan Clarke/Ruth Dawson, *Evaluation Research. An Introduction to Principles, Methods and Practice*, Thousand Oaks u.a.: Sage Publications, 1999; Daniel L. Stufflebeam/Anthony J. Shinkfield, *Evaluation Theory, Models, and Applications*, San Francisco: Jossey-Bass, 2007; zu den methodologischen Fragen siehe insbesondere E. Jane Davidson, *Evaluation Methodology Basics. The Nuts and Bolts of Sound Evaluation*, Thousand Oaks u.a.: Sage Publications, 2005.

45 Zur Unterscheidung zwischen Erklärungs- und Orientierungswissen im Kontext politischer Strategieanalyse siehe Ralf Tils, *Politische Strategieanalyse. Konzeptionelle Grundlagen und Anwendung in der Umwelt- und Nachhaltigkeitspolitik*, Wiesbaden: Verlag für Sozialwissenschaften, 2005, S. 34ff.

46 Siehe OECD, *Evaluating Peacebuilding Activities* [wie Fn. 33].

47 Für einen umfassenden Überblick über die Entwicklung verschiedener Stränge der Evaluationsforschung siehe Marvin C. Alkin (Hg.), *Evaluation Roots. A Wider Perspective of Theorists' Views and Influences*, 2. Auflage, Thousand Oaks u.a.: Sage Publications, 2013.

48 Siehe Klöti, *Inhaltliche und methodische Anforderungen* [wie Fn. 44], S. 46ff. Im Folgenden geht es also um Strategie- oder Politikevaluation, nicht aber um Rollenevaluation. Letztere fragt systematisch-bewertend danach, wie staatliche Akteure bestimmte Rollen wahrnehmen, etwa diejenigen, die mit der EU-Ratspräsidentschaft einhergehen. Siehe dazu Adriaan Schout/Sophie Vanhoonacker, »Evaluating Presidencies of the Council of the EU: Revisiting Nice«, in: *Journal of Common Market Studies*, 44 (2006) 5, S. 1051–1077.

49 Im zyklischen Politikmodell (policy cycle) durchläuft eine Intervention idealtypisch verschiedene Phasen: Problemdefinition, Setzen der politischen Agenda, Politikformulierung, Politikimplementierung und schließlich Reaktion bzw. Beendigung. Der Zyklus beginnt von neuem, wenn die Intervention abgeschlossen ist und die in ihrem Verlauf gemachten Erfahrungen in die Formulierung einer neuen eingehen. Siehe Werner Jann/Kai Wegrich, *Phasenmodelle und Politikprozesse: Der Policy Cycle*, in: Klaus Schubert/Nils C. Bandelow (Hg.), *Lehrbuch der Politikfeldanalyse*, München/Wien 2003, S. 71–104.

50 Als allgemeines methodisches Kompendium siehe European Commission, *Evalsed Sourcebook: Method and Techniques*, <http://ec.europa.eu/regional_policy/sources/docgener/evaluation/guide/evaluation_sourcebook.pdf>; als Leitfaden im Bereich Friedensaufbau siehe Thania Paffenholz/Luc Reychler, *Aid for Peace. A Guide to Planning and Evaluation for Conflict Zones*,

1. *Ex-ante-Evaluationen* setzen vor Beginn einer Intervention an, um die Ausgestaltung und Zielsetzung ihrer jeweiligen Instrumente (Programme und Projekte) zu analysieren. Geht es um Konfliktprävention oder Friedenskonsolidierung, so müssen in solch einer Designevaluation zunächst die Analyse der Konfliktursachen und die für deren Beseitigung oder Minderung eingesetzten Instrumente sowie deren Zielsetzung überprüft werden.⁵¹ Dabei sind vor allem die logische Kohärenz und die empirische Konsistenz des zugrundeliegenden *Wirkungsmodells* zu bewerten. Interventionen, die etwas verändern oder beibehalten wollen, beruhen ausdrücklich oder oft auch unausgesprochen auf einem Wirkungsmodell.⁵² Ein solches Modell, in der englischsprachigen Literatur auch als *theory of change* bezeichnet, bildet das Fundament einer jeden Programmtheorie. Es beschreibt den angenommenen Zusammenhang zwischen einer Intervention und ihrer Zielsetzung und trifft Aussagen über Notwendigkeit, Adressaten und Ablauf des Eingriffs.⁵³ Logisch *kohärent* ist ein Wirkungsmodell dann, wenn sich angesichts konzeptioneller Ausrichtung der Instrumente, zugrundeliegender kausaler Annahmen und eingesetzter Ressourcen die angestrebten Wirkungen zumindest erwarten lassen. Empirisch *konsistent* ist es, wenn es durch Befunde aus der Wirklichkeit gestützt wird oder die Erfahrung dem Wirkungsmodell zumindest nicht entgegensteht.⁵⁴

Die Designevaluation in Form einer prospektiven Programmqualitätseinschätzung erhebt den Anspruch, die konzeptionelle Qualität einer Intervention bewerten zu können, noch bevor diese stattfindet.⁵⁵ Dafür muss die Intervention allerdings so angelegt werden, dass eine methodisch akzeptable begleitende

oder abschließende Evaluation überhaupt möglich erscheint.⁵⁶

2. *Formative* oder *begleitende Evaluationen* setzen während einer Intervention selbst an, um so Rückschlüsse auf deren Wirkungen zu ermöglichen und diese Erkenntnisse in den Implementierungsprozess einzuspeisen. Damit können mögliche Änderungen oder Anpassungen bereits kurz nach Beginn einer Intervention vorgenommen werden. Betrachtet wird dabei die unmittelbare und mittelbare Umsetzung einer Intervention. Diese Art der Evaluation überschneidet sich sowohl inhaltlich als auch organisatorisch mit der begleitenden Durchführungskontrolle (monitoring) und der Rechenschaftslegung (accountability), die Effekte einer Intervention sammeln und dadurch ihren Verlauf dokumentieren, jedoch nicht systematisch aus- und bewerten.⁵⁷ In einer begleitenden Evaluation lässt sich die Programmtheorie anhand der kurz- bis mittelfristig erzeugten Wirkungen einer Intervention überprüfen. Sofern das Erkenntnisinteresse auf dem Lernprozess liegt, überlappt sich diese Form mit den »best practice«- und »lessons learned«-Formaten (auch Quasi-Evaluation). Sowohl Ex-ante- als auch formative Evaluationen zielen darauf ab, schon vor und während des Verlaufs, aber noch vor Beendigung einer Intervention zu deren Qualitätsentwicklung beizutragen.

3. *Ex-post-* oder *summative Evaluationen* erfolgen nach dem Abschluss einer Intervention. Diese wird abhängig vom Erkenntnisinteresse vor dem Hintergrund aller mit der Intervention in Zusammenhang gebrachter Effekte betrachtet. Eine summative Evaluation bewertet die kurz-, mittel- und langfristigen Wirkungen, die Wirksamkeit (effectiveness), den Nutzen sowie den Grad der Zielerreichung (success) einer Intervention. Unterteilen lassen sich die Gesamtergebnisse einer Intervention in die materiellen und immateriellen Effekte ihrer unmittelbaren Umsetzung (output), die sich mittelbar daraus ergebenden Ergebnisse (outcomes) und die erwarteten oder unerwarteten, beabsichtigten und unbeabsichtigten⁵⁸ sowie kurz-,

Baden-Baden: Nomos, 2007.

⁵¹ Siehe Michael Quinn Patton/Patricia A. Patrizi, »Strategy as the Focus for Evaluation«, in: Patricia A. Patrizi/Michael Quinn Patton (Hg.), *Evaluating Strategy*, San Francisco: Jossey-Bass, 2010 (New Directions for Evaluation Nr. 128), S. 5–28.

⁵² Siehe Peter Knoepfel/Frédéric Varone/Werner Bussmann/Luzius Mader, »Evaluationsgegenstände und Evaluationskriterien«, in: Bussmann/Klöti/Knoepfel, *Einführung in die Politikevaluation* [wie Fn. 44], S. 78–118.

⁵³ Für eine Übersicht in Bezug auf sozialen Wandel siehe Sue C. Funnell/Patricia J. Rogers, *Purposeful Program Theory. Effective Use of Theories of Change and Logic Models*, San Francisco 2011, S. 319–349 (Kapitel 11).

⁵⁴ Knoepfel/Varone/Bussmann/Mader, »Evaluationsgegenstände und Evaluationskriterien« [wie Fn. 52], S. 79–83.

⁵⁵ Siehe Cordula Reimann/Diana Chigas/Peter Woodrow, *An Alternative to Formal Evaluation of Peacebuilding: Program Quality Assessment*, Cambridge, MA: CDA Collaborative Learning Projects, Reflecting on Peace Practice Program, Oktober 2012.

⁵⁶ Siehe Cordula Reimann, *Evaluability Assessments in Peacebuilding Programming*, Cambridge, MA: CDA Collaborative Learning Projects, Reflecting on Peace Practice Program, 2012.

⁵⁷ Dennoch sind die Trennlinien eher unscharf. Siehe auch Falk Tettweiler, *Lernen in Interventionen? Evaluation am Beispiel der deutschen Afghanistan-Mission*, Berlin: Stiftung Wissenschaft und Politik, September 2011 (SWP-Studie 22/2011), S. 6.

⁵⁸ Wie etwa das Einschleppen der Cholera während der UN-Mission nach dem Erdbeben auf Haiti im Jahr 2010. Siehe Jonathan M. Katz, *The Big Truck That Went By. How the World*

mittel- und langfristigen Wirkungen (impact⁵⁹). Die Topografie des Wirkungsfeldes lässt sich in Mikro-, Meso- und Makroebene gliedern.⁶⁰ In der für die Ex-post-Evaluation zentralen Wirkungsanalyse können demnach alle mit einer Intervention logisch in Zusammenhang stehenden Effekte einer bewertenden Gesamtschau unterzogen werden. Diese liefert die Grundlage für die Wirksamkeitsanalyse, die alle identifizierten Wirkungen anhand der verfolgten Ziele misst, um Aussagen über die Effektivität zu treffen.⁶¹ Die Wirksamkeitsanalyse ist ein maßgebliches Instrument zur Bewertung, inwieweit die Ziele erreicht wurden, und damit zur Erfolgskontrolle. Als Maßstab von Effektivität können die gegenüber der Öffentlichkeit erklärten Ziele⁶² oder bestimmte Wertvorstellungen dienen. Die prominentesten Methoden zur Untersuchung der Zielerreichung sind der Vorher-Nachher-Vergleich des Soll-Zustands mit dem Ist-Zustand, die kontrafaktische Analyse im Gedankenexperiment sowie der quasi-experimentelle Gruppenvergleich.⁶³ Idealerweise gibt die Wirkungsanalyse Aufschluss über den komparativen Nutzen. Er wird im Vergleich zu den materiellen und immateriellen Kosten aller in Erwägung gezogener Interventionsarten bestimmt und von manchen als ideale Bewertungs- und Entscheidungsgrundlage für oder gegen Interventionen angesehen.⁶⁴

Came to Save Haiti and Left Behind a Disaster, Basingstoke 2010.
59 Der Begriff »impact« ist nicht einheitlich definiert. Meist werden damit alle Wirkungen auf die übergeordneten Ziele bezeichnet.

60 Siehe Gesellschaft für Evaluation e.V., Arbeitskreis »Evaluation von Entwicklungspolitik« in der DeGEval (Hg.), *Wirkungsbeobachtung und Evaluierung bei der Förderung von Demokratie und Good Governance. Leitfaden für Geberinstitutionen und GutachterInnen*, Oktober 2005, S. 6

61 Siehe Martin Quack, *Wirkungsanalysen in der zivilen Konfliktbearbeitung*, in: Christoph Weller (Hg.), *Zivile Konfliktbearbeitung. Aktuelle Forschungsergebnisse*, Duisburg: Institut für Entwicklung und Frieden, 2007 (INEF-Report Nr. 85/2007), S. 35–43 (36).

62 Siehe Jochen Hippler, *Strategische Grundprobleme externer politischer und militärischer Intervention. Unter besonderer Berücksichtigung der Krisensituationen des Nahen und Mittleren Ostens*, Duisburg: Institut für Entwicklung und Frieden, 2011 (INEF-Report Nr. 103/2011), S. 51.

63 Für eine detaillierte Übersicht siehe Gesellschaft für Evaluation e.V., Arbeitskreis »Evaluation von Entwicklungspolitik« in der DeGEval (Hg.), *Verfahren der Wirkungsanalyse. Eine Landkarte für die entwicklungspolitische Praxis*, Juli 2009, S. 15–17. Im Gegensatz zum randomisierten Kontrollgruppenvergleich wird bei einem quasi-experimentellen Gruppenvergleich keine zufällige (randomisierte) Auswahl der Teilnehmenden vorgenommen.

64 Siehe David A. Baldwin, »The Sanctions Debate and the

Ex-post-Evaluationen können auch in Form von Meta-Evaluationen stattfinden. Hier geht es darum, in einer Art kumulativer Gesamtschau die Ergebnisse einzelner Evaluationen zu analysieren und deren Qualität anhand erkenntnisrelevanter Kriterien zu bewerten.⁶⁵

Theoretische und methodologische Probleme

Evaluation ist durchaus so etwas wie die empirische Überprüfung von Theorien, sofern dabei Wirkungsannahmen untersucht werden, die den Interventionen zugrunde liegen.⁶⁶ Evaluation bestimmter Interventionen ist zwar keine Forschung, die über den Einzelfall hinausgehende Aussagen anstrebt. Aber methodisch sauber angelegt, sollte sie auch die wissenschaftliche Forschung befruchten. Umgekehrt sollte eine Programmtheorie möglichst dem Stand der wissenschaftlichen Diskussion entsprechen.⁶⁷ An der Schnittstelle zwischen Wissenschaft, Politik und Gesellschaft verortet, kann eine nach wissenschaftlichen Standards vorgenommene Evaluation sowohl wissenschaftlicher Grundlagenforschung als auch angewandter, in erster Linie problembasierter und lösungsorientierter (Auftrags-)Forschung dienen.⁶⁸

Wissenschaft, definiert als die Praxis geteilter Konventionen einer Expertengemeinschaft, bezeichnet keineswegs ein fest umrissenes Repertoire wissenschaftsphilosophischer Grundannahmen, Methoden und Instrumente zur Gewinnung von Wissen, wie es in der Diskussion oftmals suggeriert wird. Vielmehr gibt es verschiedene methodologische Zugänge zur

Logic of Choice«, in: *International Security*, 24 (1999/2000) 3, S. 80–107 (82–86).

65 Siehe Hellmut Wollmann, *Evaluation and Evaluation Research*, Madrid: Universidad Autónoma de Madrid, 2005 (Working Paper Nr. 54), S. 4; Thomas Widmer, *Meta-Evaluation. Kriterien zur Bewertung von Evaluationen*, Bern/Stuttgart/Wien 1996, S. 3f.

66 Siehe Ray Pawson, »Nothing as Practical as a Good Theory«, in: *Evaluation*, 9 (2003) 4, S. 471–490.

67 Siehe dazu Stewart I. Donaldson/Mark W. Lipsey, *Roles for Theory in Contemporary Evaluation Practice: Developing Practical Knowledge*, in: Ian F. Shaw/Jennifer C. Greene/Melvin M. Mark (Hg.), *Handbook of Evaluation. Policies, Programs and Practices*, Thousand Oaks u.a.: Sage Publications, 2006, S. 56–75.

68 Meist ist Letzteres der Fall, da Wissen bei der Evaluierung politischer Interventionen als Problemlösungsressource fungiert. Exemplarisch verdeutlicht dies ein Positionspapier der Bundesregierung: *Wissen für Entwicklung: Hochschulbildung und Wissenschaft in der deutschen Entwicklungspolitik*, Bonn: Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung, 2009 (BMZ Spezial Nr. 161).

Aneignung von Wissen, die auf unterschiedlichen philosophischen Festlegungen über die Beschaffenheit der Realität und die darüber zu erlangende Erkenntnis beruhen. All diese Zugänge können ihre eigene interne Validität besitzen und damit als Resultat wissenschaftliche Fakten (Evidenz) produzieren.⁶⁹ Die einschlägige Evaluationspraxis und die angewandten Bewertungsstandards entspringen dem jeweiligen wissenschaftsphilosophischen und erkenntnistheoretischen Verständnis. Zudem unterliegen sie dem Einfluss der jeweiligen Kultur und der Machtverhältnisse in Wissensvermittlung und Politikberatung.⁷⁰

Bei der Produktion von Evidenz, verstanden als Erkenntnis im weitesten Sinne, kann von einer *neopositivistischen*, einer *realistischen* und einer *konstruktivistischen* Methodologie gesprochen werden.⁷¹ Den entscheidenden Unterschied markiert dabei das Verhältnis von Realität und Beobachtern. Die *neopositivistische* Methodologie basiert auf einem dualistischen Verständnis, wonach Realität und Beobachter unabhängig voneinander sind.⁷² Evaluation erscheint

69 Für eine hilfreiche Typologie der verschiedenen methodologischen Zugänge siehe Patrick Thaddeus Jackson, *The Conduct of Inquiry in International Relations. Philosophy of Science and Its Implications for the Study of World Politics*, Abingdon/New York 2011.

70 Für einen Versuch, mehr Transparenz in der Politikberatung herzustellen, siehe Berlin-Brandenburgische Akademie der Wissenschaften, *Leitlinien Politikberatung*, Berlin 2008.

71 Diese Unterscheidung findet sich auch im für die Evaluation sozio-ökonomischer Aktivitäten zentralen Leitfaden der Europäischen Kommission *EVALSED: The Resource for the Evaluation of Socio-Economic Development* (online), S. 18f, <http://ec.europa.eu/regional_policy/sources/docgener/evaluation/guide/guide2012_evalsed.pdf>; siehe ferner Julian Hoerner/Paul Stephenson, »Theoretical Perspectives on Approaches to Policy Evaluation in the EU: The Case of Cohesion Policy«, in: *Public Administration*, 90 (2012) 3, S. 699–715 (702–706). Eine zweigeteilte Typologie, in der der realistische Ansatz nicht aufgeführt ist, findet sich bei Reina C. Neufeldt, »Frameworkers« and »Circlers« – Exploring Assumptions in Peace and Conflict Impact Assessment«, in: Beatrix Austin/Martina Fischer/Hans J. Gießmann (Hg.), *Advancing Conflict Transformation. The Berghof Handbook II*, Opladen/Farmington Hills: Barbara Budrich Publishers, 2011, S. 483–504.

72 Die neopositivistische Methodologie erlangte mit dem Behaviorismus nach dem Ende des Zweiten Weltkrieges eine nahezu unangefochtene Vorrangstellung in den Sozialwissenschaften, die bis in die heutige Zeit nachwirkt. Epistemologisch fußt die neopositivistische Methodologie auf dem Glauben an einen kumulativen Wissensprozess, bei dem sich eine Vielzahl von Einzeluntersuchungen zu einem sich stetig weiterentwickelnden Gesamtüberblick verdichtet. Siehe Donald T. Campbell/Julian C. Stanley, *Experimental and Quasi-Experimental Designs for Research*, Chicago: Rand McNally, 1963.

als neutraler Bewertungsvorgang außerhalb und unabhängig von persönlichen Präferenzen und gesellschaftlichen Prozessen, der Fakten als objektiv gegeben ansieht und danach Einschätzungen trifft. Kausalität wird in dem Sinn verstanden, dass eine ständige Korrelation der beobachteten Faktoren (Variablen) vorliegt. Der beobachtete Zusammenhang wird als Manifestierung eines allgemeingültigen Gesetzes (covering law) angesehen. Unabhängig vom jeweiligen Beobachter erscheinen die diesem Gesetz unterliegenden Fakten über verschiedene Kontexte hinaus generalisierbar. Die Identifikation einer eindeutigen Richtung in der Ursache-Wirkungs-Beziehung bleibt hierbei verborgen, doch wird versucht, Wirkungsketten zu identifizieren. Um Evidenz zu erzeugen, bedient sich diese Methodologie eines kontrafaktischen Kausalitätsbegriffs, der auf den Abgleich mit einer Kontroll- oder Vergleichsgruppe innerhalb eines stabil bleibenden Kontextes angewiesen ist.⁷³

Die *realistische* Methodologie dagegen lehnt das Konzept allgemeingültiger Ursache-Wirkungs-Zusammenhänge ab.⁷⁴ Ontologisch ist sie dem sogenannten wissenschaftlichen Realismus verhaftet, nach dem es eine materielle und soziale Realität gibt, die unabhängig von ihren Beobachtern und deren Interpretation existiert. Diese kann – und das ist die epistemologische Prämisse – zumindest annäherungsweise erkannt werden, auch wenn diese Erkenntnisse immer vor-

73 Der Begriff der Wirkung setzt bereits voraus, dass es eine Ursache gibt. Zu unterscheiden ist hierbei grundlegend zwischen den Definitionen Kausalität, Kausaleffekten, kausalen Erklärungen sowie der Erfassung von Kausalität. Siehe John Gerring, »Causal Mechanisms: Yes, but ...«, in: *Comparative Political Studies*, 43 (2010) 11, S. 1499–1626 (1519). Die bevorzugte Methode für den neopositivistischen Kausalitätsnachweis ist der quasi-experimentelle Kontrollgruppenvergleich. Siehe Howard White, *Theory-based Impact Evaluation: Principles and Practice*, Neu-Delhi: International Initiative for Impact Evaluation, Juni 2009 (Working Paper Nr. 3), S. 12.

74 Siehe Ray Pawson/Nick Tilley, *Realist Evaluation*, 2004; <www.communitymatters.com.au/RE_chapter.pdf>; Ray Pawson/Trisha Greenhalgh/Gill Harvey/Kieran Walshe, *Realist Synthesis: an Introduction*, Manchester: University of Manchester, Economic and Social Research Council (ESRC), Research Methods Programme (RMP), August 2004 (RMP Methods Paper Nr. 2/2004); Ray Pawson, *The Science of Evaluation. A Realist Manifesto*, Thousand Oaks u.a.: Sage Publications, 2013; Gill Westhorp (Hg.), *Realist Evaluation: An Overview*, Seminar Report, Wageningen: Centre for Development Innovation, Mai 2011; Nancy Cartwright/Jeremy Hardie, *Evidence-based Policy. A Practical Guide to Doing it Better*, New York: Oxford University Press, 2013, S. 14–88; siehe auch Ian Sanderson, »Evaluation, Policy Learning and Evidence-based Policy Making«, in: *Public Administration*, 80 (2002) 1, S. 1–22.

läufig und der Falsifizierung ausgesetzt sind. Kausalität wird aus Sicht des wissenschaftlichen Realismus in einem generativen Sinne verstanden: Es wird nach den nicht direkt beobachtbaren, aber als »real« verstandenen kausalen Mechanismen gefragt, die unter spezifischen Voraussetzungen bestimmte Wirkungen hervorbringen. Interventionen, die etwas verändern wollen, müssen als Theorien betrachtet werden. Sie sind Eingriffe in komplexe Systeme, die daraufhin analysiert werden, welche Mechanismen in bestimmten Kontexten bestimmte Ergebnismuster hervorrufen (context-mechanism-outcome pattern configuration). Aus Sicht der realistischen Evaluation lautet die grundlegende Frage: Was wirkt wie für wen in welchen Umständen und unter welchen Bedingungen?

Die konstruktivistische Methodologie folgt einem monistischen Verständnis, bei dem die Beobachter selbst Teil der von ihnen beobachteten Realität sind, so dass verschiedene Ansichten über die Realität, abhängig von den jeweils Beobachtenden, existieren können.⁷⁵ Da nach dieser Sicht Wahrnehmung – einschließlich der als rein mentales Ordnungskonstrukt begriffenen Kausalität – immer subjektiv ist, kann Evidenz nicht mehr als ein von vielen Menschen geteilter, gut begründeter »Glaubenssatz über die Welt« sein.⁷⁶ Evidenz erscheint als intersubjektiv konstruiertes Produkt, das an einen Verständnis-kontext gebunden bleibt. Zwar kann sie in Daten ausgedrückt werden, doch Aussagekraft hat sie nur im jeweiligen narrativen Verständnis- und Werte-kontext.⁷⁷ Dieser markiert damit auch jene Grenzen,

in denen sich Debatten bewegen (können). Die genaue Bedeutung von Evidenz wird kontinuierlich politisch und gesellschaftlich verhandelt und kann nur innerhalb der jeweiligen politisch sinnstiftenden Erzählung begriffen werden, in der sie verwendet wird.⁷⁸

Diese gewiss sehr groben Hinweise zeigen, dass unterschiedliche Wissenschaftsverständnisse sich auch in der Evaluationsforschung und -praxis widerspiegeln. Vorherrschend ist jedoch nach wie vor eine neopositivistische Methodologie, ausgerichtet auf die Analyse von Ursache-Wirkungs-Zusammenhängen. Sind die Kausalzusammenhänge *einfach*, ist die Wirkung einer Intervention ein abschätzbares Produkt eines vorhersehbaren Zusammenspiels weniger stabiler Komponenten wie etwa bei der Ausführung von Kochrezepten. Sind die Kausalzusammenhänge *kompliziert*, erscheint die Interventionswirkung als Zusammenspiel mehrerer stabiler Komponenten, das zwar schwer vorherzusehen, aber grundsätzlich berechenbar ist, zum Beispiel im Falle von Raketenstarts. Bei *komplexen* Kausalzusammenhängen ist die Wirkung einer Intervention durch ein unbekanntes Zusammenspiel einer Vielzahl miteinander interagierender Komponenten gekennzeichnet und daher weder vorher abzusehen noch während einer Intervention zu steuern.⁷⁹

Kann neopositivistische Evaluation bei einfachen Kausalzusammenhängen noch Evidenz – im Sinne der Wirkung von Interventionen – hervorbringen, ist dies bei komplizierten Formen schon schwieriger. Die Vorstellung eines linearen Zusammenhangs zwischen einer Intervention und ihrer Wirkung wird bereits hier mit dem Problem konfrontiert, die jeweiligen Effekte einer Intervention von denjenigen zu isolieren, die durch gleichzeitig stattfindendes Handeln anderer Akteure beeinflusst werden.⁸⁰ Schließlich entfalten sich Wirkungen auf allen Ebenen sozialer Aggregation, also auf der Mikro-, Meso- und Makroebene. Dabei können Effekte auf der Mikro- und der Mesoebene

⁷⁵ Siehe Thomas A. Schwandt, »Evaluation as Practical Hermeneutics«, in: *Evaluation*, 3 (1997) 1, S. 69–83; Thomas A. Schwandt, »Toward a Practical Theory of Evidence for Evaluation«, in: Stewart I. Donaldson/Christina A. Christie/Melvin M. Mark (Hg.), *What Counts as Credible Evidence in Applied Research and Evaluation Practice?*, Thousand Oaks u.a.: Sage Publications, 2009, S. 197–212 (199–203). Zur konstruktivistischen Methodologie in der Evaluationsforschung siehe auch Egon G. Guba/Yvonna S. Lincoln, *Fourth Generation Evaluation*, Newbury Park 1989. Zur konstruktivistischen Methodologie in der Politikfeldanalyse allgemein siehe Robert Hoppe, »Policy Analysis, Science and Politics: from ›Speaking Truth to Power‹ to ›Making Sense Together‹«, in: *Science and Public Policy*, 26 (1999) 3, S. 201–210 (205–208).

⁷⁶ Ernest R. House, »Unfinished Business: Causes and Values«, in: *American Journal of Evaluation*, 22 (2001) 3, S. 309–315 (313).

⁷⁷ Geld ist hier das paradigmatische Beispiel. Siehe Mary Poovey, *A History of the Modern Fact. Problems of Knowledge in the Sciences of Wealth and Society*, Chicago 1998. Zu dieser Unterscheidung und ihrer Relevanz für die politische Praxis siehe Sibylle Anderl, »Gute Daten, schlechte Daten«, in: *Frankfurter Allgemeine Zeitung*, 17.4.2013.

⁷⁸ Siehe Dvora Yanow, *How Does a Policy Mean? Interpreting Policy and Organizational Actions*, Washington, D.C., 1997.

⁷⁹ Diese Unterscheidung geht zurück auf Sholom Glouberman/Brenda Zimmerman, *Complicated and Complex Systems: What Would Successful Reform of Medicare Look Like?*, Ottawa: Commission on the Future of Health Care in Canada, 2002, S. 1–3.

⁸⁰ Siehe Maarten Hajer, »Policy without Polity? Policy Analysis and the Institutional Void«, in: *Policy Science*, 36 (2003) 2, S. 175–195 (186).

feststellbar sein, ohne dass diese auf die für die Zielformulierung relevante Makroebene ausstrahlen.⁸¹

Anders als bei einfachen oder komplizierten Formen lassen sich komplexe Kausalzusammenhänge nicht durch analytische Techniken wie Abstraktion oder Reduktion in Komponenten zerlegen. Vielmehr sind die Intervention und ihre Wirkung als ein Ganzes zu untersuchen. Da dieses mehr als die Summe seiner Teile ist, entfielen bei einer Zerlegung dessen spezifische Eigenschaften: das plötzliche und unvorhersehbare Auftreten unbekannter Faktoren (Emergenz), interdependente Dynamiken zwischen Akteuren und Strukturen sowie die Nicht-Linearität zwischen Faktoren und ihren Konsequenzen.⁸² Eine systematische Auseinandersetzung mit komplexen Kausalzusammenhängen findet bis auf wenige Ausnahmen bislang ausschließlich in der die Einzeldisziplinen übergreifenden Komplexitätstheorie statt.⁸³ Diese lehnt die Vorstellung ab, Ursache-Wirkungs-Zusammenhänge seien linear, proportional und voraussagbar, so dass sie auch auf die Zukunft übertragen werden können. Nicht Regelmäßigkeiten, sondern ständige Veränderung, Unvorhersehbarkeit und damit Unvorhersagbarkeit sowie die Gleichzeitigkeit von Nicht-Linearität und Linearität sind demnach die dominierenden (Un-)Ordnungsprinzipien.⁸⁴

81 So haben in Afghanistan der Aufbau ziviler Infrastruktur wie Schulen, Kliniken, Straßen oder der Energieversorgung sowie die Ausbildung von Polizeikräften dazu beigetragen, dass die staatlichen Leistungen für die Bevölkerung sich verbesserten. Doch diese Erfolge auf der Mikro- und Mesoebene bleiben auf der Makroebene ohne die erhofften Folgen für eine Stabilisierung des Landes. Siehe International Crisis Group, *Aid and Conflict in Afghanistan*, Kabul/Brüssel, 4.8.2011 (Asia Report Nr. 210). Für eine Analyse möglicher Erfolgsbedingungen von Interventionen mit dem Ziel der umfassenden Aufstandsbekämpfung siehe Matt Waldman/Thomas Ruttig, *Peace Offerings. Theories of Conflict Resolution and Their Applicability to Afghanistan*, Kabul: Afghanistan Analysts Network (AAN), 2011 (AAN Discussion Paper Nr. 1/2011).

82 Siehe Glenda H. Eoyang/Thomas H. Berkas, »Evaluating Performance in a Complex, Adaptive System (CAS)«, in: Michael R. Lissack/Hugh P. Gunz, *Managing Complexity in Organizations*, Westport 1999, S. 313–335.

83 Für eine knappe, praxisnahe Diskussion der einschlägigen Konzepte siehe OECD Global Science Forum, *Report on Applications of Complexity Science for Public Policy. New Tools for Finding Unanticipated Consequences and Unrealized Opportunities*, September 2009, sowie World Economic Forum, Global Agenda Council on Complex Systems, *Perspectives on a Hyperconnected World. Insights from the Science of Complexity*, Januar 2013.

84 Siehe Ben Ramalingam/Harry Jones/Toussaint Reba/John Young, *Exploring the Science of Complexity. Ideas and Implications for Development and Humanitarian Efforts*, London: Overseas

Development Institute, 2008 (Working Paper Nr. 285), S. 8–58.
85 Siehe Donald A. Schön, *The Reflective Practitioner*, New York: Norton, 1983, S. 42f.
86 Siehe Seva Gunitsky »Complexity and Theories of Change in International Politics«, in: *International Theory*, 5 (2013) 1, S. 35–63 (36).
87 Für eine Anwendung auf Evaluation siehe Patricia J. Rogers, »Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions«, in: *Evaluation*, 14 (2008) 1, S. 29–48.
88 Siehe Bärbel-Maria Kurth/Thomas Ziese, »Vorwort«, in: Robert Koch-Institut/Bayerisches Landesamt für Gesundheit und Lebensmittelsicherheit (Hg.), *Evaluation komplexer Interventionsprogramme in der Prävention: Lernende Systeme, lehrreiche Systeme?*, Berlin 2012 (Beiträge zur Gesundheitsberichterstattung des Bundes), S. 7.
89 Siehe Alvin M. Weinberg, »Science and Trans-Science«, in: *Minerva*, 10 (1972) 2, S. 209–222.

Während sich die Sozialwissenschaften immer stärker an Modellen der Naturwissenschaften orientierten, um wie diese allgemeingültige Gesetze und Wirkungsmechanismen zu identifizieren und Vorhersagen treffen zu können, verabschiedeten sich die Naturwissenschaften von ihrem reduktionistischen Weltzugang. Diese Hinwendung der Sozialwissenschaften zu einem reduktionistischen Ansatz ist umso erstaunlicher, als in politischen Interventionen gerade das Gegenteil jener Laborkonditionen herrscht, die für die neopositivistische Methode unerlässlich sind. Davon abgesehen sind die damit verbundenen quasi-experimentellen Methoden allenfalls dazu geeignet, relativ simple Sachverhalte zu analysieren.⁸⁵ Weil die neopositivistisch geprägte Sozialwissenschaft auf einfachen Ursache-Wirkungs-Zusammenhängen beharrt, verwundert es kaum, dass sie für die Evaluation komplexer Kausalmodelle wenig anzubieten hat.⁸⁶ Erst langsam rezipieren die sozialwissenschaftlichen Disziplinen vorhandene Alternativen.⁸⁷

Festzuhalten ist: Je einfacher die Kausalbeziehung und je niedriger das Niveau sozialer Aggregation, desto besser lässt sich eine Intervention mit Hilfe neopositivistischer Methoden evaluieren. Bei komplexen Kausalzusammenhängen und komplexen Instrumenten dagegen ist es bereits auf individueller Ebene äußerst aufwendig, Evidenz zu erzeugen.⁸⁸ Steigt die Komplexität sozialer Aggregation, wachsen die Schwierigkeiten für konventionelle Evaluation enorm. Angesichts dessen können Wirkungen von Interventionen vorher kaum abgesehen, während der Umsetzung nicht kontrolliert und im Nachhinein nur schwer nachgewiesen werden. Deshalb sind sie einer neopositivistischen Evaluation im strengen Sinne nicht mehr zugänglich.⁸⁹

Evaluation im Bereich Gewaltprävention und Friedensaufbau: Eine schwierige Herausforderung

Es sollte deutlich geworden sein, dass die Evaluation politischer Interventionen höchst anspruchsvoll ist, da sie von methodologischen Grundannahmen und deren methodischen Konsequenzen abhängt. Besonders heikel ist ihre Anwendung bei außenpolitischen Interventionen.⁹⁰ Mit diesen sollen komplexe soziale Prozesse in einem Kontext beeinflusst werden, über den nur begrenztes Wissen zur Verfügung steht und in dem Entscheidungen, anders als im eigenen politischen Gemeinwesen, nicht autoritativ durchgesetzt werden können. Zu bedenken ist auch, dass die eigene Außenpolitik mit derjenigen anderer Staaten, aber auch mit der Politik nichtstaatlicher Akteure interagiert. Aller Erfahrung nach kann Außenpolitik mit ihren sämtlichen Instrumenten gewünschte Entwicklungen oftmals bloß flankieren, denn deren (Erfolgs-)Bedingungen liegen wegen des komplexen Zusammenspiels zahlreicher Faktoren zumeist nicht im direkten Einflussbereich eines Landes oder Bündnisses. Nicht zuletzt ist der Zeitraum zu berücksichtigen, in dem Wirkungen zu erwarten und festzustellen sind. Gerade friedensaufbauende Maßnahmen brauchen einen langen Atem. Es ist schon eine große methodische Herausforderung, die unmittelbare Wirkung einzelner entwicklungspolitischer Projekte einzuschätzen. Weit schwieriger wird es, wenn die Wirkung von Interventionen bewertet werden soll, die zum Ziel haben, gewaltsame Konflikte zu verhindern und stabile Friedensstrukturen aufzubauen.⁹¹ Bei allen drei genannten Formen der Evaluation stellen

sich daher enorme grundlegende Probleme, unabhängig von der gewählten Methodologie.

Spezifische Probleme

Hinzu kommen spezifische Schwierigkeiten der einzelnen Evaluationstypen. Eine Ex-ante-Evaluation kann bereits in der Planungsphase von Interventionen stattfinden.⁹² Grundvoraussetzung ist die Analyse des Kontextes sowie des Konflikts, in den eingegriffen werden soll. Konflikte sind insofern kontextgebunden, als sie sich innerhalb kultureller Sinnzusammenhänge abspielen, die historisch gewachsen sind und in ihrer Komplexität mit einer standardisierten Analyse nur äußerst schwer zu fassen sind.⁹³ Die sozialen Beziehungen und Einstellungen der relevanten Akteure sind oft kaum zu identifizieren. Zudem erschwert es die Verflechtung des Konflikts mit dem sozialen, politischen und wirtschaftlichen Kontext, einzelne konflikttreibende Faktoren zu isolieren und zu bewerten.⁹⁴ Die für Ex-ante-Evaluationen unabdingbare Konfliktanalyse muss daher notwendig auf einer Untersuchung der Gesellschaft sowie der Machtstrukturen im Zielland aufgebaut sein, um die lokalen und regionalen Konfliktstrukturen hinreichend zu verstehen.⁹⁵ Doch vermehrtes Wissen über Konfliktquellen macht Interventionen nicht automatisch wirkungsvoller. Zu beachten ist der Kontext, der sich durch gewaltsame Konfliktaustragung verändert.⁹⁶

90 Das gilt bereits für innenpolitische Programme, bei denen mit großem personellem und finanziellem Aufwand Evaluation betrieben wird. Siehe die Teilevaluation des Kindergeldes sowie der Kinderfreibeträge, vorgenommen vom ifo Institut – Leibniz-Institut für Wirtschaftsforschung an der Universität München e.V. im Rahmen einer vierjährigen Gesamtevaluation der ehe- und familienbezogenen Leistungen: Helmut Rainer/Stefan Bauernschuster/Natalia Danzer/Timo Hener/Christian Holzner/Janina Reinkowski, *Kindergeld*, München 2013, <www.cesifo-group.de/DocDL/ifo_Forschungsbericht_60.pdf>.

91 Zu diesen Problemen siehe Larissa A. Fast/Reina C. Neufeldt, »Envisioning Success: Building Blocks for Strategic and Comprehensive Peacebuilding Impact Evaluation«, in: *Journal of Peacebuilding and Development*, 2 (2005) 2, S. 24–41.

92 Siehe Reimann/Chigas/Woodrow, *An Alternative to Formal Evaluation of Peacebuilding* [wie Fn. 55], S. 6–10.

93 Siehe Roger Mac Ginty, »Introduction: The Transcripts of Peace: Public, Hidden or Non-Obvious?«, in: *Journal of Intervention and Statebuilding*, (2013), First View (online), <www.tandfonline.com/doi/abs/10.1080/17502977.2012.727535#Uih9HNnFLWV>.

94 Siehe Fast/Neufeldt, »Envisioning Success« [wie Fn. 91], S. 25; Debra J. Rog, »When Background Becomes Foreground: Toward Context-Sensitive Evaluation Practice«, in: *New Directions for Evaluation*, 135 (2012), S. 25–40.

95 Siehe Hippler, *Strategische Grundprobleme externer politischer und militärischer Intervention* [wie Fn. 62] S. 35–41.

96 Siehe Susan L. Woodward, »Do the Root Causes of Civil War Matter? On Using Knowledge to Improve Peacebuilding Interventions«, in: *Journal of Intervention and Statebuilding*,

In eine Ex-ante-Evaluation einfließen müsste auch die Interaktion zwischen den intervenierenden Akteuren und der lokalen Gesellschaft, in die diese eingreifen.⁹⁷ Zwar identifizieren die meisten Konfliktanalysen noch die potentiellen Konfliktquellen, erfassen jedoch nicht das dynamische Zusammenspiel ihrer treibenden Faktoren,⁹⁸ das systematisch auch kaum exakt nachzuvollziehen ist.⁹⁹

Die Operationalisierung interventionsleitender Konzepte – Frieden, Konflikt, Krise – wird durch ihre Unschärfe beeinträchtigt.¹⁰⁰ Bereits die analytische Unterscheidung zwischen Konflikt und Krise ist nicht eindeutig spezifiziert.¹⁰¹ Außerdem können Konflikte sich wandeln.¹⁰² Auch wenn gewaltsam ausgetragene

Konflikte beigelegt sind,¹⁰³ besteht die Gefahr, dass sie wieder aufbrechen. Woran dies im jeweiligen Fall liegt, ist höchst umstritten, da unterschiedliche Daten voneinander abweichende Resultate in der Konfliktanalyse befördern. Aus diesem Grund scheint der Versuch zum Scheitern verurteilt, eine für die politische Praxis verlässliche Generalisierung der treibenden Faktoren zu erreichen.¹⁰⁴

Die sozio-psychologischen Konfliktquellen in komplexen Gewaltsituationen bleiben bei der Ex-ante-Evaluation meist unberücksichtigt oder unterbelichtet.¹⁰⁵ Kontext- und Konfliktanalysen befassen sich kaum mit Einwirkungen auf und langfristigen Folgen für die menschliche Psyche nach schweren Gewalterlebnissen und den daraus entstehenden, dem Friedensaufbau abträglichen Gewaltkulturen.¹⁰⁶ Es ist eine analytische Schwäche des liberalen Friedensparadigmas,¹⁰⁷ sich auf Indikatoren zu fixieren und lokales Wissen tendenziell zu vernachlässigen.¹⁰⁸ Dieses Paradigma gründet nämlich auf einem Stabilitätsverständnis, das nachhaltige soziale Dynamiken weitgehend ausblendet und eher Stillstand zum Ziel hat. Wenn dann von Fragilität die Rede ist, wird diese meist am westlichen Staatsmodell gemessen.¹⁰⁹

1 (2007) 2, S. 143–170.

97 Für einen ersten Ansatz siehe Thorsten Bonacker/Michael Daxner/Jan H. Free/Christoph Zürcher (Hg.), *Interventionskultur. Zur Soziologie von Interventionengesellschaften*, Wiesbaden: VS Verlag für Sozialwissenschaften, 2010.

98 Siehe Peter Woodrow/Diana Chigas, *Connecting the Dots: Evaluating Whether and How Programmes Address Conflict Systems*, in: Daniela Körppen/Norbert Ropers/Hans J. Gießmann (Hg.), *The Non-Linearity of Peace Processes. Theory and Practice of Systemic Conflict Transformation*, Opladen/Farmington Hills: Barbara Budrich Publishers, 2011, S. 205–228 (208).

99 »[...] programs sit in ever changing open systems, some of the mechanisms of which we can describe in models [...] but not in ways that we can manipulate or predict.« Pawson, »Nothing as Practical as a Good Theory« [wie Fn. 66], S. 486.

100 Siehe A. Betts Fetherston, »Peacekeeping, Conflict Resolution and Peacebuilding: A Reconsideration of Theoretical Frameworks«, in: *International Peacekeeping*, 7 (2000) 1, S. 190–218 (192).

101 Konflikte und Krisen sind zuerst einmal allgegenwärtige Begleiter menschlichen Zusammenlebens. Konflikte entstehen aus gegensätzlichen normativen Auffassungen und Interessen und können durch politisches oder militärisches Handeln aufgelöst werden. Die Konfliktlösungsmechanismen einer Gesellschaft, institutionalisiert in Rechtsordnungen oder dem Gewohnheitsrecht, beeinflussen jeweils den Grad der Gewaltanwendung bei der Austragung und Beilegung von Konflikten. Siehe Jeffrey A. Sluka, *The Anthropology of Conflict*, in: Carolyn Nordstrom/JoAnn Martin (Hg.), *The Paths of Domination, Resistance and Terror*, Berkeley/Los Angeles/Oxford: University of California Press, 1999, S. 18–36 (31). Krisen hingegen sind intersubjektiv geteilte Wahrnehmungen eines sich zuspitzenden Problems, dessen Bewältigung ungewiss erscheint und das eine Entscheidung mit weitreichenden Konsequenzen für den späteren Verlauf erfordert.

102 Siehe Guilain Denoëux/Lynn Carter, *Guide to the Drivers of Violent Extremism*, Washington, D.C.: United States Agency for International Development, 2009. Auf die Nicht-Linearität von Konflikten wird ausdrücklich Bezug genommen im *Conflict Assessment Framework*, Washington, D.C.: United States Agency for International Development, Juni 2012, S. 16, sowie im dazugehörigen *Conflict Assessment Framework. Application*

Guide, Washington, D.C.: United States Agency for International Development, Juni 2012.

103 Zur Beilegung und Beendigung von Konflikten existieren mehrere Datenbanken, so die Peace Agreement Digital Collection des United States Institute of Peace, <www.usip.org/publications/peace-agreements-digital-collection>, die Peace Accord Matrix der Universität Notre Dame, <<https://peaceaccords.nd.edu/>>, sowie im Conflict Data Project an der Universität Uppsala, <www.ucdp.uu.se/gpdatabase/search.php>.

104 Siehe Nicholas Sambanis, »What is Civil War? Conceptual and Empirical Complexities of an Operational Definition«, in: *Journal of Conflict Resolution*, 48 (2004) 6, S. 814–858.

105 Siehe Fetherston, »Peacekeeping, Conflict Resolution and Peacebuilding« [wie Fn. 100], S. 194f.

106 Siehe Vivienne Jabri, *Discourses on Violence: Conflict Analysis Reconsidered*, Manchester 1996. Doch gerade eine Veränderung in der sozialen Praxis ist entscheidend für die Überwindung von Konflikten. Siehe auch A. Betts Fetherston/Carolyn Nordstrom, »Overcoming Habitus in Conflict Management: UN Peacekeeping and War Zone Ethnography«, in: *Peace & Change*, 20 (1995) 1, S. 94–119 (103–113).

107 Für eine einführende Diskussion in den Begriff »peacebuilding«, der weit über »peacemaking« und »peacekeeping« hinausgeht, jedoch von »statebuilding« abzugrenzen ist, siehe Ulrich Schneckener, »Frieden Machen: Peacebuilding und peacebuilder«, in: *Die Friedens-Warte*, 80 (2005) 1–2, S. 17–39.

108 Siehe Michael Pugh, »Local Agency and Political Economics of Peacebuilding«, in: *Studies in Ethnicity and Nationalism*, 11 (2011) 2, S. 308–320 (310).

109 Siehe Stephen Baranyi/Marie-Eve Desrosiers, »Development Cooperation in Fragile States: Filling or Perpetuating

Bei der Ex-ante-Evaluation muss schließlich überprüft werden, welche Kausalzusammenhänge als Grundlage der Programmtheorie dienen, die der Intervention zugrunde liegt.¹¹⁰ Bislang wurden komplizierte Kausalzusammenhänge in den weitaus meisten Programmtheorien kaum beachtet. Stattdessen basieren Wirkungsannahmen überwiegend auf einfachen Kausalzusammenhängen. Diese werden linear in die Zukunft projiziert und liefern vermeintlich vorhersehbare Ergebnisse.¹¹¹

Ob und inwiefern sich gewünschte Entwicklungen mit einer Intervention befördern lassen, hängt von der tatsächlichen Wirkung der eingesetzten Instrumente ab. Bei der *formativen Evaluation* steht demgemäß die Wirkungsanalyse im Mittelpunkt. Hier werden die Wirkungen der eingesetzten Instrumente schon im Verlauf einer Intervention untersucht. Aufgrund ihres kontrafaktischen Kausalitätsbegriffs sind die weithin als Standard angesehenen neopositivistischen Wirkungsanalysen auf zuverlässige und gültige Basisdaten angewiesen. Doch Daten über potentielle und tatsächliche Konfliktregionen werden weder systematisch noch kontinuierlich über einen längeren Zeitraum hinweg erhoben.¹¹²

Auf der Mikroebene sind die Wirkungen einzelner, lokal begrenzter Maßnahmen noch identifizierbar. So lässt sich beispielsweise der *output* präzise bestimmen, etwa wie viele Schulen und Brunnen errichtet oder Journalisten ausgebildet wurden. Auch die *outcomes*, also der Unterricht von Schulklassen, die Versorgung

mit sauberem Trinkwasser oder die verbesserte Qualität der Medienberichterstattung, können noch recht genau festgestellt werden. Die Wirkung (*impact*) auf der Meso- und Makroebene, wo die Oberziele einer Intervention wie Frieden oder Stabilität (*peace writ large*) gemeinhin angesiedelt sind, ist jedoch so diffus, dass die Entwicklungen auf dieser Ebene nicht eindeutig auf Wirkungen einer Maßnahme zurückgeführt werden können. Ist der Nachweis von Kausalzusammenhängen bereits auf der Mikroebene mühsam, wenn er über relativ einfach messbare Aktivitäten hinausgehen soll, so scheint er auf der Makroebene unmöglich zu sein.¹¹³

Bei der *Ex-post-Evaluation* wird festgelegt, woran Gelingen oder Misslingen einer Intervention gemessen werden sollen. Dabei muss zunächst geklärt werden, auf welcher Ebene Wandel stattzufinden hat und welche Akteure zu beeinflussen sind, um gewünschte Entwicklungen voranzubringen. Dies können einzelne Individuen, Gruppen, Familien, ethnische Clans, religiöse Netzwerke oder die Gesellschaft in ihrer Gesamtheit sein. Ist Wandel auf der individuellen Ebene angestrebt,¹¹⁴ müssen Einstellungs- und Verhaltensänderungen beobachtet und nachgewiesen werden.¹¹⁵ Gleiches gilt für die Gruppenebene sowie für die Ebene des gesamten Staates. Veränderte Einstellungen sind jedoch in einer Konfliktregion schwer zu messen; es kann zudem Jahre dauern, bis eine friedensfördernde Veränderung registriert werden kann.¹¹⁶

Es besteht ein Spannungsverhältnis zwischen einer klar operationalisierbaren Zieldefinition einerseits und der Wandelbarkeit und Kontingenz von Zielen im Interventionszeitraum andererseits. Bleibt das Ziel vage, wie etwa »Reintegration von Kämpfern«,¹¹⁷ kann mit quantitativen Methoden kaum aussagekräftig

Gaps?«, in: *Conflict, Security & Development*, 12 (2012) 5, S. 443–459 (451). In der einschlägigen Literatur wird der Ruf nach einer post-liberalen Peacebuilding-Diskussion lauter. Siehe Kristoffer Lidén/Roger Mac Ginty/Oliver P. Richmond, »Introduction: Beyond Northern Epistemologies of Peace: Peacebuilding Reconstructed?«, in: *International Peacekeeping*, 16 (2009) 5, S. 587–598 (594).

110 Auf die einschlägige Literatur zu den entsprechenden Versuchen wird verwiesen in Funnell/Rogers, *Purposeful Program Theory* [wie Fn. 53], S. 70.

111 Siehe Michael Quinn Patton, *Developmental Evaluation. Applying Complexity Concepts to Enhance Innovation and Use*, New York 2011, S. 90–97.

112 Zur Problematik siehe Michael Bamberger, »The Evaluation of International Development Programs: a View from the Front«, in: *American Journal of Evaluation*, 21 (2000) 1, S. 95–102. Das beim German Institute of Global and Area Studies (GIGA) angesiedelte Projekt »Krisenfrühwarnung für das BMZ« nimmt eine »jährliche Statusanalyse von etwa 100 Ländern unter dem Gesichtspunkt der innerstaatlichen Gewaltneigung« vor. Siehe die Website des GIGA, <www.giga-hamburg.de/index.php?file=krisenindikatoren.html&folder=wissenstransfer>.

113 Siehe Christoph Spurk, »Forget Impact: Concentrate on Measuring Outcomes«, in: *New Routes*, 13 (2008) 3, S. 11–14.

114 Die Fähigkeit von Individuen, positive Veränderung herbeizuführen, wird auch mit dem Konzept des Sozialkapitals erfasst. Siehe Joseph Schumacher, »What to Measure in Peace Operations«, in: Ann Livingstone/Aleisha Arnusch (Hg.), *Measures of Effectiveness: Peace Operations and Beyond*, Clementsport: The Canadian Peacekeeping Press, 2007 (The Pearson Papers Bd. 10, Nr. 1), S. 45–59.

115 Siehe Jeffrey James/Efraim Gutkind, »Attitude Change Revisited: Cognitive Dissonance Theory and Development Policy«, in: *World Development*, 13 (1985) 10/11, S. 1139–1149.

116 Dies wird auch in den OECD-Standards konzidiert. Siehe OECD, *Evaluating Peacebuilding Activities* [wie Fn. 33], S. 68.

117 Siehe *Results: Good, but Too Local. Summary of the Cross-Section Evaluation Crisis Prevention and Peacebuilding*, Eschborn: Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ), Monitoring and Evaluation Unit, 2012, S. 3.

evaluiert werden, ob es erreicht wurde.¹¹⁸ Wandelt sich die Zieldefinition über die Zeit, können leicht punktuelle Verbesserungen oder weniger ambitionierte Indikatoren zur Erfolgsmessung herangezogen werden. Eine solche Entwicklung lässt sich am Beispiel der ISAF-Mission in Afghanistan beobachten, in deren Verlauf die Interventionsziele gesenkt wurden. Sollte zuerst noch demokratische Staatlichkeit etabliert werden, reichte es später aus, Terroranschläge zu verhindern.¹¹⁹ Multiple Zielkonflikte in und zwischen Regierungen, Nichtregierungsorganisationen und anderen Geldgebern beeinträchtigen kohärente und dauerhaft stabile Zielsetzungen. Weiterhin greift der angelegte Zeitrahmen für eine Bewertung oft zu kurz. Die Wirkungen treten, wenn sie überhaupt isoliert betrachtet werden können, meist lange nach einer Intervention zutage. Evaluationen erfassen stets nur einen Ausschnitt längerfristigen Wandels.¹²⁰ Aus Interventionen gewonnenes Wissen bezieht sich immer auf den Einzelfall und die einschlägigen Kontextbedingungen.

Begrenzte Erträge

In der Außenpolitik bewirken die zahlreichen zu berücksichtigenden Faktoren, ihr Zusammenspiel und die hohe soziale Aggregation, dass sich die Szenerie außerordentlich komplex darstellt. Die Quintessenz einer vom britischen Ministerium für internationale Entwicklung in Auftrag gegebenen Studie lautete daher, der Aufwand für den Nachweis von Wirkungen sei so groß, dass dieser auf wenige ausgewählte Interventionen begrenzt bleiben müsse.¹²¹ In der Tat lässt sich festhalten, dass der Aufwand von Evaluations-

118 Siehe Richard Bowd/Alpaslan Özerdem, »How to Assess Social Reintegration of Ex-Combatants«, in: *Journal of Intervention and Statebuilding*, (2013), S. 1–23.

119 Für ausführliche Berichte zur Fortentwicklung und Erreichung der Ziele siehe die vierteljährlichen detaillierten Berichte des Special Inspector General for Afghanistan Reconstruction (SIGAR), <www.sigar.mil/quarterlyreports/>. Für die Anpassungen der Ziele in den Mandaten des Deutschen Bundestages siehe dessen Website, <www.bundestag.de/bundestag/ausschuesse17/a12/auslandseinsaetze/auslandseinsaetze/isaf_bisherige_mandate.html>.

120 Siehe Eoyang/Berkas, »Evaluating Performance in a Complex Adaptive System (CAS)« [wie Fn. 82], S. 316.

121 Siehe Elliot Stern/Nicoletta Stame/John Mayne/Kim Forss/Rick Davies/Barbara Befani, *Broadening the Range of Designs and Methods for Impact Evaluations*, London: Department for International Development, April 2012 (Working Paper Nr. 38), S. 84f.

studien den Ertrag übersteigt, wenn sie denn überhaupt die Wirkung von Interventionen untersuchen und sich nicht auf Entscheidungsprozesse und Abläufe beschränken.¹²²

So lieferte eine Meta-Analyse von Evaluationsstudien zum Statebuilding ernüchternde Resultate:¹²³ Der »konventionelle Ansatz« in der Evaluationsforschung sei an seine Grenzen gestoßen. Die meisten dieser Studien seien methodisch fragwürdig und entsprächen nicht den Erfordernissen sozialwissenschaftlich reflektierter Forschung. Dies betreffe nicht nur die Validität der Daten, sondern grundlegende epistemologische und methodologische Fragen. Selten werde die Zuordnungslücke thematisiert, also der Kausalzusammenhang zwischen bestimmten Interventionen und beobachtbaren Entwicklungen und Effekten. Daher sei es kein Wunder, dass zwischen aufbereiteter Evidenz und den Folgerungen, die daraus gezogen werden, meist kein plausibler Zusammenhang bestehe und die Ergebnisse eher weitverbreitete Meinungen jener widerspiegeln, die die Politik betreiben, und jener, die sie bewerten. Ferner wurde kritisiert, dass oft eine überzeugende »theory of change« fehle. Zudem werde trotz hoher Kosten häufig nur wenig neues handlungsrelevantes Wissen erzeugt. Das gilt dieser Meta-Analyse zufolge auch für jene beiden Evaluationen, die noch am besten beurteilt werden: eine Studie zu Krisenprävention und Friedensaufbau im Südsudan und eine andere zu solchen Programmen in Sri Lanka. Beide Studien führen deutlich vor Augen, wie eine falsche Konfliktanalyse und unzutreffende beziehungsweise nicht klar formulierte oder fehlende Wirkungsmodelle eine Wirkungsevaluation erschweren. Sie zeigen auch, wie die potentielle Aussagekraft einer Evaluation durch eine mangelhafte Konzeption der Programme begrenzt wird.

Zehn Geberländer, darunter Deutschland, hatten die Evaluation zu ihren einschlägigen Projekten und Programmen im Südsudan in Auftrag gegeben. Der Abschlussbericht erschien im Dezember 2010. Die

122 So etwa die großangelegte Studie für die Europäische Kommission: Aide à la Décision Économique (ADE), *Thematic Evaluation of European Commission Support to Conflict Prevention and Peace-building: Final Report* (Vol. 1: Main Report, Vol. 2: Annexes), Brüssel: European Commission, September 2010.

123 In dieser Meta-Analyse wurden rund 100 Studien gesichtet und 40 näher ausgewertet. Jörn Grävingholt/Julia Leininger/Christian von Haldenwang, *Effective Statebuilding? A Review of Evaluations of International Statebuilding Support in Fragile Contexts*, Kopenhagen: Ministry of Foreign Affairs of Denmark, Juni 2012.

Evaluation lotete geradezu beispielhaft Möglichkeiten und Grenzen der Bewertung des Handelns im Aktionsfeld Krisenprävention und Friedensaufbau aus.¹²⁴ Untersucht wurde, wie die vielfältigen Aktivitäten die Dynamik des Konflikts beeinflusst haben, was Folgerungen für die weitere Ausgestaltung der Tätigkeit ermöglichen sollte. Beabsichtigt war also eine Wirkungsanalyse. Doch nicht jedes der mehr als 400 Projekte wurde unter die Lupe genommen. Vielmehr sollte die »kollektive Wirkung der internationalen Unterstützung« bewertet werden. Einige Ergebnisse sind von besonderem Interesse. So hieß es, die vorherrschende »theory of change« sei dem Konflikt nicht angemessen. Deren Prämisse lautete, der Mangel an Entwicklung im Südsudan heize den Konflikt an. Deswegen war die Arbeit überwiegend entwicklungsorientiert, was aber der Konfliktodynamik nicht entsprach, wie im Abschlussbericht beanstandet wurde. Daher waren die Hilfsprogramme, so wie sie angelegt waren, nicht geeignet, den Friedensaufbau zu befördern. Der Übergang zum Frieden wurde von den Geberländern offensichtlich als eher technischer Prozess gesehen. Eine sorgfältige Analyse der Machtstrukturen und konfliktantreibenden Faktoren lag den Aktivitäten nicht zugrunde.

Im Mittelpunkt der zweiten Studie, die in der oben aufgeführten Meta-Analyse positiv hervorgehoben wurde, standen die Strategien von zehn Geberländern und -organisationen in Sri Lanka, die dort Konfliktprävention und Friedensaufbau betreiben wollten.¹²⁵ Die Evaluation machte exemplarisch ein Defizit nicht nur dieser, sondern vieler Interventionen deutlich, die allgemein Konfliktvermeidung und Friedensaufbau zum Ziel haben: das Fehlen konkreter Wirkungsannahmen, die in eine Konfliktanalyse eingebunden sind. Unterbleibt diese, besteht die Tendenz, eine fehlende oder disparate Entwicklung – so in Sri Lanka – als Konfliktursache implizit vorauszusetzen.

In Wirkungsanalysen, die dem Stand der konventionellen Evaluationsforschung folgen, wird das einer

124 Jon Bennett/Sara Pantuliano/Wendy Fenton/Anthony Vaux/Chris Barnett/Emery Brusset, *Aiding the Peace. A Multi-donor Evaluation of Support to Conflict Prevention and Peacebuilding Activities in Southern Sudan 2005–2010*, Hove (UK): ITAD Ltd., Dezember 2010.

125 Die Studie sollte dreierlei liefern: eine Einschätzung der Strategien, eine Meta-Evaluation von Projektevaluationen und eine Bewertung der Koordinierungsaktivität. Siehe Nick Chapman/Debi Duncan/David Timberman/Kanaka Abeygunawardana, *Evaluation of Donor-supported Activities in Conflict-Sensitive Development and Conflict Prevention and Peacebuilding in Sri Lanka*, Main Evaluation Report, Final Report, Oktober 2009.

Intervention zugrundeliegende Wirkungsmodell untersucht und eine zu erwartende Wirkungskette identifiziert. Viele Evaluationen sehen sich dem Problem ausgesetzt, dass in Projekten und Programmen die Wirkungsvermutungen und Indikatoren fehlen, an denen sich die Wirkung bemessen ließe. Eine Evaluation der Tätigkeit des deutschen zivilen Friedensdienstes ergab, dass nur in einem Land (Kambodscha) ein Wirkungsmodell formuliert worden war.¹²⁶ Dass schon eine kritische Analyse beziehungsweise Rekonstruktion impliziter kausaler Wirkungsannahmen erkenntnisfördernd sein kann, zeigt die gerade erwähnte Evaluation. Sie offenbart, wie zweifelhaft die in der »conflict resolution school« der 1970er Jahre verwurzelten Annahmen sind und wie wenig die daran geäußerte Kritik reflektiert wurde. Die auf der Mikroebene individuellen Verhaltens angesiedelten Theorien postulieren, dass Friedenserziehung, Dialog zwischen Konfliktparteien und Versöhnung friedensschaffend wirken, doch es werden keine klaren Wirkungsmechanismen zwischen individuellem Einstellungswandel und Friedenskonsolidierung identifiziert.¹²⁷

Wie dieses Beispiel erkennen lässt, können Evaluationen dazu dienen, explizite oder implizite theoretische Annahmen zu überprüfen, die bestimmten Interventionen zugrunde liegen.¹²⁸ So lautet eine verbreitete Auffassung, Fortschritte in der wirtschaftlichen Entwicklung sowie auf dem Weg zu Demokratie und Rechtsstaatlichkeit trügen grundsätzlich zu Konfliktprävention und Friedensaufbau bei. Das kann, muss aber nicht sein. Interventionen in einzel-

126 Thania Paffenholz et al., *The German Civil Peace Service, Synthesis Report, Volume I: Main Report*, Bonn: Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung, April 2011, S. 7.

127 Ebd., S. 27ff.

128 Dazu auch Marie Gaarder/Jeanne Annan, *Impact Evaluation of Conflict Prevention and Peacebuilding Interventions*, The World Bank, Independent Evaluation Group, Public Sector Evaluation Department, Juni 2013 (Policy Research Working Paper Nr. 6496). Hinzuzufügen ist: wenn denn Evaluationen diese kritische Funktion überhaupt wahrnehmen. Eine Auswertung von zehn Evaluationsberichten, deren Gegenstand friedensaufbauende Projekte unterschiedlicher Organisationen in Afrika sind, präsentiert eine enttäuschende Bilanz in dieser Frage. Siehe *Linking Program Design and Evaluation in Peacebuilding: A Challenging Task. A Meta-Review of Peacebuilding Initiatives in Active Conflict Zones in Africa*. A Report from the Center for Peacebuilding and Development School of International Service, American University and Search for Common Ground, 15.1.2013.

nen dieser Bereiche können Konflikte unter Umständen sogar verschärfen.¹²⁹

Der Blick auf vorliegende Evaluationen zeigt zudem ein grundlegendes Problem: Oft haben die untersuchten Programme und Projekte nur einen sehr vagen Bezug zu den Großzielen wie Konfliktvermeidung und Friedensaufbau. In einer Studie zu Projekten im Kongo, die von einigen Geberländern und Organisationen in Auftrag gegeben wurde, lautete das kritische Resümee, nur wenige Maßnahmen könnten als direkter Beitrag zu Konfliktvermeidung und Friedensaufbau gewertet werden. Insgesamt handle es sich um eine Vielzahl eher unverbundener sektoraler Projekte.¹³⁰

129 Siehe OECD, »Encouraging Effective Evaluation of Conflict Prevention and Peacebuilding Activities. Towards DAC Guidance«, in: *OECD Journal on Development*, 8 (2008) 3, S. 7–64 (26f).

130 Amani Labda, *Peace Maybe. Joint Evaluation of Conflict Prevention and Peacebuilding in the Democratic Republic of Congo, Synthesis Report*, Ohain (Belgien): Channel Research, Juni 2011.

Grenzen des Möglichen: Was von Evaluation erwartet werden kann

Einigermaßen verlässliche Wirkungsnachweise in der Außenpolitik (aber nicht nur dort) sind sowohl finanziell und zeitlich aufwendig als auch methodisch außerordentlich schwierig.¹³¹ Wie lässt sich eine kausale Beziehung zwischen Intervention und Effekten nachweisen, also ob und inwieweit die Intervention zu den Wirkungen beigetragen hat?¹³² Zu prüfen sind konkurrierende Erklärungen für die beobachtete Entwicklung, und es muss darüber nachgedacht werden, was vielleicht ohne eine Intervention geschehen wäre. Die Wirkungsannahme, die der Intervention zugrunde liegt, muss überzeugend sein und möglichst auf einer breiteren Evidenzbasis beruhen.¹³³

Doch selbst wenn es eine gewisse Evidenzgrundlage gibt und kausale Zusammenhänge plausibel gemacht werden können,¹³⁴ heißt das nicht, dass Interventionen, die in einem bestimmten Kontext effektiv waren, es auch in einem anderen sein werden: Es kommt darauf an, ob auch in einem anderen, ähnlichen Umfeld jene oft übersehenen unterstützenden Faktoren existieren, deren es bedarf, damit eine Maßnahme die erwartete Wirkung entfalten kann.¹³⁵ Die Kontext-

gebundenheit von Wirkungen wird in der konventionellen, einer neopositivistischen Methodologie verhafteten Evaluationsforschung (deren Möglichkeiten und Grenzen in dieser Studie ausgelotet wurden) nicht angemessen reflektiert – im Unterschied zum »realistischen« Analyseansatz. Dieser wurde bislang in der Außenpolitikevaluation vernachlässigt, passt aber von seinem Grundverständnis her besser zu komplexen Interventionen in komplexen Systemen.¹³⁶ Danach hängen die Ergebnisse einer Intervention davon ab, wie diese in einem spezifischen Kontext Kausalmechanismen aktiviert.¹³⁷

Falsche Wirkungsannahmen können durch wissenschaftliche Evaluation vielleicht noch identifiziert, richtige jedoch kaum benannt werden.¹³⁸ Indem Evaluationen Wirkungsannahmen fragwürdig erscheinen lassen, können sie als Rückkoppelungsmechanismus zwischen angenommener und tatsächlicher Wirkung in gewissem Maße zu Lernerfahrungen führen.¹³⁹ Oft wird ja verlangt, Evaluationen sollten lernorientiert angelegt sein und zum Wissen darüber beitragen, was unter welchen Bedingungen in der Krisen- und Konfliktprävention wirksam sein kann, was eher nicht.¹⁴⁰

131 Zu den Problemen, verbunden mit dem Hinweis, die politischen Erwartungen an Wirkungsanalysen herunterzuschrauben, siehe auch Martin Quack, *Ziviler Friedensdienst. Exemplarische Wirkungsanalysen*, Wiesbaden: VS Verlag für Sozialwissenschaften, 2009.

132 Wenn, wie im Hinblick auf die realistische Methodologie anzunehmen ist, Wirkungen immer von einem Zusammenspiel unterschiedlicher kausaler Faktoren (»Kausalpaket«) hervorgebracht werden, dann ist »contributory cause« der angemessene Begriff. Die Intervention ist in diesem Sinne ein unzureichender, aber notwendiger Faktor in einem Gesamtpaket, das ausreicht, um die Wirkungen hervorzubringen. Siehe dazu Stern et al., *Broadening the Range of Designs and Methods for Impact Evaluations* [wie Fn. 121], S. 38–41.

133 Siehe Mark M. Rogers, *Evaluating Impact in Conflict Prevention and Peacebuilding Programs*, Cambridge, MA: CDA Collaborative Learning Projects, Reflecting on Peace Practice Program, 2012; White, *Theory-based Impact Evaluation: Principles and Practice* [wie Fn. 73].

134 Mehr wird auch bei einem anspruchsvollen Forschungsdesign nicht möglich sein. Zur Problematik der Wirkungsanalyse siehe auch Solveig Richter, *Zur Effektivität externer Demokratisierung: Die OSZE in Südosteuropa als Partner, Mahner, Besserwisser?*, Baden-Baden: Nomos, 2009, S. 47–81.

135 Zu dieser Frage siehe Nancy Cartwright, »Knowing What We Are Talking About: Why Evidence Doesn't Always Travel«,

in: *Evidence & Policy*, 9 (2013) 1, S. 97–112.

136 Siehe Lene Holm Pedersen/Olaf Rieper, »Is Realist Evaluation a Realistic Approach for Complex Reforms?«, in: *Evaluation*, 14 (2008) 3, S. 271–293.

137 Im Detail zu den mit der »Context-Mechanism-Outcome configuration« aufgeworfenen Fragen siehe Denise E. de Souza, »Elaborating the Context-Mechanism-Outcome configuration (CMOC) in Realist Evaluation: a Critical Realist Perspective«, in: *Evaluation*, 19 (2013) 2, S. 141–154.

138 Oder wie dies mit Blick auf die Leistung der Politikwissenschaft formuliert wurde: »Wenngleich Politikwissenschaft nicht in der Lage ist, »richtige« Wirkungsmodelle anzubieten, so kann sie doch »falsche« Wirkungsannahmen identifizieren.« Maull, »Wissenschaftliche Außenpolitik-Evaluation: Ein Oxymoron?« [wie Fn. 20], S. 115.

139 Dabei handelt sich um sogenanntes »double-loop learning« – im Unterschied zum »single-loop learning«, bei dem es unter anderem darum geht, ein Programm im Rahmen bestehender, nicht in Frage gestellter Annahmen zu verbessern. Zur Unterscheidung siehe Cheyanne Scharbatke-Church, *Evaluation Peacebuilding: Not Yet All It Could Be*, in: Austin/Fischer/Gießmann (Hg.), *Advancing Conflict Transformation* [wie Fn. 71], S. 459–482 (472).

140 Siehe Hal Culbertson, »The Evaluation of Peacebuilding Initiatives: Putting Learning into Practice«, in: Daniel Philpott/

Doch die Kontextgebundenheit von Interventionen verhindert eine verlässliche Generalisierung über die zu erwartenden Wirkungen.¹⁴¹ Projekt-, Programm- und Strategieevaluationen neopositivistischer Provenienz können wenig darüber sagen, ob überhaupt, wie und warum bestimmte Ansätze wirken oder nicht.¹⁴²

Wenn also Wirkungsnachweise so schwierig sind, sollte dann ganz darauf verzichtet werden? Politisch wäre dies sicher heikel. Was während einer Intervention im Einzelnen getan wird, mag auf einer Mikroebene sinnvoll und plausibel erscheinen, doch nachhaltige Legitimität gewinnen Interventionen über den Nachweis, dass ihr postulierter Beitrag umfassenderen Zielen auf der Makroebene dient.¹⁴³ Mit Projektevaluationen der üblichen Art mögen sich zwar positive Wirkungen auf lokaler Ebene feststellen lassen, mehr aber auch nicht.¹⁴⁴ Aus Gründen der Rechenschafts-

Gerard F. Powers, *Strategies of Peace. Transforming Conflict in a Violent World*, Oxford/New York: Oxford University Press, 2010, S. 65–89.

141 Siehe Rog, »When Background Becomes Foreground« [wie Fn. 94]; Michael Woolcock, »Using Case Studies to Explore the External Validity of »Complex« Development Interventions«, in: *Evaluation*, 19 (2013) 3, S. 229–248; Tobias Denskus, »Challenging the International Peacebuilding Evaluation Discourse with Qualitative Methodologies«, in: *Evaluation and Program Planning*, 35 (2012) 1, S. 148–153.

142 »From a research or social science perspective, the results of evaluations are often vague and uncertain, and evaluations rarely can make confident claims backed by solid evidence and research methodologies. Often, the main concern with evaluation results is not negative evaluations that find problems with peacebuilding programs; it is that evaluations do not reach any strong conclusions, positive or negative.« Andrew Blum, *Improving Peacebuilding Evaluation: a Whole-of-Field Approach*, Washington, D.C.: United States Institute of Peace, Juni 2011, S. 4.

143 Dazu Daniela Körppen, *Peace and Conflict Impact Assessment – Über die Utopie einer einheitlichen Methode*, in: Christoph Weller (Hg.), *Zivile Konfliktbearbeitung. Aktuelle Forschungsergebnisse*, Duisburg: Institut für Entwicklung und Frieden, 2007 (INEF-Report Nr. 85/2007), S. 27–34 (30).

144 So kommt eine Querschnittsauswertung der Evaluierungsberichte von acht Entwicklungsprojekten und -programmen des BMZ im Themenfeld Krisenprävention und Friedensaufbau zu dem Schluss: »Die Evaluierungen machen deutlich, dass viele Vorhaben zur Reduzierung von Konfliktdynamiken und friedlichem Zusammenleben auf lokaler Ebene beitragen. Beiträge zu nationalen Friedensprozessen werden eher punktuell geleistet [...]. In keinem Fall gelingt einem Vorhaben ein messbarer Beitrag zu einer Verringerung von Konfliktdynamiken und Gewalt auf nationaler Ebene. In der Tat stellen einige Evaluierungen fest, dass viele innovative Ansätze auf der Mikroebene verbleiben.« Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) (Hg.), *Synthesebericht Krisenprävention und Friedensentwicklung*, Eschborn,

legung und nachträglichen Legitimierung mag die Politik dennoch auch auf wenig aussagekräftige Evaluationen nicht völlig verzichten wollen.

Schon aus Aufwands- und Kostengründen werden Wirkungsevaluationen nur in ausgewählten Fällen möglich sein. Leitfragen für die Auswahl könnten sein: Ist die Maßnahme strategisch in dem Sinne relevant, dass von ihr eine beträchtliche Wirkung auf Konfliktprävention und Friedensaufbau erwartet werden kann? Ist das Vorhaben in dem Sinne innovativ, dass ein neuer Ansatz erprobt wird, oder handelt es sich um eine spezifische Intervention, die in anderen Kontexten wirksam war und deren Übertragbarkeit beurteilt werden soll?¹⁴⁵

Statt eine Vielzahl methodisch mehr oder minder anspruchsvoller Evaluationen zu betreiben, könnte bereits eine weniger aufwendige Programmqualitäts-einschätzung hilfreich sein und Lerneffekte mit sich bringen.¹⁴⁶ Wenn man in einer solchen prospektiven Designevaluation auf unvermeidlich ungewisse Wirkungseinschätzungen angewiesen bleiben wird, wie sollten sie dann konzipiert sein? »Systemisch«, lautet eine Antwort darauf: Nicht lineare Ursache-Wirkungs-Zusammenhänge in Wirkungsketten sollten analysiert werden, sondern wie bestimmte Interventionen sich auf ein komplexes Konfliktsystem auswirken. Wie können in einem solchen System die treibenden Kräfte beeinflusst, wie die Interaktionen verändert werden? Evaluation ist in diesem Verständnis ein dauerhaft an politische Entscheidungen gekoppelter, sowohl rückblickender als auch vorausschauender Prozess, in dessen Verlauf Interventionen den Veränderungen im Konfliktsystem angepasst werden müssen.¹⁴⁷ Dies

Dezember 2011, S. 60. Bei den untersuchten Vorhaben handelt es sich um die Wiedereingliederung von Kombattanten (Nepal, Ruanda, Sierra Leone), Friedenserziehung (Sri Lanka), Transitional Justice/Konfliktbearbeitung/Polizeireform (Guatemala), Förderung von Ernährungssicherheit, regionaler Kooperation und Stabilität (Südkaucasus), Förderung des Dialogs zwischen Staat und Zivilgesellschaft (Kolumbien), konfliktsensible sozio-ökonomische Entwicklung (Senegal).

145 Zu diesen Kriterien im Kontext der Entwicklungszusammenarbeit siehe Alexandra Caspari/Ragnhild Barbu, *Wirkungsevaluationen. Zum Stand der internationalen Diskussion und dessen Relevanz für Evaluierungen der deutschen Entwicklungszusammenarbeit*, Bonn: Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung, 2008 (Evaluation Working Papers), S. 35.

146 Siehe Reimann/Chigas/Woodrow, *An Alternative to Formal Evaluation of Peacebuilding* [wie Fn. 55].

147 Zu dieser Perspektive siehe Woodrow/Chigas, *Connecting the Dots* [wie Fn. 98]. Ähnliche Einsichten und Ansätze finden sich in der sogenannten *Developmental Evaluation* als Misch-

erfordert Monitoring-Systeme und die institutionelle Fähigkeit, Lernerfahrungen nachhaltig einzuspeisen und eingesetzte Instrumente situativ zu modifizieren. Auch ein solcher auf die Verbesserung von Entscheidungen gerichteter Prozess kann das gewünschte Ergebnis von Interventionen nicht garantieren, erhöht aber vielleicht die Chancen hierfür.¹⁴⁸

Der Ruf nach Evaluation betrifft meist die Wirkungen von Interventionen. Mit strategischer Politik-evaluation lassen sich dagegen sowohl die zentralen Inhalte als auch der Prozess der Politikformulierung und -umsetzung im Hinblick auf die strategische Kohärenz untersuchen:¹⁴⁹ Was sind die erklärten Ziele, was die eingesetzten Mittel, welches die zugrundeliegenden Annahmen? Sind die Ziele

form formativer und summativer Evaluation. Diese geht davon aus, dass Klarheit, Genauigkeit und Messbarkeit von vorher fest definierten Zielen und deren Indikatoren kontraproduktiv wirken können und erst während einer Intervention (und dann von allen Beteiligten) bestimmt werden sollten. Denn eine allzu rigide Anwendung, gerade in der Anfangszeit einer Intervention, kann deren Lernkurve beeinträchtigen und Innovation hemmen. Aufgrund nicht vorhersehbarer dynamischer Interaktionen können klare Ziele nicht immer von Beginn an definiert werden. Nicht die versprochenen Ergebnisse, sondern die Qualität der Fragen sollte daher bei der Auswahl der Projektanträge beachtet werden. Meist geben die Antragsteller vor, mehr zu wissen, als dies der Fall ist. Die Alternative *Development Evaluation* als periodische Reflexion verbindet Rückschau mit Vorausschau, um damit die Realität offener, nicht-linearer Systeme angemessener zu berücksichtigen. Siehe Quinn Patton, *Developmental Evaluation* [wie Fn. 111], S. 5, 85; Frances Westley/Brenda Zimmerman/Michael Quinn Patton, *Getting to Maybe. How the World is Changed*, Toronto 2007, S. 83f; siehe auch Tom Ling, »Evaluating Complex and Unfolding Interventions in Real Time«, in: *Evaluation*, 18 (2012) 1, S. 79–81.

148 »Man sollte wirklich versuchen, den Prozess und nicht das Ergebnis zu bewerten. Wir neigen dazu, die Qualität einer Entscheidung an die Qualität ihres Results zu knüpfen. Tatsächlich gibt es Verfahren, die mit höherer Wahrscheinlichkeit zu guten Entscheidungen führen als andere. In einer besseren Welt würden Entscheidungen nicht danach beurteilt, ob sie im Ergebnis gut oder schlecht sind.« Siehe »Die wichtigen Entscheidungen in der Euro-Krise wurden intuitiv getroffen«. Daniel Kahneman über die Illusion, gänzlich rational zu entscheiden«, in: *Internationale Politik*, 67 (2012) 6, S. 16–21 (21).

149 Einzelprojekte werden oft ohne Bezug zu einer Strategie durchgeführt. Dies hat eine Untersuchung von Peacebuilding-Aktivitäten Deutschlands, der Niederlande, Norwegens und Großbritanniens ergeben. Siehe Dan Smith, *Towards a Strategic Framework for Peacebuilding: Getting Their Act Together*, Overview Report of the Joint Utstein Study of Peacebuilding, Oslo: Ministry of Foreign Affairs, 2004 (Evaluation Report Nr. 1/2004), S. 10f.

realistisch? Stehen die eingesetzten Ressourcen im angemessenen Verhältnis zu den Zielen (oder müssen womöglich die Ziele im Lichte begrenzter Ressourcen revidiert werden)? Sind die Interventionen so fokussiert und gebündelt, dass von Anlage, Umfang und Intensität her über lokale Effekte hinaus überhaupt ein nennenswerter Beitrag zu Krisenvermeidung und Friedensaufbau erwartet werden kann?¹⁵⁰ Was wird umgesetzt, wie also sieht die tatsächliche Politik aus? Warum werden bestimmte Ziele nicht verfolgt? Das sind einige der Fragen, deren Beantwortung die Chance zu »strategischem Lernen« bietet.¹⁵¹

So verständlich der Ruf nach mehr Evaluation gerade aufgrund der schmalen Evidenzbasis für Krisenprävention und Friedensaufbau ist: Es sollte deutlich geworden sein, dass methodisch anspruchslöse, wenig aussagekräftige Evaluationen – vielleicht sollte man gar von »Pseudoevaluationen«¹⁵² sprechen – zwar die bestehende Evaluationsökonomie weiter aufblähen, jedoch nicht zu handlungsrelevantem Orientierungswissen beitragen werden. Dieses kann es im Grunde auch nur in der Form geben, dass aus dem eigenen Handeln die größtmögliche Lernerfahrung gezogen

150 Zur Problematik im britischen Kontext siehe Independent Commission for Aid Impact, *Evaluation of the Inter-Departmental Conflict Pool*, Juli 2012, wo es heißt (S. 19): »... the Conflict Pool is struggling to demonstrate real impact. In part, this is because of a lack of clarity as to what level of impact can be expected from relatively small sums directed towards influencing some of the world's most challenging conflict situations.« Der Conflict Pool ist ein gemeinsam von Außen- und Verteidigungsministerium sowie von der Internationalen Entwicklungsbehörde verwalteter Fonds zur Konfliktprävention, der 2011/12 über ein Budget von 256 Millionen britischen Pfund verfügte. Offenbar als Folge der Evaluation ist in den Strategischen Leitlinien für den Pool vom April 2013 zu lesen: »If the Pool is to make a difference on the ground some interventions need to be large enough to show significant impact on the wider conflict and stability dynamics.« Foreign and Commonwealth Office/Department for International Development/Ministry of Defence, *Conflict Pool Strategic Guidance*, April 2013, S. 19.

151 Zur »strategic policy evaluation« siehe Quinn Patton/Patrizi, »Strategy as the Focus for Evaluation« [wie Fn. 51], Zitat S. 19; siehe auch Julia Coffman/Tanya Beer, *Evaluation to Support Strategic Learning: Principles and Practices*, Washington, D.C.: Center for Evaluation Innovation, Juni 2011.

152 Dieser Begriff findet sich bei Alfred Uhl, *Methodenprobleme bei der Evaluation komplexerer Sachverhalte: Das Beispiel Suchtprävention*, in: Robert Koch-Institut/Bayerisches Landesamt für Gesundheit und Lebensmittelsicherheit (Hg.), *Evaluation komplexer Interventionsprogramme in der Prävention* [wie Fn. 88] S. 57–78 (61f).

wird.¹⁵³ Angesichts der überaus komplexen Realität von Außenpolitik kann Evaluation zwar zu einer stärker »evidenzbeeinflussten« Politik beitragen.¹⁵⁴ Doch wird es kein gesichertes Wissen darüber geben können, wie und mit welchen Mitteln sich bestimmte außenpolitische Ziele erreichen lassen.

153 »Creating a civic culture in which people understand that things seldom work as expected, especially when trying to solve complex, intractable problems, would support a shift toward learning.« Westley/Zimmerman/Quinn Patton, *Getting to Maybe. How the World is Changed* [wie Fn. 147], S. 183.

154 Wir verwenden diesen Ausdruck statt des gängigen, ein eher technokratisches Verständnis von Politik suggerierenden Begriffs der evidenzbasierten Politik. Zum Begriff der »evidence-influenced politics« siehe Kenneth Prewitt/Thomas A. Schwandt/Miron L. Straf (Hg.), *Using Science as Evidence in Public Policy*, Washington, D.C.: The National Academies Press, 2012, S. 14.