

HSFK-Report Nr. 8/2016

Auslandseinsätze evaluieren

Wie lässt sich Orientierungswissen zu humanitären Interventionen gewinnen?

Matthias Dembinski/Thorsten Gromes

Die Autoren danken der Deutschen Stiftung Friedensforschung für die Förderung des Projekts „Ein Datensatz der humanitären militärischen Interventionen nach dem Zweiten Weltkrieg“ (FP 02/14 – PS 01/12-2013), dessen Ergebnisse der Report aufgreift.

© Leibniz-Institut Hessische Stiftung Friedens- und Konfliktforschung (HSFK)

Adresse:

HSFK • Baseler Straße 27–31 • 60329 Frankfurt am Main • Deutschland

Telefon: +49 69 959104-0 • Fax: +49 69 558481

E-Mail: dembinski@hsfk.de • gromes@hsfk.de

Internet: www.hsfk.de

ISBN: 978-3-946459-06-4

6,00 €

Zusammenfassung

Für Projekte der Entwicklungszusammenarbeit oder des Zivilen Friedensdienstes ist Evaluation bereits gängige Praxis. Evaluierungen sollen darüber Aufschluss geben, ob die betrachteten Programme die gesetzten Ziele erreichen und im Sinne der Auftraggeber wirken. Die Befunde können dazu dienen, die Projekte zu bewerten und in ihrem Licht das Vorgehen künftiger Vorhaben anzupassen oder solche Programme ganz zu unterlassen. Ein im Januar 2016 dem Bundestag vorgelegter Gesetzentwurf sieht vor, nun auch Auslandseinsätze der Bundeswehr zu evaluieren. Wir begrüßen die im Gesetzentwurf genannte Absicht, aus den Erfahrungen bisheriger Auslandseinsätze „Lehren für künftige Missionen“ zu ziehen. Einsichten in Erfolgsbilanz und -bedingungen von Auslandseinsätzen scheinen umso wichtiger, als der Bundespräsident wie auch der Außenminister und die Verteidigungsministerin ein stärkeres militärisches Engagement Deutschlands fordern. Die Auslandseinsätze der Bundeswehr stecken ein weites Feld ab und reichen von Katastrophenhilfe über Evakuierungen eigener Staatsbürger bis zur Teilnahme an Anti-Terror-Einsätzen oder internationalen Missionen der Friedenserhaltung. Ein Großteil fiel in der Vergangenheit und fällt absehbar auch in Zukunft in die Kategorie der sogenannten humanitären militärischen Interventionen, bei denen die eigenen Soldaten in ein anderes Land geschickt werden, um dort unter Androhung oder Ausübung militärischen Zwangs die Bürger des Ziellands vor Gewalt zu schützen. Diese Einsätze sind besonders umstritten. Orientierungswissen über ihre Erfolgswahrscheinlichkeiten und -bedingungen ist daher umso wichtiger.

Wir befürchten, dass die vorgesehene Art der Evaluation dem Anliegen, aus bisherigen Erfahrungen zu lernen, nicht gerecht werden kann. Die Begründung des Gesetzentwurfs sieht die Evaluation als wichtigen „Baustein, um eine nachhaltige politische Unterstützung von Einsätzen im Bundestag zu erreichen“. Das lässt Zweifel daran aufkommen, ob sie ergebnisoffen geschehen wird. Zudem setzt der Gesetzentwurf auf eine Evaluation der einzelnen Einsätze, ohne deren Gesamtmenge zu betrachten und betont die Ziel- und nicht die Wirkungsorientierung der Evaluierung.

Die Evaluation eines noch laufenden Einsatzes kann dazu dienen, ihn neu auszurichten, was auch zum Abbruch der Intervention führen kann. Solche begleitenden oder Zwischen-evaluationen sind aber kaum in der Lage, längerfristige Entwicklungen im Zielland zu beobachten und zu bewerten. Dazu sind Evaluationen erforderlich, die erst einige Zeit nach Abschluss der Intervention erfolgen.

Laut Gesetzentwurf soll die Evaluation ermitteln, wieweit die im Auftrag definierten Ziele erreicht wurden. Ein solches Vorgehen kann zu Einsichten darüber führen, wie sich Abläufe verbessern ließen und welche Ziele realistisch zu erreichen wären. Gleichwohl weist eine solche zielorientierte Evaluation gravierende Nachteile auf. Mandate sind oft ungenau oder unvollständig formuliert oder verändern sich mit der Zeit, was eine fallübergreifende Bilanzierung erschwert. Vor allem hängt es dann stark von der Formulierung des Mandats ab, ob ein Einsatz als Erfolg oder Misserfolg gilt. Die gleiche Entwicklung müsste je nach Auftrag unterschiedlich bewertet werden.

Angesichts dieser Nachteile einer zielorientierten Evaluation sprechen wir uns dafür aus, ergänzend wirkungsorientiert zu evaluieren und die relevanten intendierten und nicht-intendierten Wirkungen einer humanitären militärischen Intervention zu untersuchen. Im Mittelpunkt steht dann, wieweit die Intervention ihr Kernanliegen umsetzen und die Bürger des Ziellands vor Gewalt schützen konnte. Vier einheitliche Kriterien sollten die Entwicklungen in allen Zielländern erfassen: erstens die Dauer des Gewaltkonflikts, auf den die Intervention reagiert, zweitens die Intensität der tödlichen Gewalt vor und während des Eingreifens, drittens der Rückfall in einen Gewaltkonflikt bis zu fünf Jahre nach der Intervention und viertens die Destabilisierung von Nachbarstaaten.

Die vier Maßstäbe registrieren relevante Entwicklungen in den Ziel- und Nachbarländern der Intervention, sagen aber noch nichts dazu, ob die beobachteten Veränderungen durch die Intervention verursacht wurden. Eine solche Attribution, d.h. das kausale Zurückführen einer Entwicklung auf die Intervention, ist die zentrale Herausforderung der wirkungsorientierten Evaluation. Aus einem einzelnen Fall heraus lässt sich die Attribution nur annäherungsweise bewerkstelligen. Daher empfehlen wir, über isolierte Einzelevaluationen der jeweils gerade relevanten Einsätze hinauszugehen. Bundestag und -regierung sollten einmalig eine Evaluation aller humanitären militärischen Interventionen in Auftrag geben, nicht nur der Einsätze mit deutscher Beteiligung. Die Einsichten würden dann auf einer breiteren Basis von Erfahrungen stehen und größeres Orientierungswissen bereitstellen, wenn eine Entscheidung über einen neuen Einsatz ansteht.

Die vorgeschlagene Evaluation aller Intervention umfasst zwei fallübergreifende Vergleiche. Das erste Verfahren ermittelt die durchschnittliche Wirkung der humanitären militärischen Intervention, indem es die Gruppe aller Interventionen mit einer zu konstruierenden Gruppe von Fällen ohne solche Eingriffe vergleicht. Die Vergleichsgruppe muss dabei in ihrer Zusammensetzung der Interventionsgruppe möglichst gleichen, und zwar mit Blick auf Faktoren, die Dauer und Intensität von Gewaltkonflikten beeinflussen. Das zweite fallübergreifende Verfahren vergleicht nur die Interventionsfälle miteinander, um die Konstellationen von Bedingungen von Erfolg und Scheitern zu ermitteln. Anschließend Fallstudien können helfen, die Befunde dieser fallübergreifenden Verfahren abzusichern.

Die zu erwartenden Ergebnisse einer solchen Evaluation versprechen Orientierungswissen für Entscheidungen über künftige Einsätze, verkürzen jedoch keineswegs die Verantwortung der Politiker auf ein Durchwinken einer vermeintlich alternativlosen, nicht zu hinterfragenden Beschlussvorlage. Die Befunde zur durchschnittlichen Wirkung humanitärer militärischer Interventionen geben Wahrscheinlichkeiten an, schaffen aber keine Gewissheit über die weitere Entwicklung im zu entscheidenden Fall. Die Ergebnisse der Bedingungsanalyse geben selbst keine Auskunft darüber, welche Bedingungen im akuten Fall vorliegen. Dazu fehlt es in der Regel an gesicherten Informationen über die Situation im etwaigen Zielland. Eine Evaluation der bisherigen humanitären militärischen Interventionen kann Politikern helfen, besser informiert zu entscheiden. Dabei gilt es zu bedenken, dass die Erfolgsaussicht kaum das alleinige Entscheidungskriterium ist. Dazu tritt ein ganzes Bündel außen- und innenpolitischer Erwägungen wie Bündnissolidarität, finanzielle Belastungen oder anstehende Wahlen.

Inhaltsverzeichnis

1. Die Entscheidung zur Evaluation von Auslandseinsätzen	1
2. Die Evaluation von Auslandseinsätzen: Ein sinnvolles Unterfangen?	4
2.1 Erster Einwand: Evaluationen dienen der Legitimierung	5
2.2 Zweiter Einwand: Evaluationen delegieren Verantwortung	6
2.3 Dritter Einwand: Evaluationen können keine Wirkungen erfassen	7
2.4 Vierter Einwand: Evaluationsbefunde lassen sich nicht verallgemeinern	8
3. Die Gesamtmenge humanitärer militärischer Interventionen	9
4. Forschungsstrategische Entscheidungen und ihre Konsequenzen	13
4.1 Wer sollte wann evaluieren?	13
4.2 Ziel- oder wirkungsorientiert evaluieren?	15
4.3 Woran Erfolg und Scheitern messen?	20
4.4 Wie lässt sich die Wirkung von Interventionen ermitteln?	24
5. Fazit: Für eine ergebnisoffene wirkungsorientierte Evaluation	30
Literatur	32

1. Die Entscheidung zur Evaluation von Auslandseinsätzen¹

Ein Gesetzentwurf vom Januar 2016 verpflichtet die Bundesregierung zur Vorlage von Evaluationsberichten über abgeschlossene oder zur Verlängerung anstehende Einsätze der Bundeswehr (Deutscher Bundestag 2016: 3). Der Gesetzentwurf folgt damit dem Rat einer vom Bundestag eingesetzten Kommission (Deutscher Bundestag 2015: 26–27), aber auch dem gesellschaftlicher Akteure wie der Evangelischen Kirche Deutschlands (Kirchenamt der EKD 2013), der Herausgeber des Friedensgutachtens (Schoch et al. 2007: 11–12) sowie einzelner Fachpolitiker wie insbesondere dem ehemaligen Abgeordneten Winfried Nachtwei von Bündnis 90/Die Grünen. Vorausgegangen war eine längere Kontroverse über die Grenzen und Möglichkeiten der Evaluation von Auslandseinsätzen.²

Das Stichwort „Auslandseinsätze“ deckt ein breites Spektrum verschiedener Tätigkeiten ab, von der Katastrophenhilfe und Evakuierung eigener Staatsbürger über die Beobachtung von Waffenstillständen bis hin zu Ausbildungsmissionen, Anti-Terror-einsätzen oder Einsätze zum Schutz Verbündeter. Zu all diesen Einsatzarten kann Evaluierung wichtiges Orientierungswissen bereitstellen, was gerade für umstrittene Einsätze sinnvoll wäre. Besonders umstritten ist die Beteiligung Deutschlands an sogenannten humanitären militärischen Interventionen. Da Bundestag und -regierung absehbar vor der Entscheidung stehen, ob sich Deutschland an weiteren dieser brisanten Einsätze beteiligen soll, halten wir deren Evaluation für am wichtigsten. Einsichten über Erfolg und Erfolgsbedingungen von humanitären militärischen Interventionen können dabei helfen, auf besserer Informationsbasis über ein solches Eingreifen zu entscheiden.

Die Bundesrepublik wirkt seit Anfang der 1990er Jahre an Interventionen mit, die nicht ausschließlich, aber auch humanitäre Ziele verfolgen. Den „Zweiklang, sicherheitspolitisch und humanitär geboten“ (De Maizièrè 2012), werden auch zukünftige Einsätze beherzigen. Dabei ist die seit den frühen 1990er Jahren friedenspolitisch und friedenswissenschaftlich kontrovers geführte Diskussion um humanitäre militärische Interventionen von einer Ungleichzeitigkeit geprägt. Während wir aufgrund der niederschmetternden Ergebnisse der westlichen Interventionen im Irak, in Afghanistan und in Libyen um das Risiko des Scheiterns wissen und dieses Wissen den kritischen Stimmen in einer ohnehin skeptischen Öffentlichkeit weiteren Auftrieb gibt, bereiten die Regierenden ein stärkeres militärisches Engagement *out of area* vor. Deutschland müsse sich als großer und global vernetzter Staat außen- und sicherheitspolitisch „früher, substantieller und entschiedener einbringen“, so der Bundespräsident, fast wortgleich sekundiert vom Außenminister auf der Münchener

1 Für konstruktive Kommentare danken wir Lothar Brock, Caroline Fehl, Charlotte Felbinger, Marco Fey, Ines Gundlach, Gregor Hofmann, Peter Kreuzer, Bernhard Moltmann, Dirk Peters, Stephanie Radwanski, Bruno Schoch und Jonas Wolff.

2 Weder die deutschen Progress Reports zu Afghanistan (Bundesregierung 2014) noch der UN-Expertenbericht zu Friedensmissionen (United Nations General Assembly 2015) stellen Evaluation im hier verstandenen Sinn dar.

Sicherheitskonferenz 2014.³ Diese Forderung zielt in erster Linie auf entschiedenere Maßnahmen der Konfliktprävention und Friedensschaffung. Daneben, und dies ist die neue Tonlage, geht es um ein stärkeres militärisches Engagement. All dies war noch vor der Flüchtlingskrise. Die Folgerung lautet seitdem, es müsse gelingen, das Problem an der Wurzel zu packen und die Ursachen von Flucht und Vertreibung zu bearbeiten: Staatsversagen, Bürgerkriege und die Gewalt dysfunktionaler Regierungen gegen ihre eigene Bevölkerung. Damit steht Deutschland absehbar immer wieder vor der Entscheidung, ob es Truppen einsetzen soll, um „Fremde zu retten“ (Wheeler 2000), wie das zentrale Anliegen humanitärer militärischer Interventionen umrissen wurde.

Entscheidungsträger wissen, dass derartige Einsätze scheitern und mehr Leiden verursachen können, als Gutes zu erreichen. Sie wissen auch, dass beginnende Bürgerkriege, bei denen der Versuch einer externen, zwangsbewehrten Befriedung unterbleibt, nicht einfach „ausbrennen“. Jedoch verfügen sie kaum über Informationen, die vorab Aufschluss darüber geben, ob und wie eine Intervention in einer konkreten Situation erfolgreich sein könnte.

Bevor sich Deutschland auf den in München vorgezeichneten Weg begibt, wären seine Entscheidungsträger gut beraten, entschiedene Anstrengungen zu unternehmen, um aus bisherigen Erfahrungen zu lernen. Mit der Forderung nach einer regelmäßigen Evaluation deutscher Auslandseinsätze scheint der Bundestag dem nachzukommen. Allerdings, dies legen sowohl die vorausgegangenen Debatten als auch Formulierungen des Gesetzentwurfs nahe, springt das deutsche Parlament zu kurz. Denn wie die Begründung des Gesetzentwurfs betont, soll die Evaluation dazu dienen, „eine nachhaltige politische Unterstützung von Einsätzen im Bundestag zu erreichen“ (Deutscher Bundestag 2016: 6). Es scheint demnach nicht um eine ergebnisoffene Evaluation zu gehen.

Zudem legt die Vorlage den Schwerpunkt auf einzelne Einsätze und deren Zielerreichung. Die zielorientierte Evaluation einzelner Einsätze kann sinnvoll sein, wenn es darum geht, die Kosteneffizienz und die internen Abläufe zu verbessern und zu entscheiden, ob die Intervention überhaupt fortgesetzt werden sollte. Um Orientierungswissen über die Erfolgsmöglichkeiten und die Wahrscheinlichkeiten des Scheiterns humanitärer militärischer Interventionen zu generieren, erscheint es uns allerdings notwendig, die gesamte Menge derartiger Einsätze zu bilanzieren und dabei nicht nur die Erreichung der selbstgewählten Ziele, sondern auch die darüber hinaus gehenden Wirkungen ergebnisoffen zu evaluieren.

In der Literatur zur Evaluation etwa im Bereich der Konfliktprävention ist es umstritten, ob und wie es gelingen kann, etwas über die Wirkungen von Eingriffen in komplexe politische und soziale Zusammenhänge zu erfahren (Rudolf/Lohmann 2013). In diesem Report setzen wir uns mit den Schwierigkeiten und Grenzen einer solchen Evaluation auseinander, argumentieren aber, dass sich der Versuch lohnt. Anhand von Daten über alle humanitären militärischen Interventionen seit 1945, die wir im Rahmen eines HSFK-Projekts erhoben haben, diskutieren wir zentrale Herausforderungen einer

3 Gauck (2014); Auswärtiges Amt (2014); siehe auch Bundesministerium der Verteidigung (2014).

Evaluierung. Die Literatur zur Evaluation legt es nahe, dass man beim Versuch, Erfolg zu definieren und zu erfassen, eine Reihe von Entscheidungen fällen muss. Erstens ist festzulegen, wer evaluiert und zu welchem Zeitpunkt dies geschehen soll. Zweitens geht es um die Entscheidung zwischen ziel- oder wirkungsorientierter Evaluierung. Drittens sind die erforderlichen Kriterien für Erfolg und Scheitern festzulegen, und viertens stellt sich das Problem der Attribution: Wie lässt sich ein Zusammenhang zwischen dem Eingriff und der beobachteten Veränderung nachweisen?

Die für die Evaluation der humanitären militärischen Interventionen diskutierten Vor- und Nachteile des jeweiligen Vorgehens lassen sich auf die Evaluierung anderer Sub-Kategorien von Auslandseinsätzen beziehen. Die Evaluation etwa von Ausbildungsmissionen steht vor ganz ähnlichen Herausforderungen. Für humanitäre militärische Interventionen wie auch für andere Arten von Einsätzen gilt gleichermaßen: Selbst eine ergebnisoffene, umfassende und methodisch anspruchsvolle Evaluation wird die Unsicherheitsmarge über die Erfolgsaussichten künftiger Missionen nicht auf null drücken. Aber eine Auswertung bisheriger Erfahrungen könnte sie reduzieren.

Der Report setzt sich zunächst mit Einwänden gegen eine Evaluation von Auslandseinsätzen der Bundeswehr im Allgemeinen und von humanitären militärischen Interventionen im Besonderen auseinander (Kapitel 2). Danach stellt er die Gesamtmenge humanitärer militärischer Interventionen vor (Kapitel 3). Im Hauptteil der Arbeit wollen wir entlang des skizzierten Entscheidungspfades die Vor- und Nachteile der verschiedenen Optionen diskutieren und darlegen, welche Entscheidungen wir mit Blick auf eine Vermessung von Erfolg am sinnvollsten halten (Kapitel 4). Im Schlussteil bilanzieren wir, was sich Politik und Öffentlichkeit von Evaluationsergebnissen versprechen können (Kapitel 5).

Um Missverständnissen vorzubeugen: Der Report befasst sich allein mit dem Für und Wider verschiedener Varianten der Evaluierung, präsentiert aber keine Ergebnisse der Evaluation. Auch haben unsere bereits publizierten vorläufigen Bilanzierungen humanitärer militärischer Interventionen (Gromes/Dembinski 2013; Dembinski/Gromes 2015) die im Folgenden skizzierte Evaluation nicht vorweggenommen.

2. Die Evaluation von Auslandseinsätzen: Ein sinnvolles Unterfangen?

Einer gängigen Definition zufolge ist Evaluation die „Analyse und Bewertung eines Sachverhalts, vor allem als Begleitforschung einer Innovation. [...] Evaluation [ist] Effizienz- und Erfolgskontrolle zum Zweck der Überprüfung der Eignung eines in Erprobung befindlichen Modells. [...] Evaluation ist die Einschätzung der Wirkungsweise, Wirksamkeit und Wirkungszusammenhänge“ (Brockhaus 1997: 716). Auf einen weiteren Aspekt macht eine andere Definition aufmerksam: Danach ist Evaluation „the systematic investigation of the merit or worth of an object (program) for the purpose of reducing uncertainty in decision-making“ (Mertens 1998: 219, zitiert nach Stockmann/Meyer 2014: 72). Es handelt sich um den systematischen Einsatz von Methoden, um die Effizienz, Zielerreichung und Wirksamkeit von Eingriffen in soziale Prozesse zu erheben, um daraus Folgerungen für künftige Eingriffe abzuleiten. Evaluation ist demnach sowohl Bewertungs- als auch Steuerungsinstrument (Stockmann/Meyer 2014: 15f).

Evaluation ist über den privatwirtschaftlichen Bereich hinaus in der öffentlichen Verwaltung allgegenwärtig. Dieser Trend ist verständlich, denn zum einen wollen Entscheidungsträger wissen, ob das Geld der Steuerzahler sinnvoll angelegt ist. Zum anderen sehen sie sich zunehmend in der Pflicht nachzuweisen, dass die finanzierten Programme die intendierten Ziele und Wirkungen erreichen. Evaluiert werden unter anderem schulische Unterrichtsformen, Arbeitsabläufe in Verwaltungseinheiten, Forschungsinstitute und Evaluationsvorgänge selbst. Auch die nationalen und europäischen Projekte und Programme, die in den Bereich der Außenbeziehungen fallen, werden regelmäßig evaluiert. Bei Projekten der Entwicklungspolitik ist dies seit langem Praxis und Programme, die in den Bereich der Konfliktprävention und der Friedensbildung fallen, werden seit geraumer Zeit auf den Prüfstand gestellt. Selbst der mit vergleichsweise bescheidenen öffentlichen Mitteln ausgestattete Zivile Friedensdienst wurde evaluiert (Paffenholz et al. 2011). Nun also auch die Auslandseinsätze der Bundeswehr.

Mit der Nachfrage nach methodengestützter Bewertung öffentlich finanzierter Vorhaben entwickelte sich eine breite Landschaft an Institutionen, die sich auf die Durchführung von Evaluationen spezialisieren oder deren Konzepte und Methoden entwickeln. In Deutschland wurde 1997 die Gesellschaft für Evaluation (DeGEval) als Dachverband von Organisationen gegründet, die sich in den verschiedensten Politikbereichen mit der Evaluation entweder praktisch oder begleitend befassen. Im englischsprachigen Raum erscheinen ca. ein Dutzend Fachzeitschriften zu dem Thema. Programmatisch führend im Bereich der Entwicklungszusammenarbeit und Friedensschaffung ist die OECD (Organisation für wirtschaftliche Entwicklung und Zusammenarbeit) und in Deutschland die Gesellschaft für Internationale Zusammenarbeit (GIZ). Beide erarbeiteten erste Handreichungen, werteten in Metastudien Evaluationserfahrungen aus und entwickelten auf dieser Grundlage ihre Anleitungen für Evaluationen weiter (OECD 2012; GIZ 2012). Über Kompetenzen bei der

Durchführung von Evaluationen verfügt insbesondere das Deutsche Evaluierungsinstitut der Entwicklungszusammenarbeit (DEval), das über 30 Fachkräfte beschäftigt.⁴

Trotz der Allgegenwart der Evaluation und der fortschreitenden Professionalisierung der Evaluationsagenturen ist dieses Unterfangen sowohl in der Politik als auch in der Wissenschaft nach wie vor umstritten. Dies gilt gerade für die Programme, die im Bereich der Außenbeziehungen angesiedelt sind, und hier insbesondere für die Maßnahmen im Bereich Friedensschaffung und Konfliktprävention.⁵ So war dem Gesetzentwurf vom Januar 2016 wenig überraschend eine kontroverse Debatte über die Grenzen und Möglichkeiten der methodischen und auf Handlungsorientierung zielenden Bewertung der Auslandseinsätze vorausgegangen. Bei den Debatten in den Jahren 2011, 2013 und 2015, aber auch in akademischen Diskussionen wurden vier grundsätzliche Einwände gegen die Evaluation der Auslandseinsätze der Bundeswehr geäußert: Erstens dienen Evaluationen der Legitimitätsbeschaffung, verwischen zweitens politische Verantwortung, seien drittens kaum in der Lage, Wirkungen und Wirkungszusammenhänge zu erfassen, und erlaubten viertens über den Einzelfall hinaus wenig generalisierbare Schlussfolgerungen. Wir setzen uns im Folgenden mit diesen Einwänden auseinander.

2.1 Erster Einwand: Evaluationen dienen der Legitimierung

Der erste Einwand lautet, Evaluationen seien stark von ihren Auftraggebern gesteuert, und diesen ginge es eher um Legitimationsbeschaffung als um eine ergebnisoffene Bewertung. Dieser Einwand ist nicht von der Hand zu weisen. Tatsächlich betonen Standards für Evaluationen, dass sich diese an den erklärten Zwecken sowie am Informationsbedarf der vorgesehenen Nutzer, in der Regel die Auftraggeber, orientieren sollten. Zudem seien Evaluationen so durchzuführen, dass sie eine möglichst hohe Akzeptanz der Betroffenen erreichen (DeGEval – Gesellschaft für Evaluation 2008). Auftraggeber und Betroffene sind nun in der Regel nicht daran interessiert, die Wirkungsweise der von ihnen auf die Schiene gesetzten oder durchgeführten Programme grundsätzlich in Zweifel zu ziehen. Sie wünschen sich vielmehr eine Überprüfung, die den Blick auf die Optimierung interner Abläufe richtet, und nicht eine, die die Wirksamkeit eines Programms oder den Nutzen einer Einrichtung allgemein infrage stellen könnte (Palumbo/Nachmias 1983: 76). Andere Untersuchungen kommen zu ähnlichen Befunden. Danach stünden Entscheidungsträger, Programmorganisatoren und damit auch die Evaluationsinstanzen unter einem hohen Druck, positive Ergebnisse zu präsentieren oder zumindest die grundsätzliche Nützlichkeit eines Programmes nachzuweisen. Dieser Druck beeinflusse die Aufgabenstellung der Evaluation, die Auswahl von Indikatoren und die Bewertung von Ergebnissen (Farina et al. 2013: 123; Williams et al. 2013: 7). Für die

4 Informationen sind abrufbar unter: www.deval.org.

5 Siehe die schriftliche Zusammenfassung und den Video-Mitschnitt einer Anhörung im Bundestag zur Evaluation von Krisenprävention unter: <http://bit.ly/1reLRkp> (26.4.2016).

Evaluationspraxis im Bereich der Konfliktprävention und Friedensschaffung konstatiert eine Metastudie von Peter Rudolf und Sascha Lohmann (2013: 26), Evaluierungen seien zwar „wenig aussagekräftig“, die Politik halte aber wohl aus Gründen der nachträglichen Legitimierung und Rechenschaft an ihnen fest.

Dass sich die Politik von Evaluationen eine legitimierende Funktion auch für Auslandseinsätze erhofft, zeigten die Bundestagsdebatten. So argumentierte etwa der Abgeordnete Wolfgang Götzer (CDU/CSU), Evaluationen seien geeignet, die Leistungen der Soldaten in der Öffentlichkeit sichtbar zu machen und „die längst fällige Debatte um Kosten und Nutzen von Auslandseinsätzen zu führen“ (Deutscher Bundestag 2013: 26962). In eine ähnliche Richtung, wenn auch unter politisch umgekehrten Vorzeichen, weist die Kritik der Abgeordneten Inge Höger (Die Linke): „Wir vertrauen [...] nicht darauf, dass eine umfassendere Unterrichtungspflicht die Schattenseiten von Krieg und Militär wirklich angemessen offenlegt“ (Deutscher Bundestag 2013: 26965).

Dienen Evaluationen also strukturell eher zur Legitimitätsbeschaffung? Und taugen sie allein aufgrund der Abhängigkeit von ihrem Auftraggeber und seiner Interessenlage nicht dazu, die Wirksamkeit von Programmen und Eingriffen kritisch zu hinterfragen? Eine solche Schlussfolgerung wäre deutlich überzogen. Sicherlich prägt der Auftrag die zu erwartenden Ergebnisse der Evaluation von Auslandseinsätzen. Bei der Vergabe könnte und sollte der Bundestag daher mitreden und im Sinne seines Kontrollauftrages sicherstellen, dass nicht nur interne Abläufe überprüft, sondern die intendierten und nicht-intendierten Wirkungen von Auslandseinsätzen kritisch und ergebnisoffen ermittelt werden. Eine solche wirkungsorientierte Evaluation schafft Orientierungswissen über (Miss-)Erfolge und deren Bedingungen.

2.2 Zweiter Einwand: Evaluationen delegieren Verantwortung

Ein zweiter Einwand fürchtet, die Evaluation leiste einem Trend der Bürokratisierung politischer Entscheidungen Vorschub. Die Entscheidungsfreiheit demokratisch legitimer Entscheidungsträger würde durch einen technokratischen und szientistischen Handlungsmodus ersetzt. Politische Verantwortung werde so verwischt. Diese Bedenken spielten in der Kontroverse um die systematische Überprüfung der Auslandseinsätze eine prominente Rolle. Bijan Djir-Sarai (FDP) erklärte in der ersten Debatte zu diesem Thema: „Sicherheitspolitik kann nicht ausgegliedert werden“ (Deutscher Bundestag 2011b: 12065). Ähnliche Befürchtungen sprach auch Roderich Kiesewetter (CDU/CSU) mit der Mahnung an, der Bundestag könne Sicherheitspolitik nicht „outsourcen“: „Wir müssen dazu stehen und dürfen nicht sagen: Die Wissenschaftler haben uns das empfohlen.“ (Deutscher Bundestag 2011b: 12062)

Diese Vorbehalte weisen auf einen berechtigten Einwand hin. Die Entscheidung über Auslandseinsätze der Bundeswehr betrifft den Kernbereich staatlicher Hoheitsrechte und darf nur von demokratisch legitimierten Entscheidungsträgern gefällt werden. Das ist ebenso selbstverständlich, wie dieser Einwand nicht gegen Evaluationen spricht. Denn die Vorstellung, eine Evaluierung könnte den politischen Entscheidungsträgern die Verant-

wortung abnehmen, überschätzt die Möglichkeit, mithilfe derartiger Methoden die Wirkung von Interventionen präzise abzuschätzen und Entscheidungen quasi automatisch und nur auf der Grundlage des durch Evaluationen gewonnenen Wissens zu treffen. Wie wir zeigen werden, ließen sich mit Hilfe guter Analysemethoden Muster erkennen und Wahrscheinlichkeiten angeben. Damit aber würden Evaluationen Wichtiges zu verantwortlicher Entscheidungsfindung beitragen. Denn auf derartige Informationen zu verzichten und Entscheidungen ins Blaue hinein zu treffen, wäre ebenso unverantwortlich, wie Entscheidungen einer externen Evaluationsinstanz zu überlassen. Tatsächlich spielte dieser Einwand in späteren Debatten des Bundestages kaum noch eine Rolle. Stattdessen betonten die Beiträge (Deutscher Bundestag 2013: 29691–26966), das durch Evaluationen gewonnene Hintergrundwissen sei wertvoll, um Einsätze wirksam mandatieren und kontrollieren zu können (Kiesewetter), um aus Fehlern der Vergangenheit zu lernen (Johannes Pflug, SPD) und um künftige Misserfolge zu verhindern (Omid Nouripour, Grüne).

2.3 Dritter Einwand: Evaluationen können keine Wirkungen erfassen

Der dritte Einwand bezweifelt, dass Evaluationen überhaupt Aussagen über die Wirkungen von Eingriffen in hochkomplexe soziale Prozesse machen können. Evaluationen eigneten sich eher dazu, interne Abläufe zu bewerten oder die Kosten alternativer Vorgehen abzuschätzen, könnten aber kaum die Wirkung von Interventionen ermitteln. Dieses Argument wird vor allem in der wissenschaftlichen Kritik an gängigen Evaluationsverfahren vorgebracht, die die Wirkungen von Eingriffen etwa im Bereich Konfliktprävention und Friedensbildung erfassen wollen. Diese Verfahren glichen, so Peter Rudolfs plakative Einschätzung, einem Stochern im Nebel.⁶ Zu einer ähnlich skeptischen Einschätzung kommt Hanns Maull mit Blick auf die Außenpolitikevaluierung, die seiner Auffassung nach eher dem Bereich der Kunst als dem der exakten Wissenschaft zuzuordnen sei (Maull 2008: 113).

Tatsächlich sieht sich eine Wirkungsanalyse mit einer Reihe von Schwierigkeiten konfrontiert. Sie müsste erstens ein Instrumentarium entwickeln, das in sich ständig wandelnden sozialen Systemen diejenigen mittel- und langfristigen Veränderungen beobachten kann, die möglicherweise von der Intervention hätten bewirkt werden können. Sie müsste zweitens die relevanten verzeichneten Veränderungen kausal in ein Verhältnis zur Intervention bringen. Wie wir zeigen werden, lassen sich im Rahmen von Evaluationen einzelner Maßnahmen bestenfalls Annahmen über Wirkungen falsifizieren, jedoch nicht Wirkungen nachweisen. Vielversprechender wäre der Weg über die Bewertung der Gesamtmenge eines Phänomens. Der ist allerdings aufwendig und kommt in der Evaluationspraxis nur selten zum Zuge.

6 www.bundestag.de/dokumente/textarchiv/2015/kw40_pa_zivile_krisenpraevention/387398 (26.4.2016).

2.4 Vierter Einwand: Evaluationsbefunde lassen sich nicht verallgemeinern

Der vierte Einwand bezweifelt die Generalisierbarkeit von Erkenntnissen. Selbst wenn es gelänge, ex post Aussagen über die Wirkung von Eingriffen in komplexe soziale Prozesse zu treffen, ließen sich daraus keine Folgerungen für andere Fälle ableiten. So weisen etwa Rudolf und Lohmann (2013: 6, 26) darauf hin, dass die Kontextgebundenheit einzelner Interventionen eine verlässliche Generalisierung der zu erwartenden Wirkungen unmöglich mache. Dieses Argument fand auch in den Bundestagsdebatten zu den Auslandseinsätzen einen starken Widerhall. So argumentierte beispielsweise Kiesewetter, jeder militärischer Einsatz habe seine Besonderheiten. Ähnlich äußerte sich Djir-Sarai: „Jeder Einsatz muss individuell betrachtet werden“. Reinhard Brandl (CDU/CSU) und Michael Groschek (SPD) bekräftigten, jeder Fall sei anders, ein Unikat (Deutscher Bundestag 2011b: 12061–6).

Auch dieser Einwand ist nicht falsch, muss aber in seiner Bedeutung eingeordnet werden. Sicherlich ließe der Fall A keine Rückschlüsse auf den Fall B zu, wenn beide grundverschieden wären. Je mehr Fall B dem Fall A gleicht, desto eher wären Rückschlüsse möglich. Zwar gleicht kein Auslandseinsatz dem anderen. Wenn man sich aber nicht mit Einzelfallevaluationen begnügt, sondern systematisch eine größere Gruppe von Fällen betrachtet, lassen sich Ähnlichkeiten und Muster feststellen und auf dieser Grundlage generalisierbare Aussagen treffen. Dabei wird es nie gelingen, den Ausgang einer zur Entscheidung anstehenden Intervention zu prognostizieren. Aber die systematische Auswertung der Erfahrungen würde es ermöglichen, Wahrscheinlichkeiten anzugeben und Warntafeln dann aufzustellen, wenn ähnlich gelagerte Interventionen in der Vergangenheit nicht erfolgreich waren.

Festzuhalten bleibt, dass Grenzen und Möglichkeiten der Evaluation von Auslandseinsätzen zwar weiterhin umstritten sind, im Zentrum des politischen Streits aber nicht mehr das „ob“, sondern das „wie“ steht. Zu welchem Zweck, wie und von wem sollten Evaluationen durchgeführt werden? Hier droht der Bundestag zu kurz zu springen, weil er sich auf interne Einzelfallevaluationen beschränkt, die zudem strikt zielorientiert sind und nicht zuletzt der Legitimitätsbeschaffung dienen sollen. Damit ließen sich am ehesten Abläufe prüfen, aber kaum etwas über Wirkungen aussagen.

Soll eine Evaluation die Fehler der Vergangenheit aufzeigen und Orientierungswissen für künftige Entscheidungen schaffen, dann scheint es geboten, die Gesamtmenge der Fälle humanitärer militärischer Interventionen zu untersuchen. Das nächste Kapitel stellt daher die Gesamtmenge der humanitären militärischen Interventionen nach dem Zweiten Weltkrieg vor.

3. Die Gesamtmenge humanitärer militärischer Interventionen

Die Ergebnisse jeder Gesamtevaluation hängen stark von der Definition humanitärer militärischer Interventionen ab und diese ist folglich Gegenstand des politischen Streits. Wie Grundsatzkritiker meinen, existierten diese Einsätze gar nicht, da „humanitär“ und „militärisch“ unvereinbar seien. In dieser Sichtweise kann es zwar militärische Erfolge geben, humanitäre Erfolge hingegen schließt sie aus. Dass wir der genannten Grundsatzkritik nicht folgen, machen wir mit der Vorlage dieses Reports deutlich. Gibt es keine humanitäre militärische Intervention, dann erübrigt sich jede Debatte über deren Evaluation. Zu einer solchen Debatte wollen wir aber beitragen; daher wählen wir eine Definition, die möglichst alle Varianten der humanitären militärischen Intervention erfasst und sie zugleich ausreichend gegenüber ähnlichen Einsätzen abgrenzt. Von folgender Festlegung gehen wir aus:

Bei humanitären militärischen Interventionen schicken Staaten Truppen in ein anderes Land, um unter Androhung oder Ausübung militärischen Zwangs erklärtermaßen die dortige Zivilbevölkerung zu schützen, die von einer durch Gewalt geprägten Notlage bedroht ist.⁷ Bei solchen Notlagen handelt es sich um einen bewaffneten Konflikt zwischen der Regierung und Rebellen, um Kämpfe zwischen nicht-staatlichen Gruppen oder um Übergriffe gegen die Zivilbevölkerung. Von einer humanitären militärischen Intervention sprechen wir nur, wenn alle vier definitorischen Merkmale vorliegen: 1) Auslandseinsatz, 2) Zwang, 3) von Gewalt geprägte Notlage, 4) erklärte Absicht, Fremde zu schützen.

Diese Definition besteht nicht darauf, dass humanitäre militärische Interventionen ausschließlich humanitäre Anliegen verfolgen, denn so verstanden dürfte die Gesamtmenge humanitärer militärischer Interventionen gegen null tendieren. In aller Regel sind die Motive der intervenierenden Staaten gemischt und umfassen neben der Rettung der Bürger anderer Staaten weitere Interessen. Beispielsweise kann es darum gehen, eine regionale Destabilisierung und dadurch ausgelöste Fluchtbewegungen zu verhindern. Allerdings gibt es Motive und Handlungen, die wir für unvereinbar halten mit der behaupteten Absicht, die Bürger eines anderen Landes zu retten. Solche Motive liegen vor, wenn der Intervent dem Zielland das Existenzrecht abspricht, dessen Territorium beansprucht oder verkündet, er wolle verhindern, dass ein rivalisierender Staat die Kontrolle über das Zielland übernimmt. Andere Ausschlussgründe sind gegeben, wenn der Intervent unmittelbar zuvor aus dem Zielland angegriffen wurde und ausschließlich sich selbst verteidigt oder wenn er die zu rettenden Bürger als Teil der eigenen Nation oder des eigenen Volks versteht.

Anhand welcher Quellen lässt sich einschätzen, ob ein Einsatz den Kriterien einer humanitären militärischen Intervention genügt? Militärische Auslandseinsätze entnehmen wir bis zum Jahr 2005 dem etablierten Datensatz International Military

7 Dieses Verständnis ist angelehnt an Holzgrefe (2003: 18) und Weiss (2007: 5–9).

Interventions (Pearson/Baumann 1993; Pickering/Kisangani 2009). Für den Zeitraum danach durchforsten wir die Internetseiten internationaler Organisationen und diverse Presseberichte. Das Vorliegen einer von Gewalt geprägten Notlage bestimmen wir mithilfe der Datensätze des Uppsala Conflict Data Program (UCDP), denen jeweils ein Schwellenwert von 25 Todesopfern im Kalenderjahr zugrunde liegt.⁸ Ob ein Einsatz militärischen Zwang ausübte oder androhte, leiten wir aus den Mandaten, Berichten und anderen Erklärungen der entsendenden Staaten oder Organisationen ab. Zudem werten wir die fallspezifische Sekundärliteratur aus. Auf die erklärte Absicht, Fremde zu retten, schließen wir aus Texten, die den Einsatz autorisieren, etwa Resolutionen des UN-Sicherheitsrats. Des Weiteren sind Aussagen höchster Entscheidungsträger kurz vor oder zu Beginn der Intervention relevant. Der deklarierten Intention, Zivilisten zu schützen, lassen sich ganz unterschiedliche Ziele zuordnen, nicht nur deren namentlich genannter Schutz oder die Lieferung humanitärer Hilfe. Ebenso wichtig sind Vorhaben, Gewalt zu stoppen oder zu verhindern sowie Frieden zu erzwingen, wiederherzustellen, zu sichern oder zu konsolidieren.

Allein mit der präsentierten Definition und Abgrenzung lässt sich die Gesamtmenge der humanitären militärischen Intervention nicht bestimmen, denn im Verlauf vieler Einsätze ändern sich die Ziele, Aktivitäten oder die ausführenden Parteien. Das wirft die Frage auf, ob mit jedem Wandel eine neue Intervention beginnt. Bejaht man dies, stiege die Zahl der Fälle drastisch, und die Evaluation gestaltete sich immer schwieriger, weil die erforderlichen Daten immer kürzeren Zeiträumen zuzuordnen wären. Besonders schwer fiel es, für die einzelnen der vielen aufeinanderfolgenden Interventionen Wirkungen zu erheben. Wir halten es daher für praktikabler, eine nachfolgende humanitäre militärische Intervention nur dann als eine neue aufzufassen, wenn sie deutlich anders ausgerichtet ist als die vorausgegangene. Das ist der Fall, wenn einer parteiübergreifenden Intervention ein unparteiischer Einsatz folgt oder umgekehrt.

Tabelle 1 listet die humanitären militärischen Interventionen von 1946 bis 2014 auf.⁹ Die Einsätze, in denen wir die Kriterien einer humanitären militärischen Intervention weniger eindeutig erfüllt sehen als in anderen, markieren wir kursiv als Grenzfälle. Dazu gehören Interventionen, deren Zwangscharakter sich nicht sicher bestimmen lässt, bei denen es strittig war, ob überhaupt eine von Gewalt geprägte Notlage bestand oder wie sehr die erklärten humanitären Absichten konterkariert wurden. Einsätze mit Beteiligung der Bundeswehr sind in der zweiten Spalte mit * versehen.

8 Eck/Hultman (2007); Gleditsch et al. (2002); Pettersson (2015); Sundberg et al. (2012). Daten unter: www.pcr.uu.se/research/ucdp/datasets/(26.4.2016).

9 Die Differenzen zur Fallliste in einem älteren HSFK-Report (Gromes/Dembinski 2013: 8–9) basieren auf dem längeren Zeitraum, den wir hier berücksichtigen und auf einer konsolidierten Datenbasis. Aus diesem Grund weichen wir auch von der Liste im Friedensgutachten 2015 ab (Dembinski/Gromes 2015: 81–82).

Tabelle 1: Humanitäre militärische Interventionen 1946–2014

Zielland	Hauptintervent	Beginn	Ende
Kongo	UN	15.07.1960	30.06.1964
<i>Pakistan</i>	<i>Indien</i>	<i>04.12.1971</i>	<i>16.12.1971</i>
<i>Libanon</i>	<i>Arabische Liga</i>	<i>21.10.1976</i>	<i>30.04.1979</i>
<i>Uganda</i>	<i>Tansania</i>	<i>21.01.1979</i>	<i>03.06.1979</i>
Libanon	USA	24.09.1982	31.03.1984
Sri Lanka	Indien	30.07.1987	20.03.1990
Liberia	Westafrikanische Wirtschaftsgemeinschaft	24.08.1990	28.12.1996
Irak (Kurden)	USA	07.04.1991	10.11.1997
Irak (Süden)	USA	27.08.1992	31.12.1996
Somalia	USA*	09.12.1992	31.03.1995
Bosnien-Herzegowina	UN*	12.04.1993	29.08.1995
<i>Ruanda</i>	<i>Frankreich</i>	<i>23.06.1994</i>	<i>17.07.1994</i>
Haiti	USA	19.09.1994	15.10.1994
Bosnien-Herzegowina	NATO*	30.08.1995	21.11.1995
Albanien	Italien	15.04.1997	11.08.1997
Sierra Leone	Nigeria	31.05.1997	24.05.1999
Jugoslawien (Kosovo)	NATO*	24.03.1999	10.06.1999
Sierra Leone	UN	08.07.1999	06.05.2000
Ost-Timor	Australien*	20.09.1999	22.11.1999
Sierra Leone	UN	06.05.2000	20.12.2001
DR Kongo	UN*	31.10.2000	28.03.2013
Burundi	UN	29.10.2001	04.12.2008
Côte d'Ivoire	Frankreich	22.09.2002	06.04.2005
DR Kongo	EU*	06.06.2003	01.09.2003
Salomonen	Australien	24.07.2003	13.08.2003
<i>Afghanistan</i>	<i>NATO*</i>	<i>11.08.2003</i>	<i>31.12.2014</i>
<i>Irak (nach Sturz Husseins)</i>	<i>USA</i>	<i>August 2003</i>	<i>18.12.2011</i>
Haiti	UN	29.02.2004	31.12.2005
<i>Libanon</i>	<i>UN*</i>	<i>15.09.2006</i>	<i>16.09.2006</i>
Somalia	Afrikanische Union	01.03.2007	dauert an
Sudan (Darfur)	UN*	31.12.2007	dauert an
Tschad	EU	28.01.2008	28.04.2010
Libyen	NATO	19.03.2011	23.10.2011
Côte d'Ivoire	UN	01.04.2011	11.04.2011
Südsudan	UN*	09.07.2011	dauert an
Mali	UN*	11.01.2013	dauert an
DR Kongo	UN*	28.03.2013	dauert an
Zentralafrikanische Republik	UN*	10.12.2013	dauert an
Irak (IS)	USA*	08.08.2014	dauert an

Die Angaben in Tabelle 1 basieren auf einer Datensammlung, die im Rahmen eines von der Deutschen Stiftung Friedensforschung geförderten Pilotprojekts erstellt wurde und neben einem tabellarischen Datensatz detaillierte Beschreibungen zu allen aufgelisteten Interventionen umfasst.¹⁰ Die Datensammlung machen wir demnächst öffentlich zugänglich.¹¹ Insgesamt zählen wir 39 humanitäre militärische Interventionen von 1946 bis 2014, davon sieben Grenzfälle.

Tabelle 1 gibt auch an, ob Deutschland mit eigenen Soldaten an der jeweiligen Intervention teilnahm.¹² Für diese Angabe verzichteten wir auf die Prüfung, wie weit das Mandat der deutschen Soldaten von dem der Gesamtmission abwich. Wurden etwa nur Soldaten für internationale Stäbe, den Sanitätsdienst oder das Transportwesen entsandt, werten wir das gleichwohl als Teilnahme an einer humanitären militärischen Intervention. Auch listen wir alle Beteiligungen unabhängig von ihrem absoluten oder relativen Umfang auf. Die Bandbreite ist dabei groß. Mandatierte die Bundesregierung für den Südsudan den Einsatz von nur bis zu 50 Soldaten, waren in Afghanistan zugleich mehrere Tausend Bundeswehrangehörige im Einsatz.

Wünscht die deutsche Politik Orientierungswissen über Erfolg und Scheitern humanitärer militärischer Interventionen, dann sollte sich die Evaluation nicht auf Fälle mit deutscher Beteiligung beschränken. Die humanitäre militärische Intervention als Ganze gehört auf den Prüfstand, und die gerade präsentierte Liste bietet die Grundlage dafür.

10 Zudem dokumentieren wir in diesem Projekt militärische Interventionen, die bisweilen als humanitär eingestuft werden, unserer Ansicht nach aber nicht den Kriterien genügen.

11 Großen Dank schulden wir unseren Hilfskräften Julia Haase, Julia Schaefermeyer, Anna Schwarz und Svenja Windisch, dem Praktikanten Raphael Haines, den Praktikantinnen Delina Goxho und Lisa Weis sowie den Teilnehmenden des Forschungspraktikums „Humanitäre militärische Interventionen“ an der Goethe-Universität Frankfurt: Hande Abay, Yasemin Altintop, Mira Ballmaier, Charlotte Brandes, Jan Dannheisig, Christian Diegelmann, Lena Diekmann, Markus Drews, Felix Haeckel, Fionn Harnischfeger, Lara Heckmann, Katharina Hemming, Kevin Horbach, Tanja Jacob, Svenja Jandrasits, Teresa Leien-decker, Lela Lindena, Vanessa Müller, Edina Pasztor, Christian Pogies, Alexander Quint, Julia Raab, Julia Schaefermeyer, Paul Scherer, Laura Schelenz, Karolina Schmid, Anja Siegel, Jens Stappenbeck, Maximilian Stoll, Aurélie Wallaschkowski und Marie Wittenius. Sie haben fast alle der im Folgenden präsentierten Daten zusammengetragen.

12 Dazu stützen wir uns auf die Angaben unter: www.einsatz.bundeswehr.de/portal/a/einsatzbw (26.4.2016).

4. Forschungsstrategische Entscheidungen und ihre Konsequenzen

Die Frage, ob militärische Interventionen evaluiert werden sollten, ist mittlerweile politisch offenbar entschieden. Nun geht es darum, wann und von wem was und wie evaluiert werden sollte. Für jede dieser Teilfragen diskutieren wir nun die Vor- und Nachteile der zu fällenden Entscheidungen. Wie sich zeigt, hängt die Wahl des Verfahrens davon ab, was mit der Evaluation erreicht werden soll. Wir unterstützen das Vorhaben, einen Einsatz sowohl begleitend zu bewerten als auch nach seinem Abschluss zu evaluieren, um mittelfristige Entwicklungen erfassen zu können. Des Weiteren sprechen wir uns für eine wirkungsorientierte Evaluation mithilfe einheitlicher Kriterien aus. Bei allen diesen Entscheidungen führen wir die Differenz zwischen der Evaluation eines einzelnen Falls und der aller humanitären militärischen Interventionen mit. Geht es um die kausale Attribution der beobachteten Veränderungen, vermag es vor allem eine Evaluation der Gesamtmenge, wichtiges Orientierungswissen zu generieren. Die detaillierte Analyse einzelner Interventionen soll die Ergebnisse dieser Gesamtevaluation absichern.

4.1 Wer sollte wann evaluieren?

Die ersten beiden Entscheidungen, die der Auftraggeber einer Evaluation von humanitären militärischen Interventionen fällen muss, betreffen die durchführende Instanz und den Zeitpunkt der Evaluation.

Bei der ersten Frage geht es darum, ob eine Evaluation von denjenigen Institutionen durchgeführt werden soll, die für den Einsatz verantwortlich zeichnen, also im Haus gemacht werden soll oder von einer nicht beteiligten, externen Institution. Der Gesetzentwurf fordert einen „ressortübergreifenden Evaluationsbericht“ (Deutscher Bundestag 2016: 3). In den Jahren zuvor hatten Anträge von Bündnis 90/Die Grünen eine externe Bilanzierung durch unabhängige Experten gefordert (Deutscher Bundestag 2011a: 3; Deutscher Bundestag 2012: 1).¹³ Das Plädoyer der Abgeordneten scheint zunächst plausibel, denn bei einer im Haus verantworteten und durchgeführten Evaluation stünden Unabhängigkeit und Neutralität infrage.

In der Praxis findet sich eine Reihe von Beispielen für gemischte Modelle. So unterhält die GIZ eine eigene Evaluationsabteilung und organisiert zumindest die Datensammlung, -analyse und -bewertung für das routinemäßige Monitoring und die Evaluierung von Entwicklungsprojekten im Haus und vergibt nur zentrale Evaluierungen an Externe.¹⁴ Diese Praxis weist darauf hin, dass es auch gute Gründe für interne Evaluierungen von humanitären militärischen Interventionen geben könnte. Geht es um die Überprüfung interner Abläufe oder die Kosteneffizienz alternativer Vorgehensweisen, sprächen zwei Überlegungen

13 Zumindest mit Blick auf den Afghanistan-Einsatz sprach sich auch Johannes Pflug von der SPD für eine unabhängige Evaluation aus (Deutscher Bundestag 2013: 26963).

14 Vgl. GIZ: Monitoring und Evaluierung, www.giz.de/fachexpertise/html/4987.html (26.4.2016).

für ein Vorgehen, bei dem intern Daten erhoben und bewertet werden, während externe Experten bei der Entwicklung und Anwendung der Methodik beratend zur Seite stehen. Erstens wären Evaluierungen im Haus schneller und kostengünstiger durchzuführen. Zweitens wäre die Erhebung und Bearbeitung von sensitiven Daten über Abläufe und Durchführung einfacher – gerade bei Auslandseinsätzen ein nicht unerheblicher Aspekt. Denn Daten, für die Geheimhaltungsvorschriften gelten, könnten internen Evaluatoren eher zugänglich gemacht werden als externen.

Für die externe Evaluation spräche zunächst, dass der Blick von außen Organisationen dabei helfen kann, Routinen und bürokratisierte Abläufe aufzubrechen und alternative Vorgehensweisen zu prüfen. Daneben verspricht eine externe Bewertung größere Unabhängigkeit. Dieser Vorteil kommt dann stärker zum Tragen, wenn die Evaluation auch auf die Frage des Erfolges eines Einsatzes eingehen soll. Denn dann müsste sie sich stärker gegen den Verdacht behaupten, dass Daten und Bewertungsmaßstäbe nach dem Ziel ausgewählt würden, einen möglichst großen Erfolg zu präsentieren. Allerdings ist der Vorteil größerer Neutralität bei externen Evaluationen nur ein relativer, kann doch der Auftraggeber nicht nur Erfolgskriterien und Art der Datenerhebung vorgeben, sondern auch mit der Auswahl der externen Evaluatoren die zu erwartenden Ergebnisse in eine Richtung lenken. Für dieses Problem gibt es keine vollständig überzeugende Lösung. Dennoch könnten Annäherungen an das Versprechen größerer Neutralität und Offenheit gelingen, etwa wenn alle Einsätze mit der gleichen Methodik und nach den gleichen Kriterien bewertet würden. Das würde verhindern, dass je nach Intervention genau die Maßstäbe angelegt werden, die sie in ein besonders günstiges Licht stellen.

Bei der zweiten Frage geht es darum, ob begleitend, direkt nach Abschluss eines Einsatzes oder erst nach einer gewissen Zeit evaluiert werden soll. Die Gesetzesvorlage spricht sich für eine Evaluation nach Abschluss eines Einsatzes aus, fordert aber ebenfalls bei Entscheidungen über Mandatsverlängerungen eine Bilanz des bis dahin Erreichten (Deutscher Bundestag 2016: 3). In den Debatten zuvor hatten sich die meisten Abgeordneten, wenn überhaupt, für eine zeitnahe Evaluation ausgesprochen. Allerdings erklärte Bijan Djir-Sarai (FDP): „Ein Evaluierungsbericht würde höchstens dann Sinn machen, wenn er mittel- und langfristig ausgerichtet wird. Der Erfolg der meisten Einsätze der Bundeswehr lässt sich nämlich erst im Lauf der Zeit erkennen“ (Deutscher Bundestag 2011b: 12065). Der große Vorteil der begleitenden Einzelfallevaluierung besteht darin, dass die Ergebnisse genutzt werden könnten, um laufende Einsätze zu justieren oder ggf. vorzeitig zu beenden. Ihr Nachteil: Sie geben in der Regel kaum Auskunft über Erfolg oder Misserfolg. Hingegen wäre die von den politischen Entscheidungsträgern ausgeblendete Option, Auslandseinsätze erst mit zeitlichem Verzug nach ihrem Abschluss zu evaluieren, am besten geeignet, um Wirkungen zu erfassen. Der Fall Libyen zeigt beispielhaft die Vorteile des retrospektiven Blicks. Hillary Clinton feierte die Libyen-Intervention direkt nach ihrem Abschluss als großen persönlichen und US-amerikanischen

Erfolg.¹⁵ Gareth Evans, ein friedenspolitischer Vordenker konfliktpräventiver Strategien und emeritierter Präsident der International Crisis Group, charakterisierte die Libyen-Intervention direkt nach ihrem Abschluss als „[...] a textbook case of the R2P [responsibility to protect] norm working exactly as it was supposed to“.¹⁶ Tatsächlich schienen beide Bewertungen Ende 2011 plausibel. Die Opfer des Krieges erschienen, verglichen mit den prognostizierten Opfern im Fall einer Nicht-Intervention, geringer, die intervenierenden Parteien hatten keine Verluste zu beklagen, und eine scheinbar reformorientierte Regierung hatte die Macht übernommen. Die tatsächlichen Kosten und fatalen Folgen der Intervention traten erst später hervor. Man denke etwa an den Staatszerfall, die Opfer des ab 2013 beginnenden Bürgerkrieges, die Ausbreitung des Islamischen Staates und die Ermöglichung der Fluchtbewegungen nach Europa und die Destabilisierung Malis. Dieses Beispiel veranschaulicht, was für alle Interventionen gilt: Einige ihrer Wirkungen werden erst längere Zeit nach ihrem Abschluss sichtbar.

Die wenigsten Vorteile weist unseres Erachtens eine Evaluation zum Zeitpunkt direkt nach Abschluss eines Auslandseinsatzes auf. Ihre Erkenntnisse könnten weder zur Feinsteuerung genutzt werden, noch versprechen sie Auskünfte über Wirkungen. Eine Evaluation zu diesem Zeitpunkt wäre offenbar in erster Linie dem Interesse geschuldet, schnell formal Bilanz zu ziehen und so ein Kapitel in der Geschichte der Auslandseinsätze abzuschließen.

4.2 Ziel- oder wirkungsorientiert evaluieren?

Eine Evaluation von Interventionen muss entscheiden, ob lediglich Zielerreichung oder auch intendierte und nicht-intendierte Wirkungen erfasst werden sollen. Der zielorientierte Ansatz geht davon aus, dass die intervenierenden Parteien intentional handeln, ihre Ziele deutlich und so formulieren, dass sich daraus Erfolgsgrößen ableiten lassen. Erfolg lässt sich dann daran messen, inwieweit und zu welchen Kosten die vom Auftraggeber festgelegten Ziele erreicht wurden (Peen Rodt 2014: 19). Die Forschung spricht auch von internen Erfolgskriterien. Im Fall humanitärer militärischer Interventionen finden sich die Ziele oftmals in den Einsatzmandaten. Dies gilt insbesondere für die Interventionen, die von einer internationalen Organisation wie der NATO oder den Vereinten Nationen verantwortet werden. Bei Interventionen in nationaler Verantwortung geben darüber hinaus die Deklarationen politischer Entscheidungsträger Aufschluss. Eine zielorientierte Evaluation rekonstruiert in einem ersten Schritt die vom Einsatzmandat gesetzten Ziele und wählt in einem zweiten Schritt Erfolgskriterien aus, die angeben, inwieweit die Ziele erreicht wurden.

15 Vgl. zur Rolle Clintons bei der Entscheidung zum Libyen-Einsatz und ihrer Erfolgsmeldung nach Abschluss der NATO Operationen die Artikelserie in der *New York Times* von Jo Becker und Scott Shane: Clinton, ‚Smart Power‘ and a Dictator’s Fall, *New York Times*, 27.2.2016, S. A1 und After Revolt, a New Libya ‚with Very Little Time Left‘, *New York Times*, 29.2.2016, S. A1.

16 Interview mit Gareth Evans: The R2P Balance Sheet after Libya, www.globalR2P.org/publications/205 (14.3.2016).

Der wirkungsorientierte Ansatz geht über eine zielorientierte Evaluation hinaus und soll soziale Eingriffe danach bewerten, was sie im Zielgebiet bewirken. Dabei spielt es keine Rolle, ob die Wirkungen in dem Auftrag benannt werden. Um die Wirkungen von Interventionen sichtbar zu machen, erhebt dieser Ansatz in einem ersten Schritt Veränderungen im Zielland, die sich als direkte und indirekte, intendierte und nicht-intendierte Effekte des Eingriffs denken lassen. In einem zweiten Schritt versucht er, die beobachteten Veränderungen kausal der zu evaluierenden Intervention zuzuordnen (Stockmann/Meyer 2014: 79). Bei der Bestimmung dessen, was als Erfolg oder Misserfolg zu werten ist, kann sich eine wirkungsorientierte Evaluation zwar von den Zielsetzungen der Auftraggeber leiten lassen, bestimmt aber Erfolg und Erfolgskriterien letztlich selbst. Die Forschung spricht deshalb von externen Erfolgskriterien.

In der Praxis erweist sich diese Unterscheidung der Ansätze oftmals als weniger trennscharf, als es auf den ersten Blick scheinen mag. In der Evaluationsliteratur sind denn auch unterschiedliche Begrifflichkeiten gängig. So unterscheidet etwa eine einflussreiche OECD-Richtlinie (2012: 13) zur Evaluation von Programmen der Friedenskonsolidierung zwischen

- *Outputs*: den Produkten oder Dienstleistungen, die im Rahmen einer konfliktpräventiven Maßnahme erbracht werden, wie etwa die Anzahl der gebohrten Brunnen,
- *Outcomes*: den unmittelbaren Effekten dieser *Outputs* und
- *Impacts*: den positiven oder negativen, primären oder sekundären Effekten des Eingriffs, die direkt oder indirekt intendiert oder nicht-intendiert sein können.

Eine deutliche Unterscheidung zwischen ziel- und wirkungsorientiertem Evaluieren bliebe am ehesten erhalten, wenn sich die in den Einsatzmandaten formulierten Ziele auf *Outputs* beziehen. In dem Maße, in dem sich die Mandate auf *Outcomes* oder sogar *Impacts* beziehen, verwischen die Unterschiede.

Ein Blick auf die Praxis humanitärer militärischer Interventionen mag dies verdeutlichen. Die Mandate einiger Missionen beziehen sich auf Aktivitäten der Truppe. Beispielsweise beauftragte das Mandat für die African Union Mission in Sudan (AMIS) die Truppen, einen Waffenstillstand in Darfur zu beobachten und zu verifizieren. Die tatsächliche Einhaltung des Waffenstillstands durch die Konfliktparteien kann also kein Erfolgskriterium sein, sondern nur die Beobachtung der Gewaltentwicklung durch die AMIS. Eine zielorientierte Evaluation könnte versuchen, die Anzahl der von AMIS gemachten Beobachtungen zu erfassen, in ein Verhältnis zur Größe der Truppe und des zu überwachenden Geländes zu setzen und dann auf der Grundlage von vorliegenden Erfahrungswerten, die Effektivität der Truppe und die Kosteneffizienz alternativer Beobachtungsmethoden zu bewerten. Erfolg könnte sie daran messen, ob es der Truppe gelungen ist, die Waffenstillstandsverletzungen zu erfassen.

In anderen Fällen beziehen sich die Mandate auf Wirkungen. Oftmals ist davon die Rede, die entsandte Truppe solle den Frieden wiederherstellen und Sicherheit gewähren. Zum Beispiel nennt das Mandat für die African Union Mission to Somalia (AMISOM) als

Ziel „to reduce the threat posed by Al Shabaab and other opposition armed groups“.¹⁷ Um die Erreichung dieses Zieles zu überprüfen, müsste eine Evaluation erfassen, ob AMISOM ein verändertes Bedrohungsniveau bewirkt hat.

Die Unterscheidung zwischen ziel- und wirkungsorientiertem Evaluieren ist also idealtypisch und verschwindet in dem Maße, in dem sich die Ziele einer Intervention nicht auf *Outputs*, sondern auf Wirkungen beziehen. In diesen Fällen unterscheiden sich beide Typen nur noch in der Frage, ob auch indirekte und nicht-intendierte Effekte einer Intervention erfasst werden sollen.

Letztlich hängt die Entscheidung zwischen einer zielorientierten und auf *Outputs* konzentrierten oder einer wirkungsorientierten Evaluation davon ab, was man wissen möchte. In dieser Hinsicht weisen beide Typen Vor- und Nachteile auf. Für ein zielorientiertes Vorgehen spricht, dass die Evaluation nur das erfasst, was die Intervention explizit erreichen sollte. Das entlastet eine Bilanzierung von extern definierten Zielen, die die Intervention gar nicht anstreben wollte. Ein zweiter Vorteil besteht in der einfacheren Durchführbarkeit. Die evaluierende Instanz muss selbst weniger Entscheidungen darüber treffen, welche Ziele die Auftraggeber verfolgen, was Erfolg bedeutet und welche Erfolgskriterien zu wählen sind. Wenn sich die Ziele auf *Outputs* beziehen, kann die Evaluation zudem die schwierige Frage der Attribution umgehen. Diese Argumente überzeugten offenbar viele Bundestagsabgeordnete, und auch in der Literatur wird der Erfolg militärischer Interventionen oft als Mandatserfüllung verstanden (Diehl 1994).

Diesen Vorteilen steht allerdings eine Reihe von Problemen und Nachteilen gegenüber. Erstens hängt bei diesem Vorgehen der Erfolg von der Zieldefinition ab. Formuliert ein Auftraggeber bescheidene Ziele, ist der Erfolg fast schon garantiert. Formuliert er ehrgeizige Ziele, ist es wahrscheinlicher, dass die Evaluatoren einen Misserfolg des Programms feststellen. Kurzum könnten Programme mit ähnlicher Performanz je nach Zieldefinition als Erfolg oder Misserfolg erscheinen (Peen Rodt 2014: 20). Insbesondere bei einem Vergleich mehrerer Programme, deren Ansprüche unterschiedlich weit reichen, erweist sich die Zieldefinition als die Ergebnisse dramatisch verzerrende Fehlerquelle. Die Relevanz dieses Problems wird mit einem Blick etwa auf die militärischen Missionen der Europäischen Union deutlich. Die EU entschied sich bewusst für Einsätze, bei denen das Risiko des Scheiterns gering erschien, und setzte darüber hinaus die Ziele so niedrig an, dass der Erfolg praktisch garantiert war. So erklärte ein hochrangiger Vertreter der Europäischen Union: „The EU will not launch an operation if it is not sure that it will succeed“ (Peen Rodt 2014: 24). Dagegen sollten die Einsätze anderer Staaten mit bescheidensten Mitteln hochfliegende Ziele erreichen. Beispielsweise beteiligten sich die USA 1982 mit begrenzten Truppenkontingenten an einer Multinational Force II genannten Friedensmission während des libanesischen Bürgerkrieges und verbanden damit offiziell das Ziel, die Souveränität und Autorität der libanesischen Regierung über

17 Das Mandat ist abrufbar unter: <http://amisom-au.org/amisom-mandate/> (26.4.2016).

Beirut wiederherzustellen.¹⁸ War dies bereits angesichts der Fragilität der neugebildeten Regierung, der Stärke und Gegensätze der nicht-staatlichen Kampfverbände sowie des Ausmaßes der syrischen Einmischung äußerst ambitioniert, gingen die inoffiziellen Ziele der USA noch viel weiter. Denn die Intervention sollte Impulse für eine umfassende Reform der libanesischen Armee geben, den syrischen Einfluss zurückdrängen und den Charakter Libanons als demokratisches und pro-westliches Land stärken (Malone et al. 1986).

Zweitens sind die selbst gesetzten Ziele oftmals unklar, widersprüchlich und unvollständig formuliert. Hinzu kommt, dass mehrere Beteiligte einer Intervention unterschiedliche Ziele verfolgen können oder sich die Ziele im Verlauf des Eingriffs häufig ändern können. Auch hierfür ist die bereits erwähnte Multinational Force II, an der sich neben den USA auch Italien, Frankreich und Großbritannien beteiligten, ein gutes Beispiel. Einem offiziell genannten Ziel zufolge sollten die Truppen neutral bleiben und ein sicheres Umfeld schaffen.¹⁹ Italien, der größte Truppensteller, orientierte sich bis zum Ende der Mission an diesem Teil des Auftrags. Die italienischen Truppen beschränkten sich auf Maßnahmen der Friedenserhaltung und gewährten so in ihrem Sektor erfolgreich Sicherheit für die Bevölkerung Beiruts. Dagegen verfolgten die USA und auch Frankreich viel weitergehende Ziele, manövierten sich dadurch in eine zunehmend anti-syrische Position und verstrickten sich in Kämpfe, die schließlich zum Abbruch der Mission führten.

Drittens können Auftraggeber neben den deklarierten Zielen noch Ziele anderer Art verfolgen. Die französische Operation Turquoise in Ruanda 1994 mag dies illustrieren. Die Truppen wurden im Juni des Jahres entsandt, als der Genozid längst begonnen hatte. Der Sicherheitsrat hatte die Intervention in Resolution 929 (1994) mandatiert und Frankreich zum Einsatz aller notwendigen Mittel autorisiert, um die humanitären Ziele zu erreichen. Das Mandat verpflichtete die französischen Truppen zur Neutralität und beauftragte sie, auf unparteiische Weise zu Sicherheit und Schutz von Vertriebenen, Flüchtlingen und gefährdeten Zivilisten beizutragen. Frankreich bestätigte in Erklärungen diese Zielsetzung und richtete zu diesem Zweck eine Sicherheitszone ein, die es auch gegen die heranrückende Tutsi-Streitmacht der Rwandan Patriotic Front (RPF) verteidigte. Allerdings wurden durch diese Maßnahme nicht nur Zivilisten geschützt, sondern auch den für den Genozid verantwortlichen Kämpfern und Mitgliedern der ruandischen Übergangsregierung ein Fluchtweg nach Zaire eröffnet. Angesichts der langen und engen Beziehungen Frankreichs zu den Hutu-dominierten Regierungen in Ruanda diskutieren Frankreich und die Welt seitdem, ob die „eigentlichen“ Ziele der Operation Turquoise darin bestanden, unschuldige Zivilisten oder die Verantwortlichen des Genozids zu retten. Weitere Beispiele drängen sich in diesem Zusammenhang auf. Ging es etwa in Libyen nur um den Schutz von Zivilisten, wie das UN-Mandat vorschrieb? Oder ging es

18 Exchange of Letters between the United States and Lebanon, September 25, 1982, abgedruckt in Moore (1991: 1191–1193).

19 Vereinbart wurde, „to further efforts of the Government of Lebanon to assure the safety of persons in the area and bring to an end the violence which has tragically recurred“ (Moore 1991: 1191–1193).

bald nach Beginn der Operationen um einen Regimewechsel, wie es das militärische Vorgehen der intervenierenden Staaten nahelegt? Zusammenfassend drohen die scheinbaren Vorteile der klaren Erfolgskriterien und einfacheren Durchführbarkeit zielorientierter Evaluationen also in der realen Welt schnell zu verschwinden.

Der Vorteil der Wirkungsanalyse liegt in dem Versprechen, genau die Informationen zu liefern, an denen die Auftraggeber interessiert sein sollten. Sicherlich interessiert bei humanitären militärischen Interventionen auch, ob und zu welchen Kosten die Truppe die in den Mandaten festgelegten Ziele erreicht hat. Für eine politische Bewertung unverzichtbar ist aber die Erfassung ihrer direkten und indirekten, intendierten und nicht-intendierten Wirkungen. Ein Beispiel soll dies deutlich machen: Eine Bewertung der Libyen-Mission, die lediglich fragt, ob die deklarierten Ziele des Flugverbots und des Schutzes von Zivilisten in unmittelbarer Gefahr erreicht wurden, könnte zu einem positiven Ergebnis kommen (Chivvis 2014: 174).²⁰ Dagegen macht der Blick auf weitere Entwicklungen, wenn sie sich denn der Intervention kausal zuordnen ließen, das dramatische Ausmaß des Misserfolges deutlich.

Der Nachteil einer Wirkungsanalyse ist, dass sie sehr aufwendig ist. Die evaluierende Instanz muss mehr methodische Entscheidungen treffen, die das Ergebnis beeinflussen. Dies gilt bereits für den ersten Schritt einer Wirkungsanalyse – die Erhebung von Veränderungen im Zielland. Eine Evaluation müsste die Zeitpunkte der Messungen festlegen, also die sogenannte *baseline* und den Zeitpunkt, für den Veränderungen gemessen werden. Ferner müsste sie das Zielgebiet definieren und die zu vermessenden Parameter, Skalierungen und Messmethoden festlegen. Noch aufwendiger ist der zweite und dritte Schritt – die Attribution und die politische Einordnung der Befunde. Dieses Vorhaben wirft so große Schwierigkeiten auf, dass sich viele Evaluationen ihnen erst gar nicht stellen. „[T]here is virtually no reference at all to the *attribution gap*“, heißt es in einer Meta-Evaluation zum internationalen *state-building* (Grävingholt et al. 2012: 14). Wie man mit diesem Problem umgehen kann, diskutieren wir in den Kapiteln 4.3 und 4.4.

Zusammengefasst spielt eine zielorientierte Evaluation ihre Stärken dann aus, wenn in den Einsatzmandaten Ziele präzise definiert sind und sich aus diesen möglichst direkt aussagekräftige und gut messbare *Outputs* ableiten lassen. Dann kann eine Evaluation mit überschaubarem Aufwand und geringer Subjektivität über einige Aspekte von Erfolg oder Misserfolg Auskunft geben. Eine wirkungsorientierte Evaluation ist deutlich aufwendiger, muss mehr Entscheidungen treffen und ist daher subjektiver. Ihr Vorteil: Sie kann mehr Informationen über die politisch interessante Frage erbringen, ob eine Intervention im Sinne ihres Kernanliegens Menschen gerettet hat.

20 Die Resolution 1973 des Sicherheitsrats (2011) formuliert als relevante Ziele „to establish a ban on all flights“ und „to protect civilians and civilian populated areas under threat of attack“. Eine zielorientierte Evaluation würde Luftangriffe erfassen und prüfen, ob es (a) zu Verstößen gegen das Flugverbot kam und (b) wie viele Zivilisten durch NATO-Luftangriffe ums Leben kamen und wie viele gerettet wurden. Mit der letzten Frage wäre diese Evaluation mit dem Problem der Erfassung von Wirkungen konfrontiert.

4.3 Woran Erfolg und Scheitern messen?

Die wirkungsorientierte Evaluation muss damit beginnen, Kriterien für Erfolg und Misserfolg der Intervention festzulegen, mit deren Hilfe die Entwicklungen im Zielland bewertet werden. Bei der Bundestagsdebatte zum Afghanistan-Einsatz im Dezember 2014 sprachen Redner der Regierungsparteien direkt oder indirekt eine breite Palette möglicher Kriterien für Erfolg an: Sie gedachten der ums Leben gekommenen deutschen Soldaten und Zivilkräfte, beklagten große Armut und Korruption sowie verbreitete Gewalt gegen Frauen in Afghanistan. Zugleich führten sie an, die Sicherheitslage habe sich wesentlich verbessert, man solle den ersten demokratischen Regierungswechsel in der Geschichte Afghanistans anerkennen, es gebe eine lebendige Zivilgesellschaft und höhere Einschulungsraten, und mehr Menschen als zuvor hätten Zugang zu Wasser, Strom und ärztlicher Versorgung (Deutscher Bundestag 2014: 7270–7280). Möglicherweise wollten die Redner der Regierungsparteien mit der langen Liste an Kriterien dem in der Öffentlichkeit verbreiteten Eindruck eines gescheiterten Einsatzes entgegentreten, indem sie Teilerfolge hervorhoben. Umgekehrt legen Kritiker militärischer Einsätze den Schwerpunkt auf solche Bewertungsmaßstäbe, die einen als erfolgreich wahrgenommenen Einsatz entzaubern. Der Intervention im Kosovo halten sie beispielsweise vor, sie habe den Grundkonflikt nicht gelöst, und Kosovo leide unter organisierter Kriminalität, einer überforderten Justiz sowie einer desolaten wirtschaftlichen Lage (Jöst 2009: 125–126). Hier bestimmt der politische Grundsatzstreit die Auswahl der Kriterien.

Daneben gibt es eine Reihe sachlicher Gesichtspunkte, die für oder gegen diese oder jene Kriterien zur Erfassung der Wirkungen sprechen. Bei der Evaluation eines Einzelfalles kann es sinnvoll sein, eine Vielzahl fallspezifischer Kriterien in den Blick zu nehmen, um eine möglichst umfassende Bilanz zu erstellen. Für eine wirkungsorientierte Evaluation aller humanitären militärischen Interventionen eignet sich ein umfangreicher und aus den Besonderheiten des Einzelfalles heraus generierter Katalog von Kriterien hingegen nicht. Veränderungen zu vielen Gesichtspunkten zu erheben und dann noch Ursachen zuzuordnen, wäre ein sehr aufwendiges Unternehmen, wenn es denn nicht schon daran scheitern würde, dass zu vielen Fällen die erforderlichen Daten fehlen. Eine wirkungsorientierte fallübergreifende Evaluation verlangt ein schlankes und einfach zu handhabendes Instrumentarium. Dabei sollten sich Kriterien auf das Kernanliegen humanitärer militärischer Interventionen beziehen, nämlich auf den Schutz der Bürger des Ziellands vor einer von Gewalt geprägten Notlage. Auszublenden sind weitere Beweggründe, die beim Beschluss über den Einsatz eine Rolle gespielt haben mögen, etwa das Demonstrieren von Bündnissolidarität. Diese Konzentration auf die Kernanliegen hält die Zahl der Kriterien gering und sichert so die Verfüg- und Vergleichbarkeit der Daten. Die erforderlichen Daten müssen für möglichst viele Interventionen vorliegen, valide und möglichst leicht zugänglich sein. Wichtig dabei ist, dass sich die Kriterien mithilfe einfacher, aussagekräftiger Indikatoren abbilden lassen. Aus Gründen, auf die wir im nächsten Kapitel eingehen, sollten sie auch für die Fälle mit einer von Gewalt geprägten Notlage vorliegen, in denen es keine humanitäre militärische Intervention gab. Schließlich sollten die Kriterien und Indikatoren solche Veränderungen erfassen, die sich zeitlich und räumlich nicht allzu weit von der Intervention entfernen. Dies spricht etwa dagegen, das militärische Eingreifen an der

späteren Demokratisierung im Zielland zu bewerten, wie es Pearson et al. (2013: 131–132) fordern.

Für eine wirkungsorientierte Evaluation der humanitären militärischen Interventionen schlagen wir vier Kriterien vor: 1) Das Ende oder Fortdauern der Notlage, 2) die Entwicklung der tödlichen Gewalt, 3) ein Rückfall in die Gewalt, 4) eine Destabilisierung von Nachbarländern.

Endet die von Gewalt geprägte Notlage?

Das erste Kriterium fragt, ob und wann die von Gewalt geprägte Notlage während der humanitären militärischen Intervention aufhörte. Anfangs- und Enddatum einer Intervention sind dafür mit der Dauer der jeweiligen Notlage abzugleichen. Von Gewalt geprägte Notlagen lassen sich den umfangreichen und gut gepflegten Datensätzen des Uppsala Conflict Data Program (UCDP) entnehmen. Sie erfassen unterschiedliche Arten von Gewaltereignissen, durch die in einem Kalenderjahr mindestens 25 Menschen ums Leben kommen:²¹ „Bewaffnete Konflikte“ finden zwischen Staaten statt oder zwischen der Regierung eines Landes und einer Rebellengruppe (Gleditsch et al. 2002; Pettersson 2015).²² „Nicht-staatliche Konflikte“ bezeichnen militärische Auseinandersetzungen zwischen nicht-staatlichen Gewaltakteuren ohne Beteiligung von Streitkräften einer Regierung (Sundberg et al. 2012).²³ „Einseitige Gewalt“ meint tödliche Übergriffe auf Zivilisten durch Regierungskräfte oder andere organisierte Gruppen (Eck/Hultman 2007).²⁴

Die UCDP-Datensätze dokumentieren nahezu alle Gewaltkonflikte seit 1989.²⁵ Als Nachteil könnte man dem UCDP auslegen, dass die Opferangaben für fast alle Konflikte deutlich unter den Zahlen der fallspezifischen Literatur liegen.²⁶ Auch beschränkt es sich auf Opfer direkter Gewalt, weshalb zum Beispiel konfliktbedingte Hungertote nicht in die Statistik eingehen. Die absoluten Zahlen, die das UCDP zu einem einzelnen Fall vorlegt, sind daher skeptisch zu betrachten. Geht man aber davon aus, dass das UCDP über die verschiedenen Phasen eines Konflikts hinweg und auch fallübergreifend gleichermaßen sehr zurückhaltende Angaben macht, dann ermöglichen seine Daten eine vergleichende Bilanz zum Auf und Ab tödlicher Gewalt.

Die Evaluation muss erheben, wie viel Zeit vom Beginn der Intervention bis zum Ende der Notlage vergeht. Welche Zeitspanne als kurz, noch ausreichend oder viel zu lang betrachtet wird, um von einem Erfolg zu sprechen, lässt sich nur willkürlich festlegen. Ginge

21 Siehe die Definitionen zu den folgenden Gewaltformen auf: <http://bit.ly/1GOrx04> (26.4.2016).

22 Daten unter: <http://bit.ly/1iKUfNI> (26.4.2016) und <http://bit.ly/1QiI6Dq> (26.4.2016).

23 Daten unter: www.pcr.uu.se/research/ucdp/datasets/ucdp_non-state_conflict_dataset/ (26.4.2016).

24 www.pcr.uu.se/research/ucdp/datasets/ucdp_one-sided_violence_dataset/ (26.4.2016).

25 Unberücksichtigt bleiben die Notlagen mit einer humanitären militärischen Intervention in Haiti (1994), Albanien und auf den Salomonen.

26 Zum Beispiel übertreffen die Angaben der Fallliteratur zu den zwischen 1990 und 2009 beendeten Bürgerkriegen die UCDP-Zahlen durchschnittlich um das Siebenfache.

es um die Evaluation eines Einzelfalls, könnte man fallspezifische Merkmale und Erwartungen der Öffentlichkeit oder der intervenierenden Parteien in Rechnung stellen. Beispielsweise wähte sich die Bundesregierung nach wenigen Wochen unter großem Druck, den Kosovo-Krieg zu beenden, sah sich aber in der Lage, den für die eigenen Soldaten viel gefährlicheren Afghanistan-Einsatz über Jahre hinweg fortzusetzen. Da eine Evaluation aller humanitären militärischen Interventionen einen einheitlichen Maßstab verlangt, ist eine entsprechende Festlegung von außen erforderlich. Eine unrealistisch kurze Zeitspanne von wenigen Tagen verbietet sich ebenso wie eine Frist von mehreren Jahren. In früheren Publikationen haben wir einen Schwellenwert von einem halben oder ganzen Jahr vorgeschlagen (Gromes/Dembinski 2013: 14–19; Dembinski/ Gromes 2015: 80–82).

Geht die tödliche Gewalt zurück?

Bei einer rasch beendeten Notlage kann dennoch von Erfolg keine Rede sein, wenn die Gewalt während der Intervention viel schlimmer war als vor dem Eingreifen. Umgekehrt lässt sich trotz fortdauernder Notlage eine Intervention als Erfolg einstufen, falls die Gewalt deutlich zurückging. Daher schlagen wir vor, das Kriterium der beendeten Notlage um das der Entwicklung der tödlichen Gewalt zu ergänzen.²⁷ Zur Evaluierung der Gesamtzahl humanitärer militärischer Interventionen ist dafür wieder ein einheitlicher Schwellenwert sinnvoll. Um wie viel Prozent die tödliche Gewalt je Zeiteinheit sinken oder steigen muss, um von Erfolg oder Scheitern zu sprechen, ist eine weitere willkürliche Setzung. Uns scheint es ratsam, nur größere Ausschläge nach unten oder oben zu berücksichtigen und so eine Fehlertoleranz einzubauen. In einem Beitrag für das Friedensgutachten (Dembinski/ Gromes 2015: 80–82) schlagen wir vor, sich auf Veränderungen um mindestens 30 Prozent zu beschränken.²⁸

Ein Problem dieses Indikators ergibt sich daraus, dass zu einigen Fällen die Zahlen zu Todesopfern nur für das ganze Jahr, nicht aber für kleinere Zeitabschnitte vorliegen. Für das Jahr des Interventionsbeginns lässt sich daher nicht präzise angeben, wie viele Menschen in den Tagen dieses Jahres vor und während des Eingreifens ums Leben kamen. Da das UCDP aber zu immer mehr Konflikten Gewaltereignisse mit genauerem Datum

27 Datensätze wie von UCDP müssen sich die Kritik gefallen lassen, dass sie nur die direkten Todesopfer von Kämpfen und Übergriffen erfassen, nicht aber all die Menschen, die durch indirekte Folgen wie Hunger oder Kollaps des Gesundheitssystems sterben. Auf Haushalten basierte Erhebungen versuchen daher, sowohl die direkten als auch die indirekten Opfer zu ermitteln, indem sie die Sterblichkeitsrate zu verschiedenen Zeitpunkten erheben. Solche Studien zur Demokratischen Republik Kongo (Coghlan et al. 2008) und zum Irak (Burnham et al. 2006) haben horrende Opferzahlen präsentiert, während die Befunde zum Kosovo (Spiegel/ Salama 2000) gut zu anderen Angaben zu diesem Fall passen (siehe zur Kritik der Sterblichkeitsstudien Spagat 2010). Haushaltsbasierte Sterblichkeitsraten fallen aber als Kriterium für den Erfolg humanitärer militärischer Interventionen schon deshalb aus, weil sie nur für einen kleinen Teil der Fälle vorliegen.

28 Als weiteres Kriterium scheinen sich die Entwicklungen der Flüchtlings- und Vertriebenenzahlen anzubieten. Sinkende oder stagnierende Zahlen können aber in die Irre führen, wenn zuvor Vertreibungen und Fluchtbewegungen so umfangreich waren, dass es kaum noch Menschen gibt, die ein Gebiet verlassen könnten. Für Daten zu Flüchtlingen und Vertriebenen siehe <http://bit.ly/1VPCfK6> (26.4.2016); <http://bit.ly/1sUhASA> (26.4.2016); <http://bit.ly/1gZcApE> (26.4.2016).

aufführt (Croicu/Sundberg 2015; Sundberg/Melander 2013)²⁹, wird die Bedeutung dieser möglichen Fehlerquelle abnehmen. Die zu erwartenden kleinteiligeren und präziseren Daten werden aber nichts an einem Problem ändern, das sich in Fällen mit einer langen und wechselvollen Geschichte von Gewalt stellt. Hier fällt es schwer, den Zeitraum vor der Intervention festzulegen, der mit der Zeit während des Eingreifens verglichen wird. Beginnt etwa der Vergleichszeitraum für die Intervention in Somalia mit dem bewaffneten Konflikt in der ersten Hälfte der 1980er Jahre? Oder mit dessen Eskalation 1988? Oder doch erst mit dem Sturz von Siad Barre Anfang 1991? Je nach Wahl stellen sich Ausmaß und Intensität der tödlichen Gewalt vor der Intervention anders dar, und je nach Wahl ergibt sich eine andere Bilanz für die relative Gewaltentwicklung während des Eingreifens. Erscheinen mehrere Optionen für die Wahl der Ausgangslage gleich sinnvoll, ist es geboten, die Todesrate während der Intervention mit mehreren Zeiträumen vor dem Eingreifen zu vergleichen.

Rückfall in die Gewalt

Sollte eine Intervention mit einem raschen Ende der Notlage oder einem deutlichen Rückgang der tödlichen Gewalt einhergehen, könnten Kritiker einwenden, die Gewalt sei nur kurzfristig unterbrochen und flamme später möglicherweise umso heftiger wieder auf (vgl. Jöst/Strutynski 2009: 14). Als drittes Kriterium für Erfolg oder Scheitern bietet es sich daher an, zu untersuchen, ob es innerhalb von fünf Jahren nach Ende der Intervention zu einer neuen von Gewalt geprägten Notlage kam. Ein Beobachtungszeitraum von fünf Jahren scheint willkürlich bestimmt, folgt aber der gängigen Praxis in der Forschung zum Wiederausbruch von Bürgerkriegen. Ein längerer Zeitraum verzögert für aktuelle Fälle eine Bilanzierung mit diesem Kriterium; auch können immer mehr neu auftretende Gewaltursachen ins Spiel kommen und die Attribution erschweren.

Es liegt nahe, das Auftreten einer neuen Notlage mithilfe der Daten vom UCDP zu bestimmen. Gab es einen neuen bewaffneten Konflikt oder nicht-staatlichen Konflikt oder neue einseitige Gewalt, so ist der Blutzoll dieser Ereignisse mit der beendeten Notlage zu vergleichen. Denn nach einem Krieg mit 50.000 Toten mutet ein neuer bewaffneter Konflikt mit 50 Todesopfern kaum wie ein Rückfall an, der den Misserfolg einer humanitären militärischen Intervention markiert.

Gewaltsame Notlagen in Nachbarländern?

Wie geschildert, verschärften die Folgen der Intervention in Libyen die Lage in Mali. Diese Erfahrungen regen an, solche destabilisierenden Nebeneffekte zu erfassen. Zu diesem Zweck bietet es sich an, systematisch Notlagen in Staaten zu erheben, die direkt an das Zielland der Intervention angrenzen. Relevant dafür ist die Zeit von Beginn der Intervention bis einige wenige Jahre nach ihrem Abschluss. Um die Vergleichbarkeit zu gewährleisten, ist die Zahl der dort zu beobachteten Notlagen ins Verhältnis zur Zahl der

29 Daten unter: www.ucdp.uu.se/ged/ (26.4.2016).

Nachbarländer zu setzen. Wie Notlagen auch ohne Intervention andauern, noch weiter eskalieren oder nach einigen Jahren wieder ausbrechen können, so sind destabilisierende Effekte von Notlagen auf Nachbarländer denkbar, in denen keine Intervention stattfindet. Man denke hier etwa an den Zufluss von Waffen oder veränderte demografische Gleichgewichte nach Flucht und Vertreibung (Dixon 2009: 712; Myrick et al. 2015: 5–6, 17–19). Ein relativer Nachteil dieses vierten Kriteriums gegenüber den ersten drei besteht darin, dass es sich zeitlich und räumlich von der Intervention entfernt und folglich von einer Vielzahl Faktoren beeinflusst werden kann.

4.4 Wie lässt sich die Wirkung von Interventionen ermitteln?

Dieses Kapitel diskutiert, inwiefern und auf welche Weise mehr Wissen über die politisch zentrale Frage nach den Wirkungen militärischer Interventionen generiert werden kann. Dabei stellt sich die Frage, ob es um die Wirkungen einer einzelnen Intervention oder die der Gesamtklasse gehen soll. Beide Varianten der Evaluation benutzen unterschiedliche Methoden und suchen Antworten auf unterschiedliche Fragen. Die Einzelevaluation soll für den untersuchten Fall zeigen, wie die Intervention die beobachtete Veränderung bewirkt hat. Die Analyse der gesamten Klasse humanitärer militärischer Interventionen soll hingegen die durchschnittlichen Effekte dieser Einsätze ermitteln und die Bedingungen von Erfolg und Scheitern herausarbeiten.

Im Folgenden stellen wir sowohl die Evaluation eines Einzelfalls als auch die aller humanitären militärischen Interventionen vor. Da alle Varianten spezifische Stärken und Schwächen aufweisen, empfehlen wir eine Kombination der verschiedenen Ansätze.

Zur Evaluation einer einzelnen Intervention

Ausgangspunkt der Einzelfallevaluation ist die Erhebung von relevanten Daten zur Ausgangslage im Zielland (*baseline*), zu Mandat und Aktivitäten der Intervention und zu den Veränderungen im Zielland. Welche Beobachtungspunkte sich am besten dafür eignen, mögliche Wirkungen von Interventionen zu erheben, haben wir im vorausgegangenen Kapitel diskutiert.

Um nachzuweisen, dass die Intervention die beobachteten Veränderungen verursacht hat, malt ein erster Ansatz für den einzelnen Fall aus, wie sich die Situation ohne die Intervention entwickelt hätte, und vergleicht dieses Bild mit dem tatsächlichen Geschehen. Beispielhaft zeigt sich das bei Taylor Seybolt, der die Todesrate während einer Intervention mit der Sterblichkeitsrate im Fall einer ausbleibenden Intervention vergleicht (Seybolt 2007: 31–36). Die Probleme mit solchen kontrafaktischen Argumentationen auf Basis eines einzelnen Falls wollen wir am Beispiel von zwei Aufsätzen zu der Intervention in Libyen demonstrieren, die kurz hintereinander in der renommierten Zeitschrift *International Security* erschienen sind.

Robert A. Pape (2012: 61–63) zufolge gab es vor dem Eingreifen ein großes Ausmaß von Gewalt. Gaddafis Worte, man werde von Haus zu Haus gehen, um Libyen von Ratten und Ungeziefer zu säubern, interpretiert Pape, ganz auf Linie der Interventen, als

Ankündigung eines Massakers. Die Intervention habe dieses Massaker verhindert. Hätte sie nicht stattgefunden, wären Gaddafis Repression wahrscheinlich Zehntausende Menschen zum Opfer gefallen. Alan J. Kuperman (2013: 108–123) bestreitet die zitierte Aussage Gaddafis keineswegs, hält jedoch ein anstehendes Massaker für unwahrscheinlich, da die Regierungstruppen nach Rückeroberungen im vorherigen Kriegsverlauf zurückhaltend agiert hätten. Ohne die Intervention hätte die Regierung den Krieg rasch gewonnen, und rund 1.000 Menschen wären ums Leben gekommen – deutlich weniger als die bis zu 11.500 Menschen, die durch den Krieg starben, der durch die Intervention um 30 Wochen verlängert wurde.³⁰

Wie das Beispiel zeigt, lassen sich zum selben Fall kontrafaktische Analysen mit gegenläufigen Ergebnissen konstruieren. Wir halten daher die Befunde auf Basis eines solchen Vorgehens für wenig belastbar. Zwar gibt es Gütekriterien für solche kontrafaktischen Überlegungen: Sie sollen klar und konkret sein und frei von inneren Widersprüchen, möglichst wenig an den realen Ereignissen ändern, nur begrenzt in Raum und Zeit ausgreifen und sich dabei auf etablierte, empirisch abgesicherte Theorien oder Zusammenhänge stützen (Levy 2008: 632–640; Tetlock/Belkin 1996: 16–31). Doch auch wer all das beherzigt, findet leicht Anhaltspunkte, die seine vorgefasste Ansicht stützen und so erlauben, ein stimmiges kontrafaktisches Narrativ für einen Einzelfall zu verfassen. „The fact that even experts are susceptible to crippling psychologies makes it difficult or even impossible to distinguish sound from unsound counterfactuals“, beklagt Ian S. Lustick (2010: 438–439, angelehnt an Philip E. Tetlock 2006: 146, 151).

Das vielversprechendste Verfahren, um im einzelnen Fall der Wirkung einer Intervention nachzugehen, ist das sogenannte *process tracing*. Es wird in den Sozialwissenschaften breit eingesetzt und auch als Instrument der wirkungsorientierten Evaluation empfohlen (Befani/Mayne 2014; Punton/Welle 2015). Angenommen, auf eine Intervention folgte ein schnelles Ende der Gewalt. Das *process tracing* soll dann den etwaigen Kausalmechanismus untersuchen, der die Intervention über möglichst konkret formulierte und gut überprüfbare Zwischenschritte mit der beobachteten Entwicklung verbindet (Beach/Pedersen 2013; Bennett/Checkel 2015a).³¹ Gelingt dies, spricht das für einen Beitrag der Intervention zur ermittelten Entwicklung. Die Stärke dieses Beitrags bleibt allerdings unbestimmt, solange rivalisierende oder ergänzende Erklärungen nicht entkräftet sind. In der Regel kann *process tracing* nur einen Teil dieser anderen Erklärungen zurückweisen (Bennett 2008: 712–713), weshalb sich die Wirkung einer humanitären militärischen Intervention kaum genau identifizieren lässt. Das skizzierte Vorgehen läuft damit mehr

30 Eine kürzere Version des Aufsatzes (Kuperman 2015) inspirierte sogar Redebeiträge im UN-Sicherheitsrat (UN Security Council 2015: 33).

31 Die erwähnte OECD-Richtlinie zur Evaluation (2012: 58–68, 80–86) empfiehlt, dass eine sogenannte *theory of change* die Analyse strukturieren soll. Eine solche Theorie kann dem *process tracing* eine zu überprüfende Annahme über den Kausalmechanismus liefern, der die Intervention mit der ermittelten Veränderung verbindet.

auf eine Abschätzung des Beitrags (*contribution*) als auf eine Attribution der beobachteten Entwicklung hinaus.

Die Methode weist den zusätzlichen Nachteil auf, dass sie einen erheblichen Aufwand verlangt (Schimmelfennig 2015: 101), wenn man die vorgeschlagenen Standards guter wissenschaftlicher Praxis ernst nimmt (Bennett/Checkel 2015b: 21). Wie die Erfahrung zeigt, sind dafür oftmals Jahre erforderlich, und das nur für die Analyse eines einzelnen Falls.

Die Evaluation einer einzelnen Intervention durch *process tracing* stößt an eine weitere Grenze. Ihre Ergebnisse sind auf diesen Fall begrenzt und eignen sich zunächst nicht dazu, Lehren zu ziehen, die Entscheidung über andere Interventionen informieren könnten. Aussagen über die Generalisierbarkeit der Befunde zu einem einzelnen Fall sind allenfalls dann möglich, wenn man weiß, in welcher Hinsicht der untersuchte Fall anderen gleicht oder sich von ihnen abhebt. Dafür muss man die Gesamtmenge der Interventionen und die Ausprägungen der relevanten Faktoren in allen Fällen kennen.

Die entsprechenden Probleme des *process tracing* treten auch auf, wenn es klären soll, warum nach einer Intervention die erwartete oder gewünschte Entwicklung ausgeblieben ist.

Zur Evaluation durch einen Vergleich ausgewählter Interventionen

Wie erwähnt, versucht eine Untersuchung kausaler Zusammenhänge mit Hilfe des *process tracing* in einem ganzen Bündel von potenziellen Faktoren diejenigen zu identifizieren, die tatsächlich zu der beobachteten Veränderung beigetragen haben. Um weitere Gewissheit darüber zu erlangen, ob die Untersuchung die entscheidenden Faktoren identifiziert, wäre ein Fallvergleich denkbar.

Angenommen, ein anderer Fall gleicht dem zu evaluierenden Fall weitgehend und unterscheidet sich von ihm nur mit Blick auf die Intervention und die Entwicklung der Notlage. Hier bliebe allein die Intervention, um die unterschiedlichen Entwicklungen zu erklären.³²

Die Probleme mit diesem Ansatz liegen auf der Hand. Zum einen müssten alle womöglich wichtigen Bedingungen bekannt sein und beim Vergleich berücksichtigt werden. Zum anderen dürften nur selten solche fast gleichen Fälle existieren.³³ Bestenfalls finden sich Fälle, die einander in Bezug auf einige der potenziell relevanten Faktoren ähneln. Ohne Kenntnis des Falluniversums könnte ein solcher Vergleich zweier Fälle schnell dazu führen, andere Fälle mit ähnlichen Konstellationen, aber unterschiedlichen Ergebnisse zu ignorieren und Ergebnisse des Fallvergleichs zu überschätzen. Beispielsweise argumentierte eine vergleichende Untersuchung, die Interventionen im Irak 1991 und 2003 seien in Bezug auf

32 Die Methodenliteratur diskutiert zudem Untersuchungsanlagen, in denen sich die betrachteten Fälle fast vollständig voneinander unterscheiden, aber das gleiche Ergebnis aufweisen (Berg-Schlösser/Cronqvist 2012: 113–116).

33 Als weitere Probleme solcher Vergleichsdesigns gelten, dass sie leicht das Zusammenwirken verschiedener Faktoren übersehen, von deterministischer Kausalität ausgehen und fehlerfreie Messungen voraussetzen (Lieberson 1992: 114).

relevante Faktoren ähnlich gewesen, und hätten sich deutlich in der Autorisierung durch den Sicherheitsrat unterschieden. Der Erfolg 1991 sei der internationalen Zustimmung geschuldet gewesen, der Misserfolg 2003 der fehlenden internationalen Legitimität, die der internen Opposition Auftrieb gegeben habe (Thompson 2009). Erst wenn sich der Blick weitet und Irak 2003 auch mit Fällen wie der Intervention in Afghanistan verglichen wird, wird deutlich, dass die formale internationale Legitimität allenfalls in Kombination mit anderen Faktoren über den Erfolg einer Intervention entscheidet.

Zur Evaluation der Gesamtmenge humanitärer militärischer Interventionen

Die genannten Grenzen der Einzelfalldevaluation verweisen auf den möglichen Mehrwert, die eine Untersuchung der Gesamtklasse der humanitären militärischen Interventionen hat. Wie angedeutet, klärt sie über Besonderheiten und typische Merkmale jedes Falls auf und kann so die Befunde der Evaluation einzelner und weniger Fälle verbessern. Darüber hinaus ermöglicht es eine Evaluation der Gesamtmenge, den Netto-Effekt humanitärer militärischer Interventionen abzuschätzen und die Bedingungen für Erfolg oder Scheitern herauszuarbeiten. Eine solche Gesamtevaluation wäre zunächst einmalig durchzuführen und nicht nach jeder weiteren Intervention zu wiederholen. Erst wenn sich die Menge aller humanitären militärischen Interventionen erheblich vergrößert und verändert hat, könnte sich eine weitere Gesamtevaluation lohnen. Die Analyse, ob humanitäre militärische Interventionen insgesamt eher erfolgreich waren oder gescheitert sind, müsste damit beginnen, ihre Erfolgsraten mit Blick auf die oben vorgeschlagenen Kriterien zu erheben. Allerdings würden die so gewonnenen Ergebnisse noch nicht viel aussagen. Angenommen, eine Evaluation aller humanitären militärischen Interventionen würde ermitteln, dass auf ein Drittel von ihnen ein schnelles Ende der Notlage erfolgte. Die Verfechter solcher Einsätze könnten das zunächst als enttäuschende Nachricht auffassen, würden dann aber fragen, wie oft in vergleichbaren Fällen die Gewalt ohne Intervention andauerte oder aufhörte. Eine Gesamtbilanz muss daher in einem zweiten Schritt alle Fälle humanitärer militärischer Interventionen mit Fällen ohne ein solches Eingreifen vergleichen. Erst diese Gegenüberstellung würde den durchschnittlichen Effekt der Intervention herausarbeiten. Ein solcher Vergleich wäre umso aussagekräftiger, je mehr gewährleistet ist, dass sich die eine Gruppe von der anderen allein durch die humanitäre militärische Intervention unterscheidet. Die Vergleichsgruppe einfach aus allen Gewaltkonflikten ohne Intervention oder aus einer Stichprobe aus diesen zu bilden, wäre nur unter der Annahme zulässig, dass sich Interventionen gleichmäßig über alle Arten von Fällen verteilen und nicht bei bestimmten Merkmalen der potenziellen Zielländer häufiger oder seltener auftreten. Diese Annahme scheint unrealistisch im Lichte bisheriger Arbeiten, deren Themenfeld sich mit der humanitären militärischen Intervention überschneiden. Martin Binder (2015) zeigt etwa, wie selektiv der UN-Sicherheitsrat auf humanitäre Krisen reagiert. Virginia Page Fortna (2008: 29–45) legt dar, dass bestimmte Bedingungen die Entsendung von Friedenstruppen wahrscheinlicher machen. Eine geeignete Vergleichsgruppe ist in zwei Schritten zu erstellen. Zunächst ermittelt man in den Interventionsfällen die arithmetischen Mittel und Streuungsmaße derjenigen Variablen, die Dauer, Intensität, Wiederkehr und Ansteckungseffekte der Gewalt, aber auch das Auftreten der Intervention selbst beeinflussen. Dazu zählen etwa Fläche und Bevölkerungsdichte des Ziellands, Konflikttyp oder die Anzahl der

Konfliktparteien. Anschließend wählt man die Fälle der Vergleichsgruppe so aus, dass hier das arithmetische Mittel und die Streuungsmaße all dieser Variablen denen der Interventionsgruppe möglichst gleichen, damit sich auch die Zusammensetzung der beiden Gruppen möglichst entspricht (vgl. Gilligan/Sergenti 2008 für diese Methode des Matching).³⁴

Hat man die Vergleichsgruppe gebildet, untersucht man die Gewaltentwicklung in ihr und in den Fällen mit einer humanitären militärischen Intervention. Die Ergebnisse zeigen dann den durchschnittlichen Effekt eines solchen Eingreifens. Das wäre wohlgermerkt auch ein kontrafaktisches Vorgehen, da doch die Entwicklungen in der Vergleichsgruppe genutzt werden, um auf die Entwicklungen in der Interventionsgruppe zu schließen, die es ohne ein Eingreifen gegeben hätte. Der Befund könnte etwa lauten, dass in Fällen mit einer humanitären militärischen Intervention ein schnelles Ende der Notlage um 30 Prozent wahrscheinlicher ist als in Gewaltkonflikten ohne ein solches Eingreifen. Das lässt freilich keinen Schluss über den Effekt der Intervention im Einzelfall zu. Dennoch wäre ein solcher Befund politisch relevant. Er könnte den Streit zwischen Befürwortern und Gegnern humanitärer militärischer Interventionen über deren Erfolgsraten versachlichen und Orientierungswissen bereitstellen, das künftige Entscheidungen informiert. Sollte sich etwa herausstellen, dass in der Vergangenheit ein Großteil dieser Einsätze gescheitert ist, würde dies wie ein Warnsignal bei künftigen Entscheidungen wirken. Allerdings steht ein solcher Befund umso mehr infrage, je weniger gesichert ist, dass die Konstruktion der Vergleichsgruppe alle wichtigen Bedingungen erfasst.

Die Kenntnis des durchschnittlichen Effekts aller Intervention liefert keine Einsichten darüber, unter welchen Bedingungen Einsätze erfolgreich sind oder scheitern. Zur Beantwortung dieser Frage halten wir eine separate Analyse nur der Interventionsfälle für aussichtsreich. Angesichts der (noch) geringen Zahl humanitärer militärischer Intervention scheinen uns Zweifel darüber erlaubt, ob hier quantitative Verfahren weiterhelfen. Für geeigneter halten wir eine Methode, die darauf ausgerichtet ist, das Zusammenspiel verschiedener Bedingungen zu erfassen. Die sogenannte Qualitative Comparative Analysis (Ragin 1987; Schneider/Wagemann 2012) kann herausarbeiten, welche Konstellationen mit einem Erfolg der Intervention einhergehen und welche deren Scheitern kennzeichnen. Die Evaluationsforschung ist längst auf diese Methode aufmerksam geworden (Stern et al. 2012: 24, 28).

Diese Analyse beginnt mit einer theoretisch angeleiteten Auswahl derjenigen Faktoren, die Erfolg oder Misserfolg erklären könnten. In einem zweiten Schritt erhebt sie für alle Interventionsfälle die Ausprägung dieser Faktoren und identifiziert in einem dritten Schritt systematisch die Konstellationen von Faktoren, die konsistent mit Erfolg oder Scheitern einhergehen. Sie könnte beispielsweise zum Ergebnis kommen, dass humanitäre militärische Interventionen dann erfolgreich sind, wenn sie militärisch robust ausgestattet sind *und* zugleich von einem politischen Friedensprozess begleitet werden. Die Qualitative Comparative Analysis (QCA) weist die Stärke auf, verschiedene Konstellationen von Fakto-

34 Dabei gilt es, für jedes der vier Kriterien eine Vergleichsgruppe zu konstruieren, denn mit Blick auf Rückfall in eine gewaltsame Notlage und Ansteckungseffekte können auch Dauer und Intensität zu den Einflussfaktoren gehören.

ren herauszuarbeiten, die jeweils zu Erfolg oder Misserfolg führen. Neben der genannten Konstellation könnte zum Beispiel eine unparteiische Intervention in kleinen Zielländern ebenfalls zum Erfolg führen.

Wie bei anderen Methoden auch, hängt die Aussagekraft der Befunde einer QCA von der Auswahl der berücksichtigten Faktoren ab (Amenta/Poulsen 1994). Dies ist insofern problematisch, als nur eine geringe Zahl von Faktoren sinnvoll verarbeitet werden kann (Marx/Duşa 2011). Zudem lässt die QCA nur auf den kausalen Effekt schließen, sagt aber nichts darüber aus, wie die beobachteten Konstellationen von Faktoren das jeweilige Ergebnis herbeigeführt haben. Das Vertrauen in den kausalen Effekt lässt sich stärken, wenn man zu diesem kausalen Effekt einen Mechanismus herausarbeiten kann, der die Wirkung plausibel macht. Die QCA kann das selbst nicht leisten und muss deswegen um *process tracing* ergänzt werden.

Zur Kombination verschiedener Evaluationsmethoden

Bei der Entscheidung über eine Intervention braucht die Politik Orientierungswissen darüber, ob humanitäre militärische Interventionen wirken, wie sie wirken und unter welchen Bedingungen sie Erfolge erzielen oder scheitern. Die Einzelevaluation wie auch die der Gesamtmenge kann jeweils nur einen Teil dieser Fragen beantworten. Während die Evaluation der Gesamtmenge die Fragen des „ob“ und der Bedingungen klären kann, informiert die Einzelevaluation über das „wie“. Eine Kombination von beiden Evaluationsarten und verschiedener Methoden der Gesamtevaluation maximiert das Orientierungswissen (vgl. Stern et al. 2012: 15–18, 31–35). Des Weiteren stärkt der Methodenmix die Befunde der verschiedenen Ansätze.

Das *process tracing* in einzelnen Fällen kann die Befunde der Evaluation aller Interventionen absichern. Die Ergebnisse zur Frage, ob humanitäre militärische Interventionen Gewalt beenden, eindämmen oder verschlimmern, wären belastbarer, wenn das *process tracing* in aussagekräftigen Fällen plausibel machen könnte, wie diese Wirkungen zustande kamen. Auch eine Untersuchung der Erfolgsbedingungen sollte sich nicht mit einer Feststellung von Korrelationen zufrieden geben, sondern die ermittelten Konstellationen durch ein *process tracing* in typischen Fällen auf Plausibilität prüfen. Solche typischen Fälle weisen die zuvor ermittelte Konstellation auf und sollten daher zeigen, ob sich mit ihnen die Entwicklungen hin zum Erfolg oder Scheitern nachvollziehen lassen.

So wie das *process tracing* die Evaluation der Gesamtmenge absichert, können die Befunde zur Gesamtmenge die wirkungsorientierte Evaluation künftiger Einzelfälle verbessern. Die Evaluation aller humanitären militärischen Interventionen soll zunächst einmalig erfolgen. Für künftige Fälle bleibt daher nur die Einzelevaluation, die allerdings von den Ergebnissen der Gesamtevaluation profitieren kann. Aus ihnen leiten sich verbesserte Annahmen über den Kausalmechanismus ab, die sich dann im neuen Einzelfall fokussiert untersuchen lassen. Zudem können die Befunde der Bedingungsanalyse das *process tracing* dazu anleiten, welchen Faktoren es neben der Intervention nachgehen sollte. Schließlich helfen Kenntnisse der Gesamtmenge humanitärer militärischer Interventionen dabei, den neuen Fall in Relation zu anderen zu setzen und etwa seinen Schwierigkeitsgrad zu bestimmen und dadurch politisch zu bewerten.

5. Fazit: Für eine ergebnisoffene wirkungsorientierte Evaluation

Mit diesem Report begrüßen wir die Absicht von Bundestag und -regierung, Auslandseinsätze zu evaluieren. Mit Fokus auf politisch besonders kontroverse Einsätze, nämlich humanitäre militärische Interventionen, legen wir dar, wie das Vorgehen der Evaluierung die zu erwartenden Ergebnisse prägt.

Der Gesetzentwurf sieht vor, dem Bundestag vor der Verlängerung eines Einsatzes eine Zwischenbilanz vorzulegen. Die Ergebnisse einer solchen begleitenden Evaluation können helfen, die Mission besser auszurichten, aber auch die Schlussfolgerung nahelegen, die Intervention vorzeitig zu beenden. Sie taugen aber nicht dazu, alle relevanten Entwicklungen im Zielland zu erfassen und entsprechend den Einsatz zu bewerten. Das ist erst einige Zeit nach der Intervention möglich, weshalb wir die im Gesetzentwurf dokumentierte Absicht unterstützen, den Einsatz nach seinem Abschluss zu evaluieren.

Die Bundesministerien könnten davon profitieren, externe Länder- und Methodenexpertise hinzuzuziehen, gerade wenn es darum geht, den etwaigen Wirkungen einer Intervention nachzuspüren.

Der Gesetzentwurf der Regierungsmehrheit visiert eine Evaluierung an, die überprüfen soll, wieweit die im jeweiligen Auftrag formulierten Ziele erreicht wurden. Ein solcher Ansatz verspricht wertvolle Einsichten, greift unseres Erachtens aber zu kurz. Je nach Formulierung des Mandats könnte dieselbe Entwicklung als Erfolg oder als Scheitern erscheinen. Eine Evaluierung hingegen, die die Wirkungen ins Zentrum stellt, erlaubt einen einheitlichen Maßstab und somit eine Vergleichbarkeit über einzelne Fälle hinaus. Vor allem informiert erst die wirkungsorientierte Evaluation darüber, ob humanitäre Interventionen Erfolge erzielten oder scheiterten und welche Bedingungen jeweils dafür ausschlaggebend waren.

Die Kriterien für Erfolg und Scheitern leiten wir aus dem zentralen Anliegen der humanitären militärischen Intervention ab: dem Schutz der Bürger eines anderen Staates vor einer von Gewalt geprägten Notlage. Vier Kriterien schlagen wir vor: 1) das Ende der Notlage, 2) die Entwicklung der Opferzahlen, 3) den Rückfall in die Gewalt und 4) gewaltsame Notlagen in Nachbarländern. Das Ende oder Zurückgehen der Gewalt im Zielland der Intervention zu verzeichnen, ist das Eine, etwas ganz Anderes hingegen ist es, solche Veränderungen der Intervention kausal zuzuordnen. Im Einzelfall ist diese Attribution lediglich annäherungsweise machbar, da oft selbst mit großem Aufwand das Gewicht der Intervention relativ zu anderen potenziell erklärenden Faktoren unbestimmt bleibt. Die erforderliche Kontrolle von solchen Faktoren ist effizienter fallübergreifend zu leisten. Daher sollte sich die Bilanzierung der Interventionen nicht auf voneinander isolierte Einzelevaluationen beschränken. Wir empfehlen der deutschen Politik, einmalig eine Evaluation aller humanitären militärischen Interventionen in Auftrag zu geben. Das sollte Einsätze ohne deutsche Beteiligung einschließen, um die gewonnenen Kenntnisse auf breitere Basis zu stellen.

Für eine solche Gesamtevaluation regen wir an, den durchschnittlichen Effekt humanitärer militärischer Interventionen durch einen Vergleich aller Fälle mit solchen Einsätzen mit einer sorgsam zu bildenden Gruppe von Fällen ohne ein solches Eingreifen

zu ermitteln. Zudem soll ein Vergleich der Interventionsfälle untereinander herausarbeiten, unter welchen Bedingungen oder Bündeln von Bedingungen solche Einsätze erfolgreich sind oder scheitern. Mit *process tracing* in ausgewählten Fällen lässt sich die Plausibilität dieser Befunde prüfen. Die Befunde der Gesamtevaluation ermöglichen aussagekräftigere Einzelevaluation künftiger Einsätze.

Die hier skizzierte Evaluation humanitärer militärischer Interventionen produziert keineswegs nur Ergebnisse, die sich für die Bücherregale der örtlichen Hochschule eignen, für die politische Praxis aber als nutzlos erweisen. Die Einsichten einer solchen Evaluation können und sollen Politikerinnen und Politikern die schwierige Entscheidung über Beginn und Fortsetzung einer solchen Intervention nicht abnehmen, ihnen jedoch Orientierungswissen bereitstellen. Solche Kenntnisse lassen sich aber nur gewinnen, wenn die Evaluation ergebnisoffen erfolgt und nicht auf das im Gesetzentwurf formulierte Ziel ausgerichtet wird, nachhaltige Unterstützung für die Auslandseinsätze zu sichern.

Aus der Evaluation aller bisherigen humanitären militärischen Interventionen ergibt sich keineswegs eine klare Prognose über den etwaigen Verlauf des Einsatzes, der zur Entscheidung ansteht. Quantitative Befunde informieren nur über Wahrscheinlichkeiten und schaffen keine Gewissheit über Erfolg und Scheitern in einem neuen Fall. Qualitativ gewonnene Einsichten in Bedingungen von Erfolg und Misserfolg vergangener Interventionen stoßen an die Grenze, dass es bei aktuellen Fällen neue Umstände geben kann, die bei der Evaluation noch keine Rolle spielten. Auch klären Kenntnisse früherer Interventionen nicht die Merkmale des gerade zu entscheidenden Falls auf. Oftmals ist vieles unbekannt oder Informationen sind unsicher, etwa über Anliegen und Stärke der einzelnen Konfliktparteien (vgl. Picciotto 2014: 7). Unklar bleibt darüber hinaus, welche Bedingungen nach Beginn der Intervention noch eintreten könnten, so z.B. Parteinahme externer Mächte für die Seite, gegen die sich der Einsatz richtet.

Eine Evaluation mag über die Erfolgsaussichten von humanitären militärischen Interventionen informieren, berührt damit allerdings nur einen von mehreren Aspekten, die Politiker bei ihrer Entscheidung gegeneinander abwägen. Für Grundsatzkritiker des Militärs sind große Erfolgchancen einer Intervention irrelevant, da sie militärische Mittel unabhängig vom Ausgang ihres Einsatzes ablehnen. Wer solche Einsätze nicht pauschal ablehnt, kann sich trotz guter Erfolgsaussichten gegen eine Intervention entscheiden. Das kann etwa der Fall sein, wenn keine anderen Staaten die Intervention mittragen wollen, es Priorität hat, die aufzubringenden finanziellen Mittel für andere Zwecke einzusetzen, oder die eigene Bevölkerung den Einsatz mehrheitlich ablehnt und bald Wahlen anstehen. Ein „Nein“ zu einer Erfolg versprechenden Intervention kann auch darauf beruhen, dass dem Einsatz die völkerrechtliche Grundlage fehlt. Umgekehrt können sich Politiker trotz einer hohen Wahrscheinlichkeit des Scheiterns für eine Intervention entscheiden, wenn sie Handlungsfähigkeit oder Solidarität mit Verbündeten oder den Opfern der Gewalt demonstrieren wollen. Kurz: Bei einer Entscheidung über eine humanitäre militärische Intervention stehen deren Erfolgsaussichten neben einem ganzen Bündel anderer außen- und innenpolitischer Erwägungen. Keine noch so guten Einsichten in Erfolg und Scheitern bisheriger humanitärer militärischer Interventionen nehmen den Entscheidungsträgern die schwere Last der politischen Verantwortung von den Schultern.

Literatur

- Amenta, Edwin/Poulsen Jane D.* 1994: Where to Begin. A Survey of Five Approaches to Selecting Independent Variables for Qualitative Comparative Analysis, in: *Sociological Methods & Research*, 23: 1, 22–53.
- Auswärtiges Amt* 2014: Rede von Außenminister Frank-Walter Steinmeier anlässlich der 50. Münchner Sicherheitskonferenz, 31.1.2014, <http://bit.ly/1VzkGO5> (26.4.2016).
- Beach, Derek/Pedersen, Rasmus Brun* 2013: *Process-Tracing Methods. Foundations and Guidelines*, Ann Arbor, MI: University of Michigan Press.
- Befani, Barbara/Mayne, John* 2014: Process Tracing and Contribution Analysis: A Combined Approach to Generative Causal Inference for Impact Evaluation, in: *IDS Bulletin*, 45: 6, 17–36.
- Bennett, Andrew B.* 2008: Process Tracing: a Bayesian Perspective, in: Box-Steffensmeier, Janet M./Brady, Henry E./Collier, David (Hrsg.): *The Oxford Handbook of Political Methodology*, New York, NY: Oxford University Press, 702–721.
- Bennett, Andrew B./Checkel, Jeffrey T. (Hrsg.)* 2015a: *Process Tracing. From Metaphor to Analytical Tool*, Cambridge: Cambridge University Press.
- Bennett, Andrew B./Checkel, Jeffrey T.* 2015b: Process Tracing: From Philosophical Roots to Best Practices, in: Bennett, Andrew B./Checkel, Jeffrey T. (Hrsg.): *Process Tracing. From Metaphor to Analytical Tool*, Cambridge: Cambridge University Press, 3–37.
- Berg-Schlosser/Cronqvist, Lasse* 2012: *Aktuelle Methoden der Vergleichenden Politikwissenschaft. Einführung in konfigurationelle (QCA) und makro-quantitative Verfahren*, Opladen/Farmington Hills: Verlag Barbara Budrich.
- Binder, Martin* 2015: Paths to Intervention: What Explains the UN's Selective Response to Humanitarian Crises?, in: *Journal of Peace Research*, 52: 6, 712–726.
- Brockhaus* 1997: *Die Enzyklopädie in 24 Bänden, Band 6, 20., überarbeitete und aktualisierte Auflage*, Leipzig: wissenmedia.
- Bundesministerium der Verteidigung* 2014: Rede der Bundesministerin der Verteidigung, Dr. Ursula von der Leyen, anlässlich der 50. Münchner Sicherheitskonferenz München, 31.1.2014, <http://bit.ly/1SwGSCy> (26.4.2016).
- Bundesregierung* 2014: 2014 Progress Report on Afghanistan, November 2014, <http://bit.ly/1NOjxKB> (26.4.2016).
- Burnham, Gilbert/Lafta, Riyadh/Doocy, Shannon/Roberts, Les* 2006: Mortality after the 2003 Invasion of Iraq: a Cross-Sectional Cluster Sample Survey, in: *Lancet*, 368: 1421–1428.
- Chivvis, Christopher S.* 2014: *Toppling Quaddafi. Libya and the Limits of Liberal Intervention*, New York, NY: Cambridge University Press.
- Coghlan, Benjamin/Ngoy, Pascal/Mulumba, Flavie/Hardy, Colleen/Nkamgang Bemo, Valerie/Stewart, Tony/Lewis, Jennifer/Brennen, Richard* 2008: Mortality in the Democratic Republic of Congo. An Ongoing Crisis, <http://bit.ly/1T1qsBo> (26.4.2016).

- Croicu, Mihai Catalin/Sundberg, Ralph* 2015: UCDP Georeferenced Event Dataset Codebook Version 2.0, Department of Peace and Conflict Research, Uppsala University, www.pcr.uu.se/research/ucdp/datasets/ucdp_ged/ 10.5.2016).
- De Maizière, Thomas* 2012: Internationale Verantwortung wahrnehmen – Deutschlands Rolle in der Welt von heute, Rede im Rahmen des „Forum Frauenkirche“, 1.11.2012, <http://bit.ly/1MVeD3s> (26.4.2016).
- DeGEval – Gesellschaft für Evaluation* 2008: Standards für Evaluation, Hamburg, <http://bit.ly/1VPDLMk> (26.4.2016).
- Dembinski, Matthias/Gromes, Thorsten* 2015: Humanitäre militärische Interventionen 1946–2014. Annäherungen an ein umstrittenes Thema, in: Kursawe, Janet/Johannsen, Margret/Baumgart-Ochse, Claudia/von Boemcken, Marc/Werkner, Ines-Jacqueline (Hrsg.): Friedensgutachten 2015, Münster: Lit-Verlag, 75–86.
- Deutscher Bundestag* 2011a: Antrag der Fraktion BÜNDNIS 90/DIE GRÜNEN, Drucksache 17/5099, <http://bit.ly/1VzIPF6> (26.4.2016).
- Deutscher Bundestag* 2011b: Stenografischer Bericht 105. Sitzung, Plenarprotokoll 17/105, Berlin, <http://dipbt.bundestag.de/doc/btp/17/17105.pdf> (26.4.2016).
- Deutscher Bundestag* 2012: Beschlussempfehlung und Bericht des Auswärtigen Ausschusses (3. Ausschuss) zu dem Antrag der Abgeordneten Omid Nouripour, Hans-Christian Ströbele, Marieluise Beck (Bremen), weiterer Abgeordneter und der Fraktion BÜNDNIS 90/DIE GRÜNEN – Drucksache 17/5099 –, Drucksache 17/8697, <http://dip.bundestag.de/btd/17/086/1708697.pdf> (26.4.2016).
- Deutscher Bundestag* 2013: Stenografischer Bericht 217. Sitzung, Plenarprotokoll 17/217, Berlin, <http://dip21.bundestag.de/dip21/btp/17/17217.pdf> (26.4.2016).
- Deutscher Bundestag* 2014: Stenografischer Bericht 76. Sitzung, Plenarprotokoll 18/76, Berlin, <http://dipbt.bundestag.de/doc/btp/18/18076.pdf> (26.4.2016).
- Deutscher Bundestag* 2015: Unterrichtung durch die Kommission zur Überprüfung und Sicherung der Parlamentsrechte bei der Mandatierung von Auslandseinsätzen der Bundeswehr, Drucksache 18/5000, Berlin, <http://bit.ly/1MVfeIE> (26.4.2016).
- Deutscher Bundestag* 2016: Gesetzentwurf der Fraktionen der CDU/CSU und SPD. Entwurf eines Gesetzes zur Fortentwicklung der parlamentarischen Beteiligung bei der Entscheidung über den Einsatz bewaffneter Streitkräfte im Ausland im Zuge fortschreitender Bündnisintegration, Drucksache 18/7360, <http://bit.ly/1QwXtXR> (26.4.2016).
- Diehl, Paul* 1994: International Peacekeeping, London: Johns Hopkins University Press.
- Dixon, Jeffrey* 2009: What Causes Civil Wars? Integrating Quantitative Research Findings, in: International Studies Review, 11: 4, 707–735.
- Eck, Kristine/Hultman, Lisa* 2007: Violence against Civilians in War, in: Journal of Peace Research, 44: 2, 233–246.
- Farina, Fabrizio Fitzgerald/William, Andrew/Kennedy-Chouane, Megan* 2013: Civilian and Military Evaluation and Assessment: Synergies and Differences, in: Williams, Andrew/Bexfield, James/Farina, Fabrizio Fitzgerald/de Nijs, Johannes (Hrsg.): Innovation in Operations Assessment: Recent Developments in Measuring Results in Conflict

- Environments, The Hague: NATO Communications and Information Agency, 117–144, <http://bit.ly/1Uec9Pg> (26.4.2016).
- Fortna, Virginia Page 2008: *Does Peacekeeping Work? Shaping Belligerents' Choices after Civil War*, Princeton, NJ: Princeton University Press.
- Gauck, Joachim 2014: *Deutschlands Rolle in der Welt. Anmerkungen zu Verantwortung, Normen und Bündnissen*, Rede anlässlich der Eröffnung der Münchner Sicherheitskonferenz am 31.1.2014, <http://bit.ly/1G1yaKv> (26.4.2016).
- Gilligan, Michael J./Sergenti, Ernest J. 2008: *Do UN Interventions Cause Peace? Using Matching to Improve Causal Inference*, in: *Quarterly Journal of Political Science*, 3: 2, 89–122.
- GIZ 2012: *Results: Good, But Too Local. Summary of the Cross-Section Evaluation Crisis Prevention and Peacebuilding*, Bonn/Eschborn, <http://bit.ly/1UectNK> (26.4.2016).
- Gleditsch, Nils Petter/Wallensteen, Peter/Eriksson, Mikael/Sollenberg, Margareta/Strand, Håvard 2002: *Armed Conflict 1946–2001: A New Dataset*, in: *Journal of Peace Research*, 39: 5, 615–637.
- Grävingholt, Jörn/Leininger, Julia/Haldenwang, Christian von/German Development Institute 2012: *Effective Statebuilding? A Review of Evaluations of International Statebuilding Support in Fragile Contexts*, Ministry of Foreign Affairs of Denmark, Evaluation Study 2012/3, <http://bit.ly/21e1Rj5> (26.4.2016).
- Gromes, Thorsten/Dembinski, Matthias 2013: *Bestandsaufnahme der humanitären militärischen Interventionen zwischen 1947 und 2005*, HSFK-Report 2, Frankfurt am Main.
- Holzgreffe, J.L. 2003: *The Humanitarian Intervention Debate*, in: Holzgreffe, J.L./Keohane, Robert O. (Hrsg.): *Humanitarian Intervention. Ethical, Legal and Political Dilemmas*, Cambridge: Cambridge University Press, 15–52.
- Jöst, Lena 2009: *Kosovo – Vorgeschichte und Folgen des NATO-Krieges*, in: Ruf, Werner/Jöst, Lena/Strutynski, Peter/Zollet, Nadine: *Militärinterventionen: Verheerend und völkerrechtswidrig. Möglichkeiten friedlicher Konfliktlösung*, Berlin: Karl Dietz Verlag, 101–132.
- Jöst, Lena/Strutynski, Peter 2009: *Humanitär intervenieren – aber nur mit humanitären Mitteln!*, in: Ruf, Werner/Jöst, Lena/Strutynski, Peter/Zollet, Nadine: *Militärinterventionen: verheerend und völkerrechtswidrig. Möglichkeiten friedlicher Konfliktlösung*, Berlin: Karl Dietz Verlag, 9–30.
- Kirchenamt der EKD 2013: *„Selig sind die Friedfertigen“*. Der Einsatz in Afghanistan. Aufgaben evangelischer Friedensethik, Hannover.
- Kuperman, Alan J. 2013: *A Model Humanitarian Intervention? Reassessing NATO's Libya Campaign*, in: *International Security*, 38: 1, 105–136.
- Kuperman, Alan J. 2015: *Obama's Libya Debacle. How a Well-Meaning Intervention Ended in Failure*, in: *Foreign Affairs*, 94: März/April, 66–77.
- Levy, Jack S. 2008: *Counterfactuals and Case Studies*, in: Box-Steffensmeier, Janet M./Brady, Henry E./Collier, David (Hrsg.): *The Oxford Handbook of Political Methodology*, New York, NY: Oxford University Press, 627–644.

- Lieberson, Stanley* 1992: Small N's and Big Conclusions: An Examination of the Reasoning in Comparative Studies Based on a Small Number of Cases, in: Ragin, Charles C./Becker, Howard S. (Hrsg.): What is a Case? Exploring the Foundations of Social Inquiry, Cambridge: Cambridge University Press, 105–118.
- Lustick, Ian S.* 2010: Tetlock and Counterfactuals: Saving Methodological Ambition from Empirical Findings, in: *Critical Review*, 22: 4, 427–447.
- Malone, Michael D./Miller, William H./Robben, Joseph W.* 1986: From Presence to American Intervention, in: *Survival*, 28: 5, 422–430.
- Marx, Axel/Duša, Adrian* 2011: Crisp-set Qualitative Comparative Analysis (csQCA), Contradictions and Consistency Benchmarks for Model Specification, in: *Methodological Innovations Online*, 6: 2, 103–148.
- Maull, Hanns W.* 2008: Wissenschaftliche Außenpolitik-Evaluation: Ein Oxymoron?, in: *Zeitschrift für Internationale Beziehungen*, 15: 1, 113–123.
- Mertens, Donna M.* 1998: *Research Methods in Education and Psychology. Integrating Diversity with Quantitative & Qualitative Approaches*, London: Sage.
- Moore, John N. (Hrsg.)* 1991: *The Arab-Israeli Conflict, Volume IV: The Difficult Search for Peace (1975–1988)*, Princeton, NJ: Princeton University Press.
- Myrick, Rachel/Reid, Lindsay/Kadera, Kelly M./Crescenzi, Mark J.C.* 2015: Conflict Environments and Civil War, <http://unc.live/1VPEIEg> (11.8.2015).
- OECD 2012: *Evaluating Peacebuilding Activities in Settings of Conflict and Fragility. Improving Learning for Results, DAC Guidelines and Reference Series*, OECD Publishing, <http://dx.doi.org/10.1787/9789264106802-en> (26.4.2016).
- Paffenholz, Thania et al.* 2011: *Der Zivile Friedensdienst: Synthesebericht. Evaluierungsberichte 054*, Bonn: Bundesministerium für wirtschaftliche Zusammenarbeit und Entwicklung, <http://bit.ly/1SNe0Gn> (4.5.2016).
- Palumbo, Dennis J./Nachmias, David* 1983: The Preconditions for Successful Evaluation: Is There an Ideal Paradigm?, in: *Policy Sciences*, 16: 1, 67–79.
- Pape, Robert A.* 2012: When Duty Calls. A Pragmatic Standard of Humanitarian Intervention, in: *International Security*, 37: 1, 41–80.
- Pearson, Frederic S./Baumann, Robert A.* 1993: *International Military Intervention, 1946–1988*, Inter-University Consortium for Political and Social Research, Data Collection 6035, University of Michigan, Ann Arbor, MI, Daten unter <http://bit.ly/1SGVu4m> (2.11.2015).
- Pearson, Frederic/Olson Lounsbury, Mary/Talentino, Andrea Kathryn* 2013: How Effective is International Military Intervention? The Evolution of Motives, Forms and Outcomes, in: Hegemann, Hendrik/Heller, Regina/Kahl, Martin (Hrsg.): *Studying 'Effectiveness' in International Relations. A Guide for Students and Scholars*, Opladen: Barbara Budrich Publishers, 123–142.
- Peen Rodt, Annemarie* 2014: *The European Union and Military Conflict Management. Defining, Evaluating and Achieving Success*, London: Routledge.

- Pettersson, Therése* 2015: UCDP Battle-Related Deaths Dataset Codebook, Version 5.0-2015, <http://bit.ly/1YRu0KT> (26.4.2016).
- Picciotto, Robert* 2014: Have Development Evaluators Been Fighting the Last War ... And If so, What Is to Be Done?, in: *IDS Bulletin*, 45: 6, 6–16.
- Pickering, Jeffrey/Kisangani, Emizet F.* 2009: The International Military Intervention Dataset: An Updated Resource for Conflict Scholars, in: *Journal of Peace Research*, 46: 4, 589–599, <http://bit.ly/1Wns8L0> (2.11.2015).
- Punton, Melanie/Welle, Katharina* 2015: Straws-in-the-Wind, Hoops and Smoking Guns: What can Process Tracing Offer to Impact Evaluation?, Centre for Development Impact Practice Paper 10, <http://bit.ly/1rwrJdW> (26.4.2016).
- Ragin, Charles C.* 1987: *The Comparative Method. Moving Beyond Qualitative and Quantitative Strategies*, Berkeley, CA: University of California Press.
- Rudolf, Peter/Lohmann, Sascha* 2013: Außenpolitikevaluation im Aktionsfeld Krisenprävention und Friedensaufbau, Stiftung Wissenschaft und Politik, SWP-Studie S 20, Berlin, <http://bit.ly/1SwJHDJ> (26.4.2016).
- Schimmelfennig, Frank* 2015: Efficient Process Tracing: Analyzing the Causal Mechanisms of European Integration, in: Bennett, Andrew B./Checkel, Jeffrey T. (Hrsg.): *Process Tracing. From Metaphor to Analytical Tool*, Cambridge: Cambridge University Press, 98–125.
- Schneider, Carsten Q./Wagemann, Claudius* 2012: *Set-Theoretic Methods for the Social Sciences. A Guide to Qualitative Comparative Analysis*, Cambridge: Cambridge University Press.
- Schoch, Bruno/Heinemann-Grüder, Andreas/Hippler, Jochen/Weingardt, Markus/Mutz, Reinhard* 2007: Stellungnahme der Herausgeber: Aktuelle Entwicklungen und Empfehlungen, in: dieselben (Hrsg.): *Friedensgutachten 2007*, Münster: LIT-Verlag, 3–26.
- Seybolt, Taylor B.* 2007: *Humanitarian Military Intervention. The Conditions for Success and Failure*, Oxford: Oxford University Press.
- Spagat, Michael* 2010: Ethical and Data-Integrity Problems in the Second Lancet Survey of Mortality in Iraq, in: *Defence and Peace Economics*, 21: 1, 1–41.
- Spiegel, Paul B./Salama, Peter* 2000: War and Mortality in Kosovo, 1998–99: An Epidemiological Testimony, in: *Lancet*, 355, 2204–2209.
- Stern, Elliot/Stame, Nicoletta, Mayne, John/Forss, Kim/Davies, Rick/Befani, Barbara* 2012: Broadening the Range of Designs and Methods for Impact Evaluations. Report of a Study Commissioned by the Department for International Development, DFID Working Paper 38, <http://bit.ly/1EWylEr> (26.4.2016).
- Stockmann, Reinhard/Meyer, Wolfgang* 2014: *Evaluation. Eine Einführung*, 2. Auflage, Opladen: Verlag Barbara Budrich.
- Sundberg, Ralph/Eck, Kristine/Kreutz, Joakim* 2012: Introducing the UCDP Non-State Conflict Dataset, in: *Journal of Peace Research*, 49: 2, 351–362.
- Sundberg, Ralph/Melander, Erik* 2013: Introducing the UCDP Georeferenced Event Dataset, in: *Journal of Peace Research*, 50: 4, 523–532.

- Tetlock, Philip E.* 2006: *Expert Political Judgment. How Good Is It? How Can We Know?*, Princeton, NJ: Princeton University Press.
- Tetlock, Philip E./Belkin, Aaron (Hrsg.)* 1996: *Counterfactual Experiments in World Politics. Logical, Methodological, and Psychological Perspectives*, Princeton, NJ: Princeton University Press.
- Thompson, Alexander* 2009: *Channels of Power. The UN Security Council and U.S. Statecraft in Iraq*, Ithaca, NY: Cornell University Press .
- United Nations General Assembly* 2015: *Report of the High-Level Independent Panel on Peace Operations on Uniting our Strengths for Peace: Politics, Partnership and People*, A/70/95-S/2015/446, <http://bit.ly/1UIqkgQ> (26.4.2016).
- United Nations Security Council* 1994: *Resolution 929*, <http://bit.ly/1T1sO3f> (26.4.2016).
- United Nations Security Council* 2011: *Resolution 1973*, <http://bit.ly/1TaKA4y> (26.4.2016).
- United Nations Security Council* 2015: *7389th Meeting, S/PV.7389*, <http://bit.ly/1QybQaY> (26.4.2016).
- Weiss, Thomas G.* 2007: *Humanitarian Intervention: Ideas in Action*, Cambridge: Polity.
- Wheeler, Nicholas J.* 2000: *Saving Strangers. Humanitarian Intervention in International Society*, Oxford: Oxford University Press.
- Williams, Andrew/Bexfield, James/Bell, Andrew/Pennell, Bruce* 2013: *The Rationale, Challenges and Opportunities in Operations Assessments*, in: *Williams, Andrew/Bexfield, James/Farina, Fabrizio Fitzgerald/de Nijs, Johannes (Hrsg.): Innovation in Operations Assessment: Recent Developments in Measuring Results in Conflict Environments*, The Hague: NATO Communications and Information Agency, 3–15, www.act.nato.int/images/stories/media/capdev/capdev_01.pdf (26.4.2016).