

Holger Straßheim, Oliver Schwab

Politikevaluation und Evaluationspolitik

Zusammenfassung

Es gibt inzwischen kaum ein Politikfeld, in dem Evaluierung nicht in irgendeiner Form praktiziert wird. Die Evaluierungslandschaft zeichnet sich dabei sowohl durch Vereinheitlichungs- als auch Pluralisierungstendenzen aus. In unterschiedlichen Evaluierungsarrangements kommen spezifische Akteure, Praktiken und Standards zusammen. Zudem handelt es sich nicht um politikfreie Räume. Vielmehr sind diese „Varieties of Evaluation“ selbst Ausgangspunkt, Medium und Gegenstand von Politiken der Evaluation, also der Auseinandersetzung um legitime Formen der Politikbewertung. Der Themenschwerpunkt unternimmt eine Bestandsaufnahme, indem er die Formen und Folgen von Politikevaluation als Evaluationspolitik sowohl in theoretisch wie empirisch ausgerichteten Beiträgen nachzeichnet. In interdisziplinärer Absicht werden aktuelle Befunde der Politik- und Verwaltungswissenschaft, der Governance- und Policy-Analyse, der politischen Soziologie und der Soziologie der Bewertung vorgestellt.

Schlagwörter: Evaluierung, Evaluationspolitik, Wirkungen, Wertungen, Kausalität

Abstract

The Evaluation of Politics and the Politics of Evaluation

Instruments and modes of evaluation are expanding across policy areas and jurisdictions. The emerging landscape can be characterized by dynamics of both standardization and pluralisation. Forming distinct arrangements of evaluation, particular actors, practices and standards are closely interlinked. In and between these “varieties of evaluation”, a multiplicity of actors such as policy makers, administrators, experts, consultants or stakeholders is struggling over legitimate modes of appraising and valuating public policy. This special issue brings together theoretical as well as empirical studies on the politics of policy evaluation. It presents findings from a broad interdisciplinary background including public policy and governance, comparative policy analysis, administration research, political sociology and the sociology of valuation and evaluation.

Keywords: Evaluation, Politics of Evaluation, Causality, Valuation, Impact

1 Einleitung

Verfahren der Evaluierung haben ursprünglich im Fahrwasser von New-Public-Management-Reformen, in den vergangenen zwanzig Jahren jedoch vor allem im Zuge der Bestrebungen einer „evidenzbasierten Politik“ auf nationaler, internationaler und transnationaler Ebene eine hohe Verbreitung erfahren (Pollitt & Bouckaert, 2011; Straßheim, 2017b; Vedung, 2008). Es gibt inzwischen kaum ein Politikfeld, in dem

Evaluierung nicht in irgendeiner Form praktiziert wird. Die Evaluierungslandschaft zeichnet sich dabei sowohl durch Vereinheitlichungs- als auch Pluralisierungstendenzen aus, in deren Ergebnis sich alleine in Deutschland ein großes Spektrum unterschiedlicher Arrangements¹ herausgebildet hat. Die Bandbreite reicht von der anlassbezogenen Evaluierung experimenteller Politik (z. B. die Evaluierung der Hartz-Reformen in der Arbeitsmarktpolitik, siehe auch Petra Kaps in diesem Band), über die Etablierung von Evaluierungsfunktionen innerhalb von Organisationen (z. B. bei der Evaluierung von Lehre in den Hochschulen oder den Peer-Review-Verfahren der Deutschen Forschungsgemeinschaft (DFG), siehe auch Cornelia Schendzielorz und Martin Reinhart in diesem Band) und die häufig praktizierte Evaluierung einzelner fachpolitischer Programme bis zur Einrichtung von Strukturen mit spezifischen Evaluierungskompetenzen (z. B. das Deutsche Institut für Entwicklungspolitik, siehe auch Jörg Faust in diesem Band). Mit den Arrangements variieren auch Organisationsformen, Ansätze, Methoden und Designs der Evaluierungen erheblich. Es ist davon auszugehen, dass wir es mit „Varieties of Evaluation“ (Schwab, 2017) zu tun haben, die sich in Wechselwirkung mit ihrem jeweiligen Kontext entwickelt haben und jeweils spezifische Akteurskonstellationen, Evaluierungspraktiken und Standards aufweisen. Diese Evaluierungsarrangements bilden keine politikfreien Räume. Vielmehr sind sie selbst Ausgangspunkt, Medium und Gegenstand von Politiken der Evaluation, also der Auseinandersetzung um legitime Formen der Politikbewertung (Bjørnholt & Larsen, 2014; Boswell, 2018; Bovens, Hart, & Kuipers, 2008; Stephenson, Schoenefeld & Leeuw, 2019; Weiss, 1993 [1973]).

Evaluierungsarrangements sind durch eine besondere Positionierung im Verhältnis zur Wissenschaft einerseits und zu Politik und Verwaltung andererseits sowie durch bestimmte Anforderungen gekennzeichnet (DeGEval, 2017): Zum einen sind Evaluierungen durch ein systematisches Vorgehen gekennzeichnet und empirisch fundiert, womit sie anschlussfähig an (fach-)wissenschaftliche Diskurse sind. Zum anderen sind sie auf eine Bewertung des Nutzens im Hinblick auf einen bestimmten Verwertungskontext oder der allgemeiner verstandenen Güte des Gegenstandes ausgerichtet, womit sie mit politisch-administrative Verfahren koppelbar werden.² Evaluierungen stellen Transparenz über die Bewertungskriterien her. Sie können sich auf unterschiedliche Gegenstände beziehen (Projekte, Programme, Politiken, Organisationen). Es ist diese besondere Positionierung zwischen Wissenschaft und Politik/Verwaltung, die Evaluierungsprozesse prägt und die Entwicklung von Evaluierungsarrangements für Einflüsse von beiden Seiten öffnet (Bovens, Hart & Kuipers, 2008). Jenseits der normativen Orientierung durch die DeGEval - Gesellschaft für Evaluation (DeGEval) als Fachgesellschaft ist in der Praxis die Verwendung des Begriffs „Evaluierung“ genauso vielfältig wie die sich herausbildenden Evaluierungsarrangements, und die begriffliche Auffächerung ist eng mit dieser Differenzierung verknüpft.

Derzeit gibt es – vor allem im deutschsprachigen Raum – wenig Forschung und damit wenig Erkenntnisse über die institutionelle und kulturelle Einbettung von Evaluationsarrangements einerseits und über die Formen und Folgen von Evaluationspolitiken andererseits. Dies jedoch wäre eine zentrale Voraussetzung, um im inter- und transnationalen Vergleich wie auch im Vergleich von Politikfeldern Aussagen über zukünftige Entwicklungen der Evaluierungslandschaft und über Bedingungen der Lernfähigkeit und Reflexivität von Politikevaluierungen und Evaluierungspolitiken treffen zu können. Unter anderem ergeben sich dabei folgende Forschungsfragen:

- Welche *Varieties of Evaluation* im Sinne distinkter Arrangements von Akteurskonstellationen, Evaluierungspraktiken und Standards lassen sich identifizieren? Durch welche Merkmale unterscheiden sich diese Evaluierungsarrangements? Handelt es sich um gebietskörperschaftliche Varianten, die etwa mit bestimmten Verwaltungskulturen einhergehen? Oder konstituieren sich die Varianten in bestimmten gesellschaftlichen Sektoren und Politikfeldern? Was sind die Mechanismen der Konstitution, Transformation oder auch des Verschwindens solcher Arrangements?
- Inwiefern erfüllt sich die Hoffnung, aus Evaluierungsverfahren Impulse für politische Lernprozesse zu gewinnen? Führen bestimmte Ausprägungen der Evaluationspolitik statt zu einer kritischen Beurteilung zu einer ‚Disziplinierung und Zählung‘ kritischer Perspektiven? Wohnt bestimmten Evaluationspolitiken eine Tendenz der Verhüllung von Politikversagen und nicht intendierten Nebeneffekten inne? Unter welchen Bedingungen erhellen bzw. verdunkeln Evaluationen das politische Verständnis von komplexen Problemlagen?

Der Themenschwerpunkt unternimmt eine Bestandsaufnahme zu diesen Fragestellungen und Forschungslinien, indem er die Formen und Folgen von Politikevaluation als Evaluationspolitik sowohl in theoretisch wie empirisch ausgerichteten Beiträgen nachzeichnet. Er präsentiert in interdisziplinärer Absicht aktuelle Forschungen der Politik- und Verwaltungswissenschaft, der Governance- und Policy-Analyse, der politischen Soziologie und der Soziologie der Bewertung.³ Diese Einführung entwickelt zunächst auf der Grundlage der Beiträge und der einschlägigen Literatur mögliche Ansätze und Perspektiven einer Analyse von Politikevaluation als Evaluationspolitik. Dazu werden in einem ersten Schritt Grundlagen und Konzepte von Evaluation als politischer Praxis vorgestellt. In einem zweiten Schritt sollen dann Erkenntnisse über die Ausprägung und den Wandel der *Varieties of Evaluation* zusammengetragen werden. In einem dritten Schritt stellen wir knapp die Beiträge vor. Abschließend skizzieren wir Dimensionen und Typen der Politisierung und des Politischen in Evaluationen. Diese reichen von einer öffentlich und kontrovers auf der „Vorderbühne“ ausgetragenen Evaluationspolitik über deren symbolische oder sogar simulative Aufführung, einer „stillen“ Evaluationspolitik auf der „Hinterbühne“ bis hin zu einer Entpolitisierung von Evaluationen.

2 Evaluation als politische Praxis

Vor nunmehr 47 Jahren plädierte Carol Weiss (Weiss, 1993; 1993 [1973]) in ihrem für ein ganzes Genre der Forschungsliteratur vorausweisenden Artikel „Where Politics and Evaluation Research Meet“ dafür, Evaluationen als einen politischen Prozess zu begreifen. Sie wendet sich damit gegen ein rational-instrumentelles Modell der Trennung zwischen Evaluation und Entscheidung, das Evaluationen als Orte der Produktion von objektiven, politikfreien Fakten auffasst, auf deren Grundlage politische Entscheidungen optimiert werden können (dazu auch Bovens, Hart & Kuipers, 2008). Evaluationen sind jedoch laut Weiss aus drei Gründen mit Politik eng verwoben: Erstens bilden die Objekte von Evaluationen – Politikprogramme, Instrumente und andere Interventionen – gewissermaßen Epizentren politischer Spannungen, Kontroversen und Verhandlungen.

gen (1993 [1973], pp. 94-97). In ihnen manifestieren sich die Positionen konkurrierender, um Einfluss ringender Akteure mit teilweise weitreichenden sozialen und politischen Folgewirkungen. Über die Lebenszeit eines Programmes hinweg geraten Evaluationen direkt in das Kräftefeld dieser Auseinandersetzungen. So divergieren unter Umständen die Interpretationen der Programmziele zwischen Politik und Verwaltung einerseits und der Evaluierenden andererseits. Mitunter kommt es zu Zielkonflikten. Veränderungen der Programmgestaltung oder -ausrichtung ordnen das politische Kräftefeld neu und sorgen so dafür, dass sich die an der Evaluation Beteiligten mit geänderten Erwartungskonstellationen konfrontiert sehen. Alice O’Connors (2001) historische Analyse sozialpolitischer Evaluationen in den USA zeigt auf eindrückliche Weise, wie spätestens mit den *Welfare*-Reformen in den 1980er Jahren Expertinnen⁴ und Evaluierende von einem Wandel des politischen und ideologischen Klimas gleichsam überholt und teilweise marginalisiert wurden – einem Wandel, dem sie allerdings zuvor durch einen rivalisierenden Wettbewerb um die einflussreicheren Evaluationsmethoden („analytic warfare“) den Weg bereitet hatten.

Zweitens finden die Ergebnisse von Evaluationen in Entscheidungsprozessen häufig keine oder eine nur schwer nachverfolgbare Verwendung (Beck & Bonß, 1984; Caplan, Morrison & Stambaugh, 1975; Weiss, 1980; 1993 [1973], pp. 97-99). Das liegt zum einen an den Zeithorizonten und Rhythmen politischer Entscheidungen und Verfahren, die von denen der Evaluationsforschung abweichen. Wahl- oder Budgetzyklen erzeugen einen Handlungs- und Rechtfertigungsdruck, dem sich politische Akteure kaum entziehen können und der sich nicht ohne weiteres mit den Abläufen von Evaluationsprojekten synchronisieren lässt. Zum anderen folgen politisch-administrative Prozesse eigenen Relevanzen und Prioritäten, die mit darüber bestimmen, ob und in welcher Form Evaluierungsergebnisse zur Geltung kommen. Um resonanzfähig zu sein, müssen Evaluationsberichte in die Präferenzhierarchien und Darstellungslogiken von Ministerien oder Behörden übersetzt werden. Außerdem können im Zuge von geänderten legislativen oder administrativen Interessenlagen die ursprünglichen Ziele und Bewertungskriterien in einem anderen Licht erscheinen. Der Evaluation selbst wird dann möglicherweise eine geringere (oder eine höhere) Bedeutung zugemessen als erwartet. Evaluationen sind auch deswegen notorisch mit „Übersetzungsproblemen“ konfrontiert (Schwab, 2013).

Weiss hebt jedoch noch einen dritten Aspekt hervor, der sich nicht auf den politischen Kontext – den Gegenstand oder die Wirkung – von Evaluationen bezieht und den eigentlichen Kern ihrer Argumentation ausmacht: Evaluationen sind implizit politisch (Weiss, 1993 [1973], pp. 100-106). Weiss überschreitet damit die Grenze zwischen Evaluation und Entscheidung und verortet das Politische im Evaluieren selbst (zur Überwindung des Trennungsmodells Jasanoff, 2011). Es liegt in der Natur von Geltungsaussagen in Evaluationen, dass sie immer gleichzeitig auch eine politisch-wertende Haltung („political stance“, Weiss, 1993 [1973], p. 94) einnehmen. Durch die Bewertung einiger Programme als problematisch und anderer als erfolgreich, durch die Akzeptanz bestimmter Programmziele und die Hinnahme ihrer Prämissen wirken Evaluationen legitimierend. Die Fokussierung auf spezifische Kontextfaktoren und Wirkungen einer Intervention blendet andere Problemdimensionen und Strukturbedingungen aus. Was als methodische Notwendigkeit ausgegeben wird, hat möglicherweise weitreichende politische Implikationen, weil es die Aufmerksamkeit und Wahrnehmungsweisen politischer Akteure durch Schemata strukturiert (dazu Scott, 1998 und

Regine Paul in diesem Band). Das ist unvermeidbar und sollte nicht zur Suche nach neuen Technologien und Methoden der Neutralisierung und Objektivierung verleiten. Vielmehr, so schlussfolgert Weiss, sollte es um eine Sensibilisierung für die politische Dimension des Evaluierens gehen, damit für alle Beteiligten die politischen Implikationen und Nebenfolgen des Evaluierens sichtbar, kommunizierbar und strategisch bearbeitbar werden (zu politiksensiblen Evaluationskonzepten Bovens, Hart & Kuipers, 2008; Fischer, 1999).

Auch die neuere kritische Politikforschung greift auf dieses grundlegend politische Verständnis von Evaluationen zurück, wenn sie sich – wie vor mehr als vier Dekaden auch Weiss – gegen ein nach wie vor wirksames rational-instrumentelles Evaluationsverständnis wendet (Adelle, Jordan & Turnpenny, 2012; Bjørnholt & Larsen, 2014; Bovens, Hart & Kuipers, 2008; Stephenson, Schoenefeld & Leeuw, 2019). Vor allem der Aufschwung der evidenzbasierten Politik seit Ende der 1990er Jahre hat zu einer (Wieder-)Belebung des kybernetischen Modells geführt, wonach Evidenzen als augenscheinlich überzeugende und verlässliche Wissensbestände so systematisch in den Politikkreislauf einzuspeisen sind, dass es zu einer kontinuierlichen Verbesserung und Wirkungssteigerung politischer Maßnahmen und Entscheidungen kommt (Straßheim, 2018a). Demgegenüber gelten Evaluationen im kritischen Strang der Forschung als inhärent politischer Akt – oder besser eine Verkettung politischer Akte – in denen sich Wertungen und Beurteilungen manifestieren (Bovens, Hart & Kuipers, 2008, p. 1).

Evaluationen weisen in mehrfacher Hinsicht Charakteristika eines politischen Handelns auf, das auf kollektiv verbindliche Ordnung zielt und dazu (Auswahl-)Entscheidungen trifft (Fischer & Gottweis, 2012; Nullmeier & Pritzlaff, 2009): Erstens sind Evaluationen eng mit kollektiv verbindlichen Regelungen und Programmen verzahnt, die erst in der Ausgestaltung des Verfahrens, den Prozeduren und Methoden, den Berichts- und Bewertungskriterien ihre konkrete Umsetzung erfahren. Dabei kommt es immer auch zu Abweichungen, Neubewertungen und Ausnahmetatbeständen. Evaluationen ermöglichen deswegen eine meist mikroskopische, mitunter aber pfadabweichende Rekalibrierung existierender Regel- und Programmbestände, die in sich selbst bereits den Anspruch kollektiver Verbindlichkeit trägt und deswegen auch besonders zäh umstritten sein kann – selbst wenn es scheinbar nur um Kleinigkeiten geht (Vedung, 2008). Zweitens können sich in Kontroversen über Zieldefinitionen und Indikatorenkonstruktionen sehr unterschiedliche Auffassungen über das Verständnis beispielsweise von Arbeitsmärkten, Strukturwandel oder Leistungserbringungen widerspiegeln. Wer sich durchsetzt, bekräftigt in der Konsequenz eine spezifische politische Perspektive und schafft so mögliche Anknüpfungspunkte für die Rechtfertigung (oder Kritik) legislativer oder administrativer Entscheidungen (Fischer, 2003). Drittens definieren Evaluationen mit der Festlegung von Zielgruppen, der Unterteilung von Interventions- und Kontrollgruppen, der Bestimmung einer sozialen Ebene, auf der Wirkungen gemessen werden und der Annahme über Wirkungsbeziehungen und -bedingungen die (Kollektiv-)Subjekte politischen Handelns. Damit verbunden sind oft Zuweisungen von Handlungsverantwortungen. Im Aufstieg der randomisierten Kontrollstudien in der Armuts- und Wohlfahrtsstaatsforschung sehen beispielsweise Kritiker eine Verschiebung politischer Verantwortungszuschreibungen von der strukturellen Ebene auf die Ebene der individuellen Betroffenen (so bereits Straßheim, 2018b und Faust in diesem Band; Weiss, 1993 [1973], p. 105). Evaluationen setzen sich insofern aus Sequenzen von einzelnen Handlungen oder Kommunikationen zusammen, die in wertender Weise

Regelungs-, Kollektivitäts-, Verbindlichkeits- und Konfliktbezüge herstellen, bekräftigen oder mit Akzeptanz versehen. Evaluation ist in diesem Sinne eine inhärent politische Praxis (Nullmeier & Pritzlaff, 2009, S. 13-16).

Wie Schendzielorz und Reinhart in diesem Band anhand von Peer Review-Verfahren zeigen, können Evaluationen durch die sequentielle Verklammerung von Wertungs- und Geltungsbehauptungen eine hohe Akzeptanz und Autorität erreichen. Sie bilden eine Art Scharnier zwischen evaluatorischen und epistemischen Aussagen, indem sie verschiedenartige Praktiken der Wertzuschreibungen und -abwägung in eine Infrastruktur von Evaluationsprozeduren einbetten und mit politisch-administrativer, aber zugleich auch fachlicher Gültigkeit ausstatten. Evaluationen profitieren insofern von der „Legitimation durch Verfahren“ (Luhmann, 1989 [1969]).⁵ Natürlich stehen Evaluationsergebnisse im politischen Prozess unter Vorbehalt, werden angezweifelt oder bestritten. Es gibt auch gute Gründe, anzunehmen, dass der permanente Druck zur Performanzmessung und zum Leistungsvergleich langfristig die Autoritäts- und Vertrauensbasis von Evaluationen in der Öffentlichkeit erodieren lässt (Boswell, 2018; Straßheim & Kettunen, 2014).

Die Fabrikation von Vertrauen ist insofern fragil: Evidenzen können niemals aus sich selbst heraus überzeugen, sondern müssen in Bezug zu einem laufenden Argumentations- und Bewertungsprozess gesetzt werden (Majone, 1989). Akteure verweisen in ihren Einschätzungen aufeinander (oder grenzen sich voneinander ab), werfen ihre Reputation in die Waagschale oder berufen sich auf Professionsstandards. Evaluationsberichte und *Policy Papers* bilden nicht nur eine Entscheidungs- und Bewertungsgrundlage, sondern gewinnen auch ein symbolisches Potential, indem sie für den Abschluss eines im Nachhinein nur schwer rekonstruierbaren Prozesses der Verhandlung von Definitionen, des Streits um Schlussfolgerungen und des Feinschliffs von Formulierungen stehen. Wenn sie dieses symbolische Potential erreichen, dann entwickeln sich bestimmte Evaluationsverfahren wie etwa *Peer Reviews* zu regelrechten Erfolgsgeschichten, die zur Nachahmung anregen und über den konkreten Anwendungskontext hinaus „Meta-Stabilität“ erreichen (Schendzielorz und Reinhart in diesem Band). Auf solche Evaluationstechniken wird dann immer wieder und in sehr unterschiedlichen Kontexten zurückgegriffen und gerade durch ihren symbolischen Gehalt zu einem gewissen Grade gegen Kritik abgesichert; zugleich erfahren sie je nach institutionellen und organisatorischen Rahmenbedingungen eine sehr unterschiedliche Anwendung.

3 Varieties of Evaluation

In bemerkenswerter Übereinstimmung zeichnen die Beiträge dieses Bandes das Bild einer expandierenden Evaluationslandschaft, die sich einerseits durch Professionalisierungstendenzen und Vereinheitlichungen auszeichnet, andererseits jedoch durch eine breite Vielfalt in der Institutionalisierung und Umsetzung von Evaluationsverfahren und -praktiken (insbesondere die Beiträge von Wolfgang Meyer und Reinhard Stockmann wie auch Thomas Widmer).

Hellmut Wollmann und andere sehen die Entwicklung dieser Landschaft durch drei Wellen gekennzeichnet (Stern, 2009; Wollmann, 2003): Schon Ende des 19. Jahrhunderts hatten sich Allianzen zwischen wissenschaftlichen und sozialpolitischen Akteuren gebildet, die dann die Legitimations- und Wissensbasis für spezifische Vorstellun-

gen über die Leistungsfähigkeit, Interventionstiefe und Extensivität von Wohlfahrtsstaaten lieferten (Bulmer, 1991; Wittrock, Wagner, & Wollmann, 1991). Die Expansion des Wohlfahrtsstaates in den 1960er und 1970er Jahren (erste Welle) ist nicht denkbar ohne die massive Pluralisierung solcher Arrangements der Politikberatung und -evaluierung und jener als „epistemic drift“ bezeichneten Bewegung, die wissenschaftliche und politische Qualitäts- und Relevanzkontrollen immer stärker miteinander verzahnte (Elzinga, 1985; Wagner, Hirschhorn Weiss, Wittrock & Wollmann, 1991). Eine zweite Welle von Evaluationen seit dem Ölpreisschock Mitte der 1970er Jahre war vor allem durch Budgetkürzungen und Kostensenkungen motiviert. Die „modelling wars“ zwischen US-Denkfabriken in den späten 1970er und frühen 1980er Jahren waren das Resultat dieser zweiten Expansionswelle. Sie mündeten in den Bemühungen von Experten und Beratungseliten, durch die Bewaffnung mit Mikrosimulationen, Modellierungen und ersten randomisierten Kontrollstudien die Wirksamkeit Kosteneffizienz staatlicher, vor allem sozialstaatlicher Maßnahmen in großem Umfang zu bewerten und so einen Anteil an der „poverty research industry“ zu gewinnen (O'Connor, 2001). Die dritte Welle von Evaluationen in den späten 1980er Jahren im Zuge der *New-Public-Management-Reformen*, seit Ende der 1990er Jahren eng verbunden mit der „evidence based policy“-Bewegung, führte zu einer verstärkten Institutionalisierung und einer systematischen, durch Leitlinien gestützten Verbreitung entsprechender Instrumente (Krapp, Pannowitsch & Heinelt, 2015; Straßheim, 2018a).

Wollmanns historische Analyse wäre aus heutiger Sicht durch eine vierte Welle zu ergänzen, welche seit mindestens drei Dekaden die ersten drei Wellen begleitet, verstärkt und partiell überrollt: Eine zunehmende Zahl inter- und transnational agierender Organisationen, Expertennetzwerke und -gemeinschaften prägen die Expansion von Evaluationen (Adelle & Weiland, 2013; Stephenson, Schoenefeld & Leeuw, 2019; Stockmann, Meyer & Taube, 2020 und Stockmann & Meyer in diesem Band; Stone, 2019; Straßheim, 2017b). Dazu gehört zum einen die Evaluationspolitik der Europäischen Union und hier insbesondere der Kommission. Durch die Implementierung von Evaluationsinstrumenten wie etwa *Impact Assessments* seit 2002, die Entwicklung eines Evaluationssystems zur Wirkungskontrolle, der „Smart Regulation“- und „Better Regulation“-Agenda, durch Expertennetzwerke, Training, Leitlinien und Standards unterstützt, beeinflusst und forciert sie entsprechende Bemühungen in den Mitgliedstaaten (Eußner & Blomeyer, 2020; Stern, 2009). Als besonders einflussreich hinsichtlich der Stärkung von Evaluationsverfahren hat sich die Kohäsionspolitik erwiesen, wo durch die direkte Einbindung nationaler Behörden im Rahmen der geteilten Mittelverwaltung Impulse besonders gut zu vermitteln sind.⁶ Des Weiteren sorgen internationale Organisationen wie die Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) oder die Weltbank in Form von *Peer Reviews* nationaler Evaluationen, durch eigene Leitlinien und Vergleiche guter Praktiken für eine grundsätzlich freiwillige, in ihrer Bindungswirkung jedoch nicht zu unterschätzende Verbreitung von Evaluationen (Morth, 2005; Radaelli, 2004). Hinzu kommen Expertennetzwerke, epistemische Gemeinschaften, Think Tanks, philanthropische Organisationen, Consulting-Firmen, transnationale Rechtsanwaltskanzleien und eine Vielzahl weiterer Organisationen an den Schnittstellen von Politik, Wissenschaft, Wirtschaft und Zivilgesellschaft (Botzem, 2012; Quack, 2016; Stone, 2012, 2019; Straßheim & Loer, 2019). Mit ihnen formiert sich eine „globale Agora“ (Stone, 2012) der Evaluationspolitik, die in ihren ebenenübergreifenden Zusammenhängen und Auswirkungen derzeit noch zu wenig

verstanden ist. Zuletzt haben Arno Simons und Jan-Peter Voss (Béland, Howlett, & Mukherjee, 2018; Simons & Voss, 2018) darauf hingewiesen, dass zunehmend auf der inter- und transnationalen Ebene mit Akteurskonstellationen (sogenannten „instrument constituencies“) zu rechnen ist, die in der Produktion, Fabrikation und Verbreitung von Evaluationsinstrumenten ihre kollektive Aufgabe sehen – und dies weniger, um Lösungen für Probleme bereitzustellen, als vielmehr im Sinne John Kingdons (1984) mit dem Ziel, für die von ihnen entwickelten Lösungen passende Probleme zu finden.

Die sich herausbildende Evaluationslandschaft ist divers und umfasst sehr verschiedenartige, voneinander teilweise abgrenzbare, teilweise einander beeinflussende und sich überschneidende Arrangements. Grundsätzlich lassen sich drei Perspektiven der Analyse und des Vergleichs dieser Varieties of Evaluation unterscheiden (Leeuw & Furubo, 2008; Schwab, 2017; Stephenson, Schoenefeld & Leeuw, 2019):

(1) Ansätze des Vergleichs von Regierungs- und Governancesystemen betonen die prägende Kraft politisch-administrativer Institutionen, die Rolle föderaler Ordnungen für die Einheitlichkeit von Evaluationen, den Einfluss des Wahlsystems auf evaluationspolitisch relevante Akteurskonstellationen und Politikstile wie auch die Verwaltungs- und Rechtstraditionen (Berg-Schlosser & Müller-Rommel, 1991; Murswieck, 1994; Rueschemeyer & Skocpol, 1996; Welch & Thompson, 1980). In einer policyanalytischen Variante dieses Ansatzes treten die strukturellen Fähigkeiten und Kapazitäten („policy analytical capacity“) von Regierungen und Verwaltungen in den Blick (Howlett, 2009). Diese umfassen die Fähigkeiten von Regierungen, Evaluationsforschung und eigene Evaluationskapazitäten etwa in Ressortforschungseinrichtungen bereitzustellen, komplexere statistische Methoden und Evaluationsinstrumente zu entwickeln und eine entsprechende Professionalisierung zu ermöglichen, Interessengruppen und andere politische Akteure in Evaluationsprozeduren einzubeziehen, die Ergebnisse in Form eigener Leitlinien zu reflektieren, sie innerhalb des politisch-administrativen Systems und in unterschiedlichen Ressorts zu implementieren, zu kontrollieren und letztlich auch in der Öffentlichkeit zu kommunizieren. Da diese strukturellen Merkmale in den verschiedenen Politikfeldern unterschiedlich ausgeprägt sind, erklärt dies die Entstehung und Differenzierung sehr verschiedener Evaluierungsarrangements innerhalb nationaler Ordnungen.

Die Ergebnisse der Analysen von Wolfgang Meyer und Reinhard Stockmann wie auch von Thomas Widmer in diesem Band sind auch in dieser Hinsicht äußerst aufschlussreich, weil sie sowohl auf international vergleichender wie nationaler bzw. subnationaler Ebene den Einfluss von Regierungssystemen stark relativieren. Dies lässt sich anhand des Schweizer Paradoxons (siehe Widmer in diesem Band) illustrieren, einem System, dem aufgrund der fragmentierten Staatsstrukturen, des Milizsystems und der Rückkopplungsfunktion der direkten Demokratie traditionell eine geringe Evaluierungskapazität und ein ebenso geringer Evaluierungsbedarf zugeschrieben wurde – und das dennoch als Evaluationspionier einen ungewöhnlich hohen Grad der Institutionalisierung und Implementierung von Evaluationen aufweist.

(2) Ansätze der vergleichenden Analyse von Akteurskonstellationen und -gemeinschaften sind konzeptuell ähnlich angelegt wie Forschung zu epistemischen Gemeinschaften und advokatorischen Koalitionen (Haas, 2013; Leeuw & Furubo, 2008; Sabatier & Weible, 2014; Wittrock, Wagner & Wollmann, 1991). Jan-Eric Furubo und Frans Leeuw gehen in ihrem Konzept der Evaluationssysteme davon aus, dass Evaluierungsar-

rangements sich genau dann ausdifferenzieren, wenn (a) Regierungs- oder Nichtregierungsorganisationen andere Organisationen mit der Umsetzung von Evaluationen beauftragen oder sie zumindest darin fördern, (b) sich in diesem Zuge ein kollektiv geteiltes Glaubenssystem („shared epistemology“) über die Gründe, Formen und Ziele von Evaluationen stabilisiert, (c) diese Allianz auf Dauer angelegt ist und (d) sich ihre Aktivitäten auf die Nutzung der Ergebnisse und Informationen aus Evaluationen in politischen Entscheidungs- und Implementationsprozessen beziehen (Leeuw & Furubo, 2008, pp. 159-160). Im Ergebnis konfigurieren sich aus dieser Sicht die *Varieties of Evaluation* also nicht entlang von Räumen oder Systemen des Regierens, sondern entlang von epistemisch vergemeinschafteten Akteurskonstellationen. Dementsprechend lassen sich Evaluationssysteme durch ihre paradigmatische, auf gemeinsamen Annahmen beruhende Verknüpfung spezifischer Verfahren, Instrumente und Akteure voneinander abgrenzen.

Systeme der Performanzmessung etwa sind eng mit der Entwicklung des *New Public Managements* verbunden, werden auf nationaler Ebene durch Ministerien und Audit-Organisationen gefördert und auf internationaler Ebene durch die OECD unterstützt. Dabei spielt die Verdichtung auf ausgewählte Kennzahlen und die damit einhergehende Verringerung und teilweise Ausblendung von Komplexitäten eine wichtige Rolle. Das dahinter stehende Glaubenssystem verbindet Annahmen über die anreizfördernde Wirkung von Performanzanalysen und -vergleichen mit der Behauptung, dass politische und gesellschaftliche Akteure die daraus resultierenden Informationen tatsächlich zur Optimierung von Entscheidungen nutzen und dabei nichtintendierte Nebeneffekte prinzipiell kontrollierbar sind (Bjørnholt & Larsen, 2014; Leeuw & Furubo, 2008, p. 161). In einem gewissen Kontrast dazu stehen Systeme der (quasi-)experimentellen Evaluation, die randomisierte Kontrollstudien als „Goldstandard“ betrachten und in einem mittlerweile weltumspannenden Netzwerk von Expertenorganisationen wie der Campbell Collaboration oder dem Abdul Latif Jameel Poverty Action Lab die Wirkungsanalyse von Interventionen auf der Mikroebene betreiben (Donovan, 2018 und Faust in diesem Band): Sie fokussieren stark auf Verhaltensweisen und Mechanismen, die auf Ebene der einzelnen Untersuchungseinheiten greifbar werden. Mit dieser Fokussierung blenden sie noch stärker als Systeme der Performanzmessung Komplexitäten und Zusammenhänge aus. Der Kontrast zu Systemen der Performanzmessung wird dann stärker, wenn die RCT-Bewegung sich mit verhaltensökonomischen Expertenzirkeln verbindet und auf dieser Grundlage Annahmen über rationales, informiertes Entscheiden durch experimentelle Erkenntnisse über verzerrte Rationalitäten ersetzt werden (Straßheim, 2017a). Petra Kaps zeichnet in ihrem Beitrag für diesen Band nach, wie sich im Zuge der Hartz-Reformen advokatorische Koalitionen der experimentellen Politikgestaltung herausbilden. Dabei hängt die Ausprägung dieser Koalitionen und ihre Durchsetzungsfähigkeit in Deutungskämpfen besonders von der institutionellen Ausgestaltung des Evaluationsverfahrens, der Umstrittenheit des jeweiligen Deutungssystems wie auch vom Organisationsgrad und der Machtposition der beteiligten Akteure ab.

(3) Schließlich konzentrieren sich Ansätze der Evaluationskulturen auf die mehr oder weniger unhinterfragten Praktiken und Prozeduren, die damit verbundenen impliziten Normen, Prämissen und Wertordnungen und die darauf beruhenden Infrastrukturen und Institutionen des Evaluierens (vgl. für unterschiedliche Herangehensweisen Wid-

mer und Schendzielorz & Reinhart in diesem Band). Ortwin Renn (1995) hatte bereits Mitte der 1990er Jahre darauf hingewiesen, dass sich Länder wie die USA, Großbritannien oder Deutschland in ihren Stilen der Politikberatung und -evaluation stark unterscheiden. In seiner Klassifikation unterscheidet er zwischen einem „adversarial“ bzw. „pluralist style“ der öffentlichen Wissensproduktion und Politikbewertung wie in den USA, einem „embodied“ bzw. „service-based style“, der wie in Großbritannien durch ministerielle „in house“-Einrichtungen und enge Konsultationsbeziehungen geprägt ist, und einem „corporatist style“ hochgradig institutionalisierter Beziehungen zwischen verschiedenen gesellschaftlichen Interessengruppen, die in Enquetekommissionen, parlamentarischen Ausschüssen und Verhandlungsrunden zu Erkenntnissen und Bewertungen gelangen (z. B. Deutschland, Schweden, teilweise die Niederlande). Daran anschließend argumentiert Sheila Jasanoff (2005; 2011), dass Arrangements der Expertise und Evidenzproduktion in engem Zusammenhang mit (national-) spezifischen Epistemologien stehen – institutionalisierten und weitgehend selbstverständlichen Praktiken der Produktion öffentlich relevanten Wissens, die aus dem historischen Zusammenspiel sozio-technischer, wissenschaftlicher und politischer Kulturen hervorgegangen sind und mit deren Hilfe die Mitglieder einer Gesellschaft bestimmen, ob bislang ungeprüfte Geltungsansprüche im Prozess kollektiv bindender Entscheidungen Berücksichtigung finden sollen (Straßheim, 2013).

Die vergleichende Analyse von Evaluationskulturen und Epistemologien kann erheblich zum Verständnis jener Selbstverständlichkeiten beitragen, mit denen ganz bestimmte Modi der Evaluation in einigen Kontexten öffentliche Akzeptanz finden, in anderen Kontexten dagegen nicht. Das lässt sich beispielsweise gut an der Debatte über experimentelle und verhaltensökonomisch informierte Politiken nachvollziehen, die in Großbritannien und auch in den USA Zustimmung erfahren, während sie in Deutschland auf heftigen Widerstand in einer breiten Öffentlichkeit stoßen (Straßheim, Jung, & Korinek, 2015; für den Bereich der Entwicklungspolitik den Beitrag von Faust in diesem Band). Und dennoch haben Typologien wie die von Renn oder Jasanoff nur begrenzte Aussagekraft, weil sie gleichsam gefrorene Wissenslandschaften präsentieren. Das zeigen die Beiträge in diesem Band übereinstimmend ganz deutlich: Die *Varieties of Evaluation* unterliegen einem hochgradig dynamischen Wandel, der auch für interne Differenzierungen und Konflikte sorgt. So kommen Meyer und Stockmann zu dem Schluss, dass die Ausprägung von Evaluationen weniger durch die institutionellen Rahmenbedingungen der politischen Systeme oder strukturelle Besonderheiten gekennzeichnet seien als vielmehr durch Entwicklungsprozesse und Reformen in der Regierungspraxis – eine durch Wandel und Diversität gekennzeichnete Situation, die auch Widmer für die Schweiz konstatiert. Deutlich einflussreicher scheinen – so zeigen auch die Beiträge von Widmer und Kaps – politikfeldspezifische Unterschiede und Akteurskonstellationen zu sein.

Die drei genannten Ansätze der *Varieties of Evaluation* schließen einander nicht aus, sie bedürfen aber der wechselseitigen Befruchtung und der konzeptuellen Weiterentwicklung – insbesondere unter den Bedingungen der „postnationalen Konstellation“ (Habermas, 2001). Praxistheorien und Ansätze der neueren Wertungs- und Bewertungssoziologie erscheinen hier als besonders vielversprechend, um Evaluationspolitiken in ihrer Dynamik und Vielgestaltigkeit zu erfassen (Lamont, 2012; Ringel & Werron, 2019; Schendzielorz & Reinhart in diesem Band).

4 Die Beiträge

Wolfgang Meyer und *Reinhard Stockmann* haben in einer umfassenden Bestandsaufnahme die Institutionalisierung von Evaluierung in Europa untersucht. Sie analysieren die formale Verankerung in Legislative und Exekutive sowie die Nutzung von Evaluationsergebnissen. Die institutionelle Verankerung wie auch die Nutzung von Evaluierungen variiert europaweit stark, ohne dass sich ein einheitliches Entwicklungsmodell herauskristallisieren würde. Insgesamt ist die Institutionalisierung in den Niederlanden und der Schweiz am weitesten fortgeschritten. Für die Institutionalisierung der Evaluierung in den nationalen Systemen spielen deren Strukturmerkmale nur eine geringe Rolle. Wichtiger ist die Regierungspraxis und hier vor allem das Ausmaß, in dem die Regierungen Ideen aus *New Public Management* und evidenzbasierter Politik aufgreifen. Die Parlamente treten als Treiber der Entwicklung und Nutzer der Evaluierung nur im Ausnahmefall prominent in Erscheinung. Nicht für die Institutionalisierung, aber für die Professionalisierung im Bereich Evaluierung ist außerdem die Größe des Landes von Bedeutung. Größe begünstigt die Entwicklung von Fachcommunities. Übergreifend treten Trends zur Verbreitung und Professionalisierung, aber auch teils nachlassende Dynamik gerade in den Ländern mit relativ weit entwickelten Evaluierungssystemen und eine Art Ausfransung von Evaluierungsverfahren in Richtung Controlling und Audit sowie eine generelle Entwertung von Evidenz zutage.

Wo Wolfgang Meyer und Reinhard Stockmann ein breites Spektrum im Ländervergleich in den Blick nehmen, betrachtet *Thomas Widmer* den Fall der Schweiz in der Tiefe und fragt auf Grundlage der Ergebnisse eines umfassenden Forschungsprogramms nach Wechselwirkungen zwischen Politik und Evaluierung. Die Verankerung der Evaluationsfunktion ist in der Schweiz auf Bundesebene stärker als auf Kantonsebene, wo sich vor allem Unterschiede zwischen Politikfeldern, weniger zwischen Kantonen zeigen. Hinter der hohen Zahl formal verankerter Evaluationsklauseln stehen neben regionalen Differenzierungen auch inhaltlich unterschiedliche Ausgestaltungen. In den für die Schweiz typischen Verfahren direkter Demokratie scheinen Evaluationsergebnisse insgesamt wenig Eingang zu finden; sie können sich aber auf die Diskursqualität auswirken, wenn auch eher über Expertinnen als über die breite Bevölkerung. Die parlamentarische Nutzung ist deutlich erkennbar und geht über die reine Kontrollfunktion hinaus, ist aber insgesamt wenig systematisch. Evaluationsprozesse werden eher genutzt als Evaluationsergebnisse. Im Hinblick auf die Rolle institutioneller Faktoren bestätigt die Analyse der Schweiz im Wesentlichen den internationalen Vergleich: Institutionen können nur in Teilen das Bild erklären, das sich insbesondere hinsichtlich der Praxis von Evaluierungsarrangements zeigt. Auch wenn zu keinem anderen Land so systematische Analysen zur Entwicklung der Evaluierungskultur vorliegen, tun sich hier weiter vielfältige Forschungsfragen auf. Widmers Apell, die Heterogenitäten und insbesondere die Erklärungsfaktoren für die Herausbildung unterschiedlicher Evaluationskulturen besser zu verstehen, kann somit auch für die internationale Perspektive gelten.

Mit *Jörg Fausts* Beitrag zur Bedeutung rigoroser Wirkungsevaluierung in der Entwicklungszusammenarbeit wird eine sektorale Perspektive eingenommen. Damit treten auch die Besonderheiten des konkreten Politikfeldes in den Vordergrund – dort haben Meyer und Stockmann sowie Widmer wichtige Einflussfaktoren für die Gestaltung von Evaluierungsarrangements vermutet. Faust zeigt, dass einerseits grundlegende Eigenschaften der Entwicklungszusammenarbeit die Rolle von Evaluierungen in diesem Po-

litikfeld stärken und das sie dort in besonderem Maße ihre Erkenntnis-, Lern- und Kontrollfunktion zeigen kann. An den Auseinandersetzungen um die Rolle rigoroser Wirkungsevaluierungen, die Faust strukturiert nachzeichnet, wird die Rolle von Einflussfaktoren stärker konturiert, die bereits in den vorangegangenen Beiträgen anklangen: Verschiedene inhaltliche Diskussionslinien entlang derer der Stellenwert rigoroser Wirkungsevaluierungen unterschiedlich bewertet wird, treten zu Tage. Aber auch die Rolle verschiedener Akteursgruppen, insbesondere auch der internationalen Geberorganisationen sowie die Abhängigkeit von Eigenschaften des Evaluierungsgegenstandes und dem Evaluierungsansatz werden neben weiteren thematisiert.

Petra Kaps konzentriert sich in ihrem Beitrag auf Experimente in der deutschen Arbeitsmarktpolitik und deren Evaluierung. Anhand der Kombilohn-Experimente (1998-2005), der Experimente zur Pauschalierung von Sozialhilfeleistungen (1999-2004) und des Experiments zur Erprobung von Eingliederungsmodellen (2005-2008) geht sie den *Varieties of Evaluation* nach. Sie verfolgt dabei eine auf Akteurskonstellationen bezogene Perspektive: Auf der Grundlage des Advokaten-Koalitionen-Ansatzes lotet sie insbesondere die politischen Deutungskämpfe und den Einfluss der institutionellen Verfahrensgestaltung aus. In der Tat kann Kaps nachweisen, dass die institutionelle Einbettung – die Dezentralisierung des Evaluationsauftrags, der Grad der Institutionalisierung des Evaluationsauftrags, aber auch die Eigenschaften des zu evaluierenden Programmes und die Verfügbarkeit von Datenquellen – einen Einfluss auf die Einheitlichkeit und interne Abstimmung von Evaluierungsarrangements haben. Zudem unterliegen die Advokaten-Koalitionen im Evaluierungsprozess unterschiedlich heftigen politischen Deutungskämpfen, die jeweils von der institutionellen Ebene, der Umstrittenheit des jeweiligen Politikprogrammes sowie dem Organisationsgrad und der Machtposition der handelnden Akteure abhängen.

Aus einer bewertungssoziologischen Perspektive erschließen *Cornelia Schendzielorz* und *Martin Reinhart* jene spezifischen Praktiken, die in Peer-Review-Verfahren zur Stabilisierung von *Varieties of Evaluation* führen. Zudem leuchten sie die politische Dimension dieser Arrangements aus, indem sie die Steuerungs- und Legitimationsfunktion dieser Verfahren rekonstruieren und damit auch die wissenschaftliche (Selbst-)Regierung durch *Peer Reviews*. Die empirische Grundlage bildet eine fallvergleichende Analyse von *Peer Reviews* im Rahmen der Förderung von Sonderforschungsbereichen (SFB) durch die DFG, im Rahmen von Zeitschriften und im Rahmen von Berufungsverfahren. Als Zwischenbefund identifizieren Schendzielorz und Reinhart typische Bewertungspraktiken, die als elementare Bestandteile von *Peer Reviews* sowohl die Binnendifferenzierung als auch prozedurale Anpassungsfähigkeit von *Peer Reviews* ausmachen. *Peer Reviews* entwickeln „Meta-Stabilität“, gerade weil sie Unsicherheiten und auch die Vielfalt von Personenkonstellationen, Arbeitsteilung und Rollentrennung durch Prozessualität und Sequenzialität kanalisieren und reduzieren – all das unter einem symbolisch vereinheitlichenden Label. Im Zusammenhang damit zeigen Schendzielorz und Reinhart auch, dass die in diesen Verfahren ablaufenden, mitunter widersprüchlichen und nicht kohärenten Kriterien zuzuordnenden Bewertungspraktiken sich auf die Vereinheitlichung berufen können, die *Peer Reviews* gleichsam als die „Stimme der Wissenschaft“ leisten. So ermöglichen *Peer Reviews* Adaption wie auch Legitimation nach innen wie außen. Sie arbeiten wie ein Scharnier, das „wissenschaftliche und politische Qualitäts- und Legitimationserfordernisse ineinander überführt und füreinander kompatibel macht“ (Schendzielorz und Reinhart).

Diese politische Multifunktionalität von Evaluationen adressiert auch *Regine Paul* in ihrem konzeptuellen Beitrag zu Ex-ante-Evaluationen, der dieses Themenheft abschließt. Darin entwickelt sie in intensiver Auseinandersetzung mit der Policy-Forschungsliteratur eine interpretative Perspektive, die Evaluationen als institutionell eingebettete Instrumente begreift, die auf selektive Weise Deutungshorizonte nicht nur erschließen, sondern politische Deutungen und Erwartungskonstellationen überhaupt erst ermöglichen. Im Zentrum von Pauls Beitrag steht eine Typologie, die idealtypisch drei verschiedene, mit Ex-ante Evaluationen verbundene Logiken ausweist und damit unterschiedliche Thesen der Literatur zu Policy-Instrumenten miteinander verknüpft: Unter Umständen kann dasselbe Instrumentarium je nach Kontextbedingungen instrumentelle, legitimatorische oder machbezogene Funktionen übernehmen. Paul führt damit nicht nur Literaturen unterschiedlicher Provenienz zusammen. Vielmehr zeigt sich gerade in den Kombinationen scheinbar gegenläufiger Rationalitäten die Vielgestaltigkeit der Evaluationspolitiken.

5 Evaluationspolitik auf der Vorder- und Hinterbühne

Überblickt man die hier versammelten Beiträge, so erscheinen Evaluationen in der Tat als eine in sich politische Praxis, die durch Widersprüchlichkeiten und Differenzierungsbewegungen gekennzeichnet ist und doch ein erhebliches Expansionspotential aufweist. Evaluationen sind deswegen inhärent politisch, weil mit ihnen Behauptungen über kollektive Verbindlichkeit und durchaus nicht unumstrittene, potentiell konflikt-hafte und immer kontingente, auch anders mögliche Wertungen verbunden sind. Akzeptiert man diese dem vorliegenden Themenschwerpunkt zugrunde liegende These, dann kann man behelfsmäßig und sicher nur vorläufig zu einer Systematisierung von Evaluationspolitik in ihren Dimensionierungen und Erscheinungsformen gelangen. Diese Systematisierung kann je nach Theorie der Politik und des Politischen auch anders ausfallen (Bedorf & Röttgers, 2010; Marchart, 2010; Nullmeier & Pritzlaff, 2009; Zürn, 2018).

Wir gehen hier davon aus, dass Politik mindestens zwei Dimensionen aufweist: Sie erschließt immer einen Raum des auch anders Möglichen, trifft explizit oder implizit eine Auswahl und ist insofern zumindest potentiell umstritten und konflikt-haft. Dort, wo Notwendigkeitskonstruktionen und Automatismen den Diskurs bestimmen, verschwindet Politik (Greven, 2010). Zum anderen realisiert sich Politik in einer Öffentlichkeit, die in sich hochgradig divers sein kann und doch zumindest vorläufig durch Politik als ein Kollektivsubjekt des gemeinsamen Handelns und Kommunizierens erfahrbar wird (Arendt, 2017 [1950]; Rancière, 2002). Je begrenzter diese Öffentlichkeit und das potentiell relevante Akteursspektrum, desto geringer der Spielraum für Politik. Im Ergebnis postulieren wir zwei Dimensionen – Konfliktpotential und Akteursspektrum –, die uns Aufschluss über mögliche Erscheinungsformen von Evaluationspolitik geben (siehe auch Zürn, 2013).

Tabelle 1: Erscheinungsformen der Evaluationspolitik

		Akteursspektrum	
		<i>breit</i>	<i>eng</i>
Konfliktpotential	<i>hoch</i>	(1) Evaluationspolitik auf der Vorderbühne	(2) Evaluationspolitik auf der Hinterbühne
	<i>niedrig</i>	(3) Symbolische/simulative Evaluationspolitik	(4) Entpolitisierung von und durch Evaluationen

Quelle: Eigene Darstellung.

In *Tabelle 1* sind diese Erscheinungsformen abgetragen. Wir unterscheiden zwischen (1) Evaluationspolitik auf der Vorderbühne, (2) Evaluationspolitik auf der Hinterbühne, (3) symbolischer bzw. simulativer Evaluationspolitik und (4) einer Entpolitisierung der Evaluationspolitik:

(1) Evaluationspolitik findet auf der Vorderbühne statt, wenn wissenschaftliche Expertinnen, Akteure aus Parlament, Regierungen und Ministerien, zivilgesellschaftliche und wirtschaftliche Akteure öffentlich um gemeinsame Bewertungsgrundlagen und -kriterien ringen und die Konsequenzen von Evaluationsergebnissen für kollektiv verbindliche Entscheidungen ausloten (Bovens, Hart & Kuipers, 2008; zur Metapher der Vorder- und Hinterbühne Goffman, 1959; Hilgartner, 2000). Solche im eigentlichen Sinne politischen Evaluationen sind mitunter konfliktreich und lösen häufig Kaskaden von Folge- und auch Gegenevaluationen aus (Bovens, Hart & Kuipers, 2008). Als „öffentliches Drama“ (Hilgartner, 2000) liefern sie wissenschaftlichen und politischen Akteuren eine Bühne für den Gewinn (oder Verlust) von Autorität und Reputation, ermöglichen Deliberation und Austausch über fundamentale Werte und Normen (Fischer, 1999) – sind aber auch der Ort für Machtasymmetrien, Verantwortungsverschiebungen und „blame games“ (Hood, 2007). Insofern handelt es sich bei Evaluationspolitik auf der Vorderbühne auch keineswegs um ein Ideal herrschaftsfreier Kommunikation, sondern eher um einen Spielraum für vielerlei Varianten politischer Praxis.

(2) Evaluationspolitik auf der Hinterbühne betrifft demgegenüber nur einen schmalen Akteurskreis im Arkanbereich parlamentarischer und administrativer Verhandlungsarenen, in Expertennetzwerken und eher geschlossenen Akteurskonstellationen (Bovens, Hart & Kuipers, 2008). Diese Form der Evaluationspolitik unterliegt möglicherweise den Evaluationszyklen und -rhythmen von Politikprogrammen, sie kann aber auch im Zuge politischer Krisen, der Untersuchung möglicher Fehlentscheidungen, Neuwahlen oder Rücktritten auftreten und ist insofern potentiell konfliktreich. Im Grunde ist sie der Normalfall bei Verhandlungen über Indikatoren und Verfahrensstandards; insofern handelt es sich auch um eine Notwendigkeit effektiver Evaluationspraxis und eine unweigerliche Begleiterscheinung der Professionalisierung und Institutionalisierung von Evaluationen. Einige Befunde, die auf die starken Unterschiede von Evaluierungsarrangements zwischen Politikfeldern hinweisen (siehe Meyer und Stockmann sowie Widmer in diesem Band) legen die Vermutung nahe, dass die Ausgestaltung der Evaluierungsarrangements nicht nur von Seiten der Politik und Verwaltung, sondern auch von den jeweils mit den *Policies* verbundenen wissenschaftlichen Fachgemeinschaften

getrieben ist. Die Professionalisierung kann auch in technokratische Strukturen umschlagen (Fischer, 1990; Straßheim & Kettunen, 2014). Dies führt dann zu Wissensmonopolen, politischen Schematismen und mit großer Wahrscheinlichkeit zu jenen Formen des Politikversagens, die James Scott (Scott, 1998) in seiner Analyse von Regierungsprogrammen eindrucksvoll geschildert hat. Widmer betont in diesem Zusammenhang die Rolle von evaluationsfachlicher Reflexion in Ergänzung zu fachpolitischen Wissensbeständen.

(3) Evaluationen können für die politische Kommunikation durchaus eine wichtige symbolische Funktion ausfüllen, weil sie im Bestfall Orte des kollektiven Austauschs und der deliberativen Reflexion demokratischer und politischer Wertgrundlagen sind – selbst wenn die Ergebnisse nicht unmittelbar umsetzbar sind (Fischer, 2009). Wenn Evaluationspolitik jedoch zur „Evaluitis“ (siehe Meyer & Stockmann in diesem Band) und damit zu einem Ritual verkommt, das in kulissenhafter Weise lediglich die Rechtfertigung von Politikprogrammen sicherstellen soll, kann man von *symbolischer* Evaluationspolitik sprechen (im Sinne von Edelman, 1964). Evaluationen werden dann zu Maschinerien der Fabrikation von Vertrauen, wenn Rankings, Benchmarkings und Performanzmessungen in alle denkbaren Bereiche des öffentlichen Lebens ausgedehnt werden. In einer solchen „Audit Society“ (Power, 1997) findet schleichend eine Entfremdung statt, weil in paradoxer Weise immer mehr Menschen den Technologien der Bewertung ausgesetzt sind, ohne abweichende Interessen oder Vorstellungen artikulieren zu können (Boswell, 2018). Evaluationen, die tatsächlich politisch folgenreich sind, finden dann unter Umständen lediglich auf der Hinterbühne statt, die sorgfältig von der Öffentlichkeit abgeschirmt ist. Im Gegensatz dazu überschreitet *simulative* Evaluationspolitik die Grenzen zwischen Vorder- und Hinterbühne und wirkt kollektivierend (Blühdorn, 2013). Hier gibt es kein Hinterbünnenspiel mehr. Längst haben Bürger wie Eliten die Erwartungen an eine gemeinsame Wertgrundlage verloren. Evaluationen dienen nur noch der Simulation eines funktionierenden Gemeinwesens – ein postdemokratisches Spektakel, das auf allseitige Zustimmung und Beteiligung rechnen kann, weil – wie dies Ingolfur Blühdorn, Felix Butzlaff, Michael Deflorian, Mirijam Mock und Daniel Hausknost (2020) jüngst anhand der Umwelt- und Klimapolitik nachgewiesen haben – damit der wohlthuende Eindruck kollektiver Handlungsfähigkeit erzeugt werden kann, ohne die Prämissen (aber auch Paradoxien) einer spätkapitalistischen Konsumgesellschaft aufgeben zu müssen.

(4) Wenn Evaluationen weder eine Öffentlichkeit erreichen noch in irgendeiner Weise Anknüpfungspunkte für die Suche nach politischen Alternativen bieten, dann unterliegen sie Prozessen der Entpolitisierung. Ihre Ergebnisse bleiben marginal, sie laufen als Beiwerk eines behördlich-administrativen Apparats ab oder werden nur selektiv wahrgenommen (zur Nutzung von Evaluationsergebnissen Widmer und Faust in diesem Band). Evaluationen können jedoch auch selbst zu Mechanismen der Mobilisierung gesellschaftlicher Entpolitisierungsdynamiken werden. Ihre Ergebnisse werden in solchen Fällen zur Verleugnung von Alternativen, zur Konfliktreduktion und zur Behauptung von Entscheidungsnotwendigkeiten genutzt (Greven, 2010). Dies mag fallweise immer wieder vorkommen und es lässt sich vermuten, dass Entpolitisierungsstrategien in Politikfeldern mit einem besonders hohen Legitimationsdruck häufiger zu beobachten sind (Fischer, 2009). In den vergangenen Jahren wurden im Zeichen einer instrumentellen Regierungstheorie jedoch auch systematische Überlegungen angestellt, die auf eine

weitreichende Entpolitisierung und Entdemokratisierung politischer Systeme im Zeichen einer Stärkung der Wissenseiten hinauslaufen (Brennan, 2016; Min, 2015; Straßheim, 2020). In „Epistokratien“ sollen die Schwächen der Demokratie durch monopolisierte Bewertungs- und Leistungsanalysen ausgeglichen werden. Befürworter schlagen vor, Expertenkommissionen in bestimmten Bereichen ein Vetorecht gegenüber parlamentarischen Entscheidungen einzuräumen („epistocratic veto“), das Wahlrecht auf jene zu beschränken, die in Sachfragen ausreichend informiert sind („restricted suffrage“) oder Wählerstimmen je nach individuellem Abschneiden in einem Wissenstest zu gewichten („weighted voting“) (Brennan, 2016). Auch wenn solche Überlegungen wohl wenig Umsetzungschancen haben, zeigen sie doch, dass die Entpolitisierung von und durch Evaluationen gravierende Folgen haben kann. Es bleibt weiteren Forschungen vorbehalten, mögliche Zusammenhänge zwischen solchen Dynamiken der Entpolitisierung und Entdemokratisierung einerseits und populistischen Tendenzen (Meyer & Stockmann in diesem Band) andererseits auszuloten.

Welche Konsequenzen haben diese Beobachtungen für die Evaluationspraxis? Frank Fischer (1999) hat vor mehr als zwanzig Jahren ein deliberatives Evaluationssystem entwickelt, das die öffentliche und konfliktorische Reichweite von Evaluationen stufenweise erhöht und immer auch die Reflexion auf die zugrundeliegenden politischen Ideen und kollektive Legitimation erzwingt. Mark Bovens, Paul t’Hart und Sanneke Kuipers (2008) suchen nach Wegen der Vermittlung zwischen programmatischen, effektivitätsbezogenen Evaluationen und politischen, die normativen Grundlagen betreffenden Evaluationen. In beiden Fällen geht es um die Vermeidung von Entpolitisierungen und um die Sicherung einer Pluralität von Kriterien des Evaluierens in einer möglichst breiten Öffentlichkeit (Majone, 1989, p. 183). Möglicherweise muss die Frage ähnlich beantwortet werden wie es jüngst Julia Hamann, David Kaldewey und Julia Schubert in einem preisgekrönten Aufsatz zur Relevanz von Wissenschaft getan haben (Hamann, Kaldewey & Schubert, 2019). Es gehe letztlich nicht um standardisierte Bewertungsverfahren, sondern um einen sich beständig ausweitenden „Pool“ von Bewertungsperspektiven, aus dem je nach Situation und Position ausgewählt werden könne – „Autonomie wäre dann, kurz gesagt, die Freiheit zur Heteronomie“ (Hamann, Kaldewey & Schubert, 2019, S. 27). In der Unübersichtlichkeit der Evaluationslandschaft liegt insofern auch eine Chance. Die vergleichende Erforschung der *Varieties of Evaluation* könnte, einem Observatorium ähnlich, diese Landschaft für die Praxis erschließen, die Vielfalt der Evaluationspolitiken sichtbar machen und so Autonomieresourcen bereitstellen (Haus, Korinek, & Straßheim, 2018). Dieser Band will dazu einen Beitrag leisten.

Anmerkungen

- 1 Unter Arrangements verstehen wir hier ein Bündel von Aspekten, welche die Evaluierungspraxis in einem abgrenzbaren Gegenstandsbereich charakterisieren: Diese Arrangements umfassen institutionelle Aspekte, Akteursstrukturen, aber auch Präferenzen für bestimmte Ansätze und Methoden sowie epistemologische Grundannahmen. Unterschiedliche Perspektiven zur Analyse dieser Arrangements stellen wir Teil 3 dieser Einleitung vor.
- 2 Die Positionierung zwischen Wissenschaft und Politik/Verwaltung kommt auch darin zum Ausdruck, dass in den Standards für Evaluationen neben Genauigkeits- und Durchführbarkeitsstandards auch jeweils ein Block von Standards zu Fairness und Nützlichkeit definiert wird.

- 3 Die Beiträge gehen auf die zweitägige Konferenz „Politikevaluation und Evaluationspolitik“ am Wissenschaftszentrum Berlin für Sozialforschung am 17. und 18. November 2018 zurück. Wir danken den Teilnehmerinnen und Teilnehmern für rege Diskussionen und den Kommentatorinnen und Kommentatoren sowie den Gutachterinnen und Gutachtern für zahlreiche Anregungen. Unser Dank gilt zudem der Stiftung Mercator für ihre großzügige Unterstützung.
- 4 Aus Gründen der besseren Lesbarkeit wird im Folgenden alternierend die männliche und weibliche Sprachform verwendet. Selbstverständlich sind alle Geschlechter gleichermaßen gemeint.
- 5 Was auch dazu führt, dass die Behauptung, „evaluiert“ zu haben, zur Legitimation politisch nutzbar ist, ohne dass notwendigerweise tatsächlich Prozesse durchlaufen werden, die systematisch den Wert oder Nutzen eines Gegenstandes analysieren würden. Letztlich kann der Evaluationsbegriff hier politisch als Symbol zur Legitimationsbeschaffung verwendet werden (siehe auch den Beitrag von Paul in diesem Band).
- 6 Während ansonsten Evaluierungspraktiken stark mit den originär europäischen Politiken, wie etwa der Innovationspolitik verknüpft waren, hat in den letzten Jahren eine regulativ vermittelte Evaluierungsverpflichtung auch in anderen Politikfeldern, wie etwa der europäischen Beihilfenpolitik zu einer primärrechtlich gefassten und damit unmittelbar EU-weit geltenden Evaluierungspflicht geführt.

Literatur

- Adelle, Camilla, Jordan, Andrew & Turnpenny, John (2012). Proceeding in parallel or drifting apart? A systematic review of policy appraisal research and practices. *Environment and Planning C: Government and Policy*, 30 (3), 401-415.
- Adelle, Camilla & Weiland, Sabine (2013). Policy assessment: the state of the art. *Impact assessment and project appraisal*, 30 (1), 25-33.
- Arendt, Hannah (2017 [1950]). Was ist Politik? In Ursula Ludz (Hrsg.), *Hannah Arendt: Was ist Politik? Fragmente aus dem Nachlass* (S. 9-12). München/Berlin: Piper.
- Beck, Ulrich & Bonß, Wolfgang (1984). Soziologie und Modernisierung. Zur Ortsbestimmung der Verwendungsforschung. *Soziale Welt*, 35 (3), 389-397.
- Bedorf, Thomas & Röttgers, Kurt (2010). *Das Politische und die Politik*. Frankfurt a.M.: Suhrkamp.
- Béland, Daniel, Howlett, Michael & Mukherjee, Ishani (2018). Instrument constituencies and public policy-making: an introduction. *Policy and Society*, 37 (1), 1-13.
- Berg-Schlosser, Dirk, & Müller-Rommel, Dieter (1991). *Vergleichende Politikwissenschaft*. Opladen: Leske + Budrich.
- Bjørnholt, Bente & Larsen, Flemming (2014). The politics of performance measurement: 'Evaluation use as mediator for politics'. *Evaluation*, 20 (4), 400-411.
- Blühdorn, Ingolfur (2013). *Simulative Demokratie. Neue Politik nach der postdemokratischen Wende*. Frankfurt a.M.: Suhrkamp.
- Blühdorn, Ingolfur, Butzlaff, Felix, Deflorian, Michael, Mock, Mirijam & Hausknost, Daniel (2020). *Nachhaltige Nicht-Nachhaltigkeit: Warum die ökologische Transformation der Gesellschaft nicht stattfindet*. Bielefeld: transcript.
- Boswell, Christina (2018). *Manufacturing Political Trust. Targets and Performance Measurement in Public Policy*. Cambridge: Cambridge University Press.
- Botzem, Sebastian (2012). *The Politics of Accounting Regulation. Organizing Transnational Standard Setting in Financial Reporting*. Cheltenham, UK/Northampton, MA: Edward Elgar.
- Bovens, Mark, Hart, Paul 't & Kuipers, Sanneke (2008). The Politics of Policy Evaluation. In Robert E. Goodin, Michael Moran, & Martin Rein (Eds.), *The Oxford Handbook of Public Policy* (pp. 319-335). Oxford.
- Brennan, Jason (2016). *Against Democracy*. Princeton: Princeton University Press.
- Bulmer, Martin (1991). National contexts for the development of social-policy research: British and American research on poverty and social welfare compared. In Peter Wagner, Carol Hirschhorn Weiss, Björn Wittrock, & Hellmut Wollmann (Eds.), *Social Sciences and Modern States. National Experiences and Theoretical Crossroads* (pp. 148-167). Cambridge: Cambridge University Press.

- Caplan, Nathan, Morrison, Andrea & Stambaugh, Russel J. (1975). *The Use of Social Science Knowledge in Policy Decision at the National Level. A Report to Respondents*. Ann Arbor: Ann Arbor Center for Research on the Utilization of Scientific Knowledge.
- DeGEval Gesellschaft für Evaluation e.V. (2017). *Standards für Evaluation*. Verfügbar unter: <https://www.degeval.org/degeval-standards/standards-fuer-evaluation/>.
- Donovan, Kevin P. (2018). The rise of the randomistas: on the experimental turn in international aid. *Economy and Society*, 47 (1), 27-58.
- Edelman, Murray J. (1964). *The Symbolic Use of Politics*. Illinois: Illini Books.
- Elzinga, Aant (1985). Research, Bureaucracy and the Drift of Epistemic Criteria. In Björn Wittrock & Aant Elzinga (Eds.), *The University Research System. The Public Policies of the Home of Scientists* (pp. 191-220). Stockholm: Almquist & Wiksell International.
- Eußner, Ansgar, & Blomeyer, Roland (2020). Evaluation Practice in the European Union. In Reinhard Stockmann, Wolfgang Meyer, & Lena Taube (Eds.), *The Institutionalisation of Evaluation in Europe* (pp. 463-479). London: Palgrave.
- Faust, Jörg (2020). Rigorose Wirkungsevaluierung – Genese, Debatte und Nutzung in der Entwicklungszusammenarbeit. *dms – der moderne staat*, 13 (1), 61-80.
- Fischer, Frank (1990). *Technocracy and the Politics of Expertise*. Newbury Park/London/New Delhi: Sage.
- Fischer, Frank (1999). *Evaluating Public Policy*. Chicago: Nelson-Hall.
- Fischer, Frank (2003). *Reframing Public Policy*. Oxford: Oxford University Press.
- Fischer, Frank (2009). *Democracy & Expertise. Reorienting Policy Inquiry*. Oxford: Oxford University Press.
- Fischer, Frank & Gottweis, Herbert (Eds.). (2012). *The Argumentative Turn Revisited: Public Policy as Communicative Practice*. Durham/London: Duke University Press.
- Goffman, Erving (1959). *The Presentation of Self in Everyday Life*. New York: Anchor Books.
- Greven, Michael (2010). Verschwindet das Politische in der politischen Gesellschaft? Über Strategien der Kontingenzerleugnung. In Thomas Bedorf & Kurt Röttgers (Hrsg.), *Das Politische und die Politik* (S. 68-88). Frankfurt a.M.: Suhrkamp.
- Haas, Ernst B. (2013). Epistemic Communities. In Joel Krieger (Ed.), *The Oxford Companion to Comparative Politics (Volume 1)* (pp. 351-359). Oxford/New York: Oxford University Press.
- Habermas, Jürgen (2001). *The Postnational Constellation*. Cambridge: MIT Press.
- Hamann, Julian, Kaldewey, David & Schubert, Julia (2019). Ist gesellschaftliche Relevanz von Forschung bewertbar und wenn ja, wie? *Forschung und Gesellschaft*, 14, 13-27.
- Haus, Juliane, Korinek, Rebecca-Lea & Straßheim, Holger (2018). Expertise im Nexus. Von der Verwendungs- zur Vernetzungsforschung. In Nico Lüdtke & Anna Henkel (Hrsg.), *Das Wissen der Nachhaltigkeit. Herausforderungen zwischen Forschung und Beratung* (S. 63-88). München: Oekom Verlag.
- Hilgartner, Stephen (2000). *Science on Stage: Expert Advice as Public Drama*. Stanford (CA): Stanford University Press.
- Hood, Christopher (2007). What Happens When Transparency Meets Blame Avoidance. *Public Management Review*, 9 (2), 191-210.
- Howlett, Michael (2009). Policy analytical capacity and evidence-based policy-making: Lessons from Canada. *Canadian Public Administration*, 52 (2), 153-175.
- Jasanoff, Sheila (2005). *Designs on Nature: Science and Democracy in Europe and the United States*. Princeton (NJ): Princeton University Press.
- Jasanoff, Sheila (2011). The Practices of Objectivity in Regulatory Science. In Charles Camid, Neil Gross, & Michèle Lamont (Eds.), *Social Knowledge in the Making* (pp. 307-338). Chicago/London: University of Chicago Press.
- Kaps, Petra (2020). Formen der Evaluation experimenteller Politikgestaltung. *dms – der moderne staat*, 13 (1), 81-100.
- Kingdon, J.W. (1984). *Agendas, Alternatives, and Public Policies* (2. ed.). Boston: Little Brown and Company.

- Krapp, Max-Christopher, Pannowitsch, Sylvia & Heinelt, Hubert (2015). *Wissenspolitik und politischer Wandel. Veränderung der deutschen und britischen Arbeitsmarktpolitik seit den 1990er Jahren*. Baden-Baden: Nomos.
- Lamont, Michèle (2012). Toward a comparative sociology of valuation and evaluation. *Annual Review of Sociology*, 38 (1), 201-221.
- Leeuw, Frans L. & Furubo, Jan-Eric (2008). Evaluation Systems: What Are They and Why Study Them? *Evaluation*, 14 (2), 157–169.
- Luhmann, Niklas (1989 [1969]). *Legitimation durch Verfahren* (2. Aufl. ed.). Frankfurt a.M.: Suhrkamp.
- Majone, Giandomenico (1989). *Evidence, Argument, and Persuasion in the Policy Process*. New Haven/London: Yale University Press.
- Marchart, Oliver (2010). *Die politische Differenz*. Frankfurt a.M.: Suhrkamp.
- Meyer, Wolfgang & Stockmann, Reinhard (2020). Institutionalisierung der Evaluation in den politischen Systemen Europas. Eine vergleichende Analyse. *dms – der moderne staat*, 13 (1), 24-43.
- Min, John B. (2015). Epistocracy and democratic epistemology. *Politics in Central Europe*, 11 (1), 91-112.
- Morth, Ulrika (Ed.) (2005). *Soft Law In Governance And Regulation: An Interdisciplinary Analysis*. Edward Elgar.
- Murswiek, Axel (1994). *Regieren und Politikberatung*. Opladen: Westdeutscher Verlag.
- Nullmeier, Frank & Pritzlaff, Tanja (2009). Zu einer Theorie politischer Praktiken. *Österreichische Zeitschrift für Politikwissenschaft*, 38 (1), 7-22.
- O'Connor, Alice (2001). *Poverty Knowledge: Social Science, Social Policy, and the Poor in Twentieth-Century U.S. History*. Princeton: Princeton University Press.
- Paul, Regine (2020). Analyse and rule? A conceptual framework for explaining the variable appeals of ex-ante evaluation in policymaking. *dms – der moderne staat*, 13 (1), 124-142.
- Pollitt, Christopher & Bouckaert, Geert (2011). *Public Management Reform: A Comparative Analysis – New Public Management, Governance, and the Neo-Weberian State*. Oxford: Oxford University Press.
- Power, Michael (1997). *The Audit Society. Rituals of Verification*. Oxford: Oxford University Press.
- Quack, Sigrid (2016). Expertise and Authority in Transnational Governance. In Roger Cotterrell & Maksymilian Del Mar (Eds.), *Authority in Transnational Legal Theory. Theorising Across Disciplines*. (pp. 361-386). Cheltenham, UK/Northampton, MA: Edward Elgar.
- Radaelli, Claudio M. (2004). The diffusion of regulatory impact analysis – Best practice or lesson drawing. *European Journal of Political Research*, 43 (5), 723-747.
- Rancière, Jacques (2002). *Das Unvernehmen*. Frankfurt a.M.: Suhrkamp.
- Renn, Ortwin (1995). Styles of using scientific expertise: a comparative framework. *Science and Public Policy*, 22 (3), 147-156.
- Ringel, Leopold & Werron, Tobias (Eds.) (2019). *Rankings – Soziologische Fallstudien*. Wiesbaden: VS Verlag.
- Rueschemeyer, Dietrich, & Skocpol, Theda (Eds.) (1996). *States, Social Knowledge, and the Origins of Modern Social Policies*. New Jersey: Princeton Univ. Press.
- Sabatier, Paul, & Weible, Christopher M. (Eds.) (2014). *Theories of the Policy Process (third edition)*. Boulder: Westview.
- Schendzielorz, Cornelia & Reinhart, Martin (2020). Die Regierung der Wissenschaft im Peer Review. *dms – der moderne staat*, 13 (1), 101-123.
- Schwab, Oliver (2013). Aus Daten Sinn machen. Übersetzungsprobleme im Performanzmanagement. *dms – der moderne Staat*, Sonderheft 1/2013 „Wissen und Expertise in Politik und Verwaltung“ (Hg. Sabine Kropp & Sabine Kuhlmann), 103-125.
- Schwab, Oliver (2017). Varieties of Evaluation? Zur Rolle von Evaluierung in politisch-administrativen Systemen. In Sabine Kuhlmann & Oliver Schwab (Hrsg.), *Starke Kommunen – wirksame Verwaltung. Fortschritte und Fallstricke der internationalen Verwaltungs- und Kommunal-forschung* (S. 199-225). Wiesbaden: VS Verlag.

- Scott, James C. (1998). *Seeing Like A State. How Certain Schemes to Improve the Human Condition Have Failed*. New Haven/London: Yale University Press.
- Simons, Arno & Voss, Jan-Peter (2018). The concept of instrument constituencies: accounting for dynamics and practices of knowing governance. *Policy and Society*, 37 (1), 14-35.
- Stephenson, Paul J., Schoenefeld, Jonas J. & Leeuw, Frans L. (2019). The Politicisation of Evaluation: Constructing and Contesting EU Policy Performance. *Politische Vierteljahresschrift*, 60, 663–679.
- Stern, Elliot (2009). Evaluation policy in the European Union and its institutions. *New Directions for Evaluation*, 123, 67-85.
- Stockmann, Reinhard, Meyer, Wolfgang & Taube, Lena (2020). *The Institutionalisation of Evaluation in Europe* London: Palgrave.
- Stone, Diane (2012). *Knowledge Actors and Transnational Governance: The Private-Public Policy Nexus in the Global Agora*. London: Palgrave Macmillan.
- Stone, Diane (2019). *Making Global Policy*. Cambridge: Cambridge University Press.
- Straßheim, Holger (2013). Politische Expertise im Wandel: Zur diskursiven und institutionellen Einbettung epistemischer Autorität. *dms – der moderne Staat, Sonderheft 1/2013 „Wissen und Expertise in Politik und Verwaltung“* (Hg. Sabine Kropp/Sabine Kuhlmann), 65-87.
- Straßheim, Holger (2017a). Die Globalisierung der Verhaltenspolitik. *Jahrbuch Normative und institutionelle Grundfragen der Ökonomik, Band 16: Kapitalismus, Globalisierung, Demokratie*, 211-242.
- Straßheim, Holger (2017b). Trends towards evidence-based policy formulation. In Michael Howlett & Ishani Mukherjee (Eds.), *Handbook of Policy Formulation* (pp. 504-521). Cheltenham, UK/Northampton, MA: Edward Elgar.
- Straßheim, Holger (2018a). Policy as a body of expertise. In Hal K. Colebatch & Rob Hoppe (Eds.), *Handbook on Policy, Process and Governing* (pp. 89-108). Cheltenham, UK/Northampton, MA: Edward Elgar.
- Straßheim, Holger (2018b). Verhaltenspolitik im Wohlfahrtsstaat: Zur Mikrofokussierung in der Sozialpolitik(forschung). *Sozialer Fortschritt*, 67 (8-9), 759-782.
- Straßheim, Holger (2020). De-biasing Democracy. Behavioural public policy and the post-democratic turn. *Democratization*, 27 (3), 461-476.
- Straßheim, Holger, Jung, Arlena & Korinek, Rebecca-Lea (2015). Reframing Expertise: The Rise of Behavioural Insights and Interventions in Public Policy In Ariane Berthoin Antal, Michael Hutter, & David Stark (Eds.), *Moments of Valuation. Exploring Sites of Dissonance* (pp. 249-268). Oxford: Oxford University Press.
- Straßheim, Holger & Kettunen, Pekka (2014). When does evidence-based policy turn into policy-based evidence? Configurations, contexts and mechanisms. *Evidence & Policy*, 10 (2), 259-277.
- Straßheim, Holger & Loer, Kathrin (2019). Global Health Promotion: Who Are the Policymakers and What Are Their Interests. In Benjamin Ewert & Kathrin Loer (Eds.), *Behavioural Policies for Health Promotion and Disease Prevention* (pp. 73-92). London: Palgrave Macmillan.
- Vedung, Evert (2008). *Public Policy and Program Evaluation*. New Brunswick, NJ: Transaction Publishers.
- Wagner, Peter, Hirschhorn Weiss, Carol, Wittrock, Björn & Wollmann, Hellmut (Eds.) (1991). *Social Sciences and Modern States. National Experiences and Theoretical Crossroads*. Cambridge: Cambridge University Press.
- Weiss, Carol (1993). Politics and Evaluation: A Reprise with Mellow Overtones. *Evaluation Practice*, 14 (1), 107-109.
- Weiss, Carol (1993 [1973]). Where Politics and Evaluation Research Meet. *Evaluation Practice*, 14 (1), 93-106.
- Weiss, Carol H. (1980). Knowledge creep and decision accretion. *Knowledge: Creation, Diffusion, Utilization*, 8 (2), 381-404.
- Welch, Susan & Thompson, Kay (1980). The Impact of Federal Incentives on State Policy Innovation. *American Journal of Political Science*, 24, 715-729.
- Widmer, Thomas (2020). Wechselwirkungen von Politik und Evaluation: Befunde aus der Schweiz. *dms – der moderne staat*, 13 (1), 44-60.

- Wittrock, Björn, Wagner, Peter & Wollmann, Hellmut (1991). Social science and the modern state: policy knowledge and political institutions in Western Europe and the United States. In Peter Wagner, Carol Hirschhorn Weiss, Björn Wittrock, & Hellmut Wollmann (Eds.), *Social Sciences and Modern States. National Experiences and Theoretical Crossroads* (pp. 28-85). Cambridge: Cambridge University Press.
- Wollmann, Hellmut (2003). Evaluation in public-sector reform. Trends, potentials and limits international perspective. In Hellmut Wollmann (Ed.), *Evaluation in Public-Sector Reform. Concepts and Practice in Internationale Perspective* (pp. 231-258). Cheltenham, UK/Northampton, MA: Edward Elgar.
- Zürn, Michael (2013). Politisierung als Konzept der Internationalen Beziehungen. In Michael Zürn & Matthias Ecker-Ehrhardt (Hrsg.), *Die Politisierung der Weltpolitik. Umkämpfte internationale Institutionen* (S. 7-35). Frankfurt a.M.: Suhrkamp.
- Zürn, Michael (2018). *A Theory of Global Governance. Authority, Legitimacy, and Contestation*. Oxford: Oxford University Press.

Anschriften der Autoren:

Prof. Dr. Holger Straßheim (Korrespondenzautor), Universität Bielefeld, Fakultät für Soziologie, PF 100131, 33501 Bielefeld, E-Mail: holger.strassheim@uni-bielefeld.de.

Dr. Oliver Schwab, IfS Institut für Stadtforschung und Strukturpolitik GmbH, Lützowstr. 93, D-10785 Berlin, E-Mail: schwab@ifsberlin.de.