

Aus Evaluationen lernen?

WERNER BUSSMANN¹

1. Einleitung

Im Jahre 1636 schrieb Descartes im "Discours de la méthode" eine Passage, die berühmt geworden ist und zugleich den Interpreten viel Mühe bereitet hat:

Ma seconde maxime était d'être le plus ferme et le plus résolu en mes actions que je pourrais, et de ne suivre pas moins constamment les opinions les plus douteuses, lorsque je m'y serais une fois déterminé, que si elles eussent été très assurées. Imitant en ceci les voyageurs qui, se trouvant égarés en quelque forêt, ne doivent pas errer en tournoyant tantôt d'un côté, tantôt d'un autre, ni encore moins s'arrêter en une place, mais marcher toujours le plus droit qu'ils peuvent vers un même côté, et ne le changer point pour de faibles raisons, encore que ce n'ait peut-être été au commencement que le hasard seul qui les ait déterminés à le choisir; car, par ce moyen, s'ils ne vont justement où ils désirent, ils arriveront au moins à la fin quelque part où vraisemblablement ils seront mieux que dans le milieu d'une forêt.²

Descartes' Maxime kann in der Alltagssprache wie folgt zusammengefasst werden:

¹ Deutsche Fassung eines Vortrags, den der Verfasser an einem von der Universität Genf und dem Nationalen Forschungsprogramm "Wirksamkeit staatlicher Massnahmen" organisierten Seminar am 8. Oktober 1992 in Crans gehalten hat.

² Descartes 1976, S. 24/25.

Aus Evaluationen lernen?

WERNER BUSSMANN¹

1. Einleitung

Im Jahre 1636 schrieb Descartes im "Discours de la méthode" eine Passage, die berühmt geworden ist und zugleich den Interpreten viel Mühe bereitet hat:

Ma seconde maxime était d'être le plus ferme et le plus résolu en mes actions que je pourrais, et de ne suivre pas moins constamment les opinions les plus douteuses, lorsque je m'y serais une fois déterminé, que si elles eussent été très assurées. Imitant en ceci les voyageurs qui, se trouvant égarés en quelque forêt, ne doivent pas errer en tournoyant tantôt d'un côté, tantôt d'un autre, ni encore moins s'arrêter en une place, mais marcher toujours le plus droit qu'ils peuvent vers un même côté, et ne le changer point pour de faibles raisons, encore que ce n'ait peut-être été au commencement que le hasard seul qui les ait déterminés à le choisir; car, par ce moyen, s'ils ne vont justement où ils désirent, ils arriveront au moins à la fin quelque part où vraisemblablement ils seront mieux que dans le milieu d'une forêt.²

Descartes' Maxime kann in der Alltagssprache wie folgt zusammengefasst werden:

¹ Deutsche Fassung eines Vortrags, den der Verfasser an einem von der Universität Genf und dem Nationalen Forschungsprogramm "Wirksamkeit staatlicher Massnahmen" organisierten Seminar am 8. Oktober 1992 in Crans gehalten hat.

² Descartes 1976, S. 24/25.

Wenn jemand im Wald verloren ist, dann lohnt es sich nicht, den Weg zu erkunden. Um aus dem Dunkel herauszufinden, ist vielmehr - durch Zufall oder Überlegung - irgendein Weg einzuschlagen und unbeirrt ans Ende zu gehen.

Descartes' zweite Maxime ist ein Plädoyer gegen ein suchendes, tastendes Vorgehen. Descartes hätte, wenn er heute leben würde, wohl nicht viel übrig für Evaluationen³ (soweit sie das Dunkel des Waldes zu erhellen suchen und sich nicht im Licht der Sonne augenfällig ergeben).

Die philosophische Gegenposition zu Descartes wird bezogen von Kant⁴, Locke und Hume. Ich möchte auf das Zitieren der entsprechenden Passagen verzichten, aber kurz auf einen "Klassiker" der neueren Zeit hinweisen, nämlich das Plädoyer des Evaluationspioniers Donald T. Campbell für eine experimentelle Politik:

*"The United States and other modern nations should be ready for an experimental approach to social reform, an approach in which we try out new programs designed to cure specific social problems, in which we learn whether or not these programs are effective, and in which we retain, imitate, modify, or discard them on the basis of apparent effectiveness on the multiple imperfect criteria available."*⁵

Kurz zusammengefasst und interpretiert: Campbell befürwortet eine Mischung aus Experimenten und darauf beruhendem Lernen; dabei hat das Kriterium der Wirksamkeit ausschlaggebende Bedeutung.

Nun hat diese amerikanische Politik einer experimentierenden (Sozial-) Politik im Jahr 1992 mit den Aufständen in Los Angeles und anderen Städten offenkundig Schiffbruch erlitten. Damit verbunden ist die Frage, ob auch das dahinterstehende Modell einer experimentierenden Politik entweder falsch ist oder falsch angewendet wurde.

³ Vgl. Elster 1979, S. 57.

⁴ Kant 1964, S. 708.

⁵ Campbell 1972, S. 167.

Diesen Fragen möchte ich nachgehen. Ich möchte ermitteln, ob das Modell experimenteller Politik auf genügend soliden Grundlagen beruht, und dabei namentlich die beiden folgenden Fragen behandeln:

- können mit Evaluationen zuverlässige Kenntnisse über die Wirkung von Massnahmen beschafft werden?
- sind diese Kenntnisse auch für die Erarbeitung neuer Massnahmen hilfreich?

Das Hauptgewicht meiner Überlegungen wird auf informatorischen Aspekten beruhen, d.h. der Frage, ob die für staatliches Lernen erforderlichen Informationen durch Evaluationen grundsätzlich beschafft werden können. Auf die organisationellen und politischen Voraussetzungen bzw. Hindernisse für eine optimale Verwendung von Informationen werde ich nicht eingehen. Damit gehe ich notgedrungen von einer eher positiven Einschätzung der Möglichkeiten von Evaluationen aus. Aber indem ich alle Möglichkeiten von Evaluationen im Hinblick auf die Vision einer experimentierende Politik auslote, werde ich gleichzeitig auch die Grenzen dieses Instruments aufzeigen können.

2. Wie zuverlässig sind Kenntnisse über die Wirkungen bestehender Massnahmen?

Können wir zuverlässige Kenntnisse über die Wirkungen bestehender Massnahmen beschaffen und wenn ja, wie? Können wir - um beim Bild von Descartes zu bleiben - Licht ins Dunkel des Waldes bringen? Damit diese Frage nicht abstrakt bleibt, soll sie mit einigen Beispielen konkretisiert werden:

- Inwieweit wird die herrschende Scheidungspraxis durch das geltende Scheidungsrecht geprägt?
- Wie hat sich die Pflicht zur Umweltverträglichkeitsprüfung (UVP) auf das Bauverfahren ausgewirkt?
- Zu welchen Änderungen im Angebot an Restaurants und Bars würde eine Aufhebung kantonaler Bedürfnisklauseln im Gastgewerbe führen?

Die Schwierigkeiten der Wirkungsermittlung können verglichen werden mit dem Versuch, am Radio aus einem Gewirr mehrerer Sender den Sender "Beromünster" herauszufiltern. In der Tat ist der Staat ja nur einer unter vielen Sendern und zudem - dies eine Erkenntnis aus zahlreichen Wirkungsstudien - für individuelles Handeln gar nicht dermassen bestimmend. Wir müssen, um die Wirkungen staatlichen Handelns zu ermitteln, den Einfluss des Staats aus dem "Lärm" zahlreicher anderer Sender herauszufiltern versuchen.

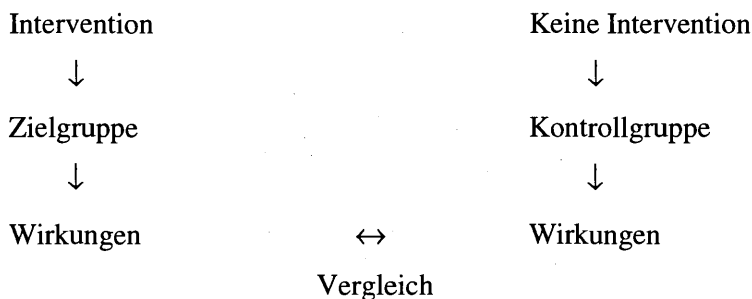
Zur Ermittlung der Wirksamkeit staatlicher Massnahmen bedient sich der Alltagsverstand der Hilfe von Vergleichen ("vorher war es besser"; "bei denen funktioniert es nicht so gut wie bei uns"; "bei solcher Unterstützung würde es uns auch besser gehen"). Evaluationen als systematische Wirkungsermittlungen und -beurteilungen beruhen in vielen Fällen ebenfalls auf Vergleichen, wobei namentlich in den USA die Palette der Vergleichsmöglichkeiten durch den Einbezug von Sozialexperimenten systematisch erweitert wurde. Ich möchte vier grundlegende Formen des Vergleichs erwähnen: Die Wirksamkeit einer Massnahme kann einmal mit dem Zustand beim Ausbleiben dieser Massnahme verglichen werden, dann lässt sie sich mit dem Zustand vergleichen, bevor sie eingeführt wurde; Massnahmen lassen sich weiter mit ähnlichen Massnahmen, die andernorts ergriffen wurden, vergleichen, und schliesslich ist auch eine Kombination dieser Vergleichsmöglichkeiten denkbar.

2.1 Der Vergleich von Situationen mit und ohne Massnahme

Die Gegenüberstellung der Situationen mit und ohne Massnahme gilt in der Evaluation als der eigentliche Königspfad, ist aber mit vielerlei praktischen Schwierigkeiten verbunden. Ihre idealtypische Ausprägung ist das Sozialexperiment, bei welchem eine bestimmte Massnahme bei einem Teil der Bevölkerung getestet und die Wirkungen mit jenen Wirkungen verglichen werden, die sich in Kontrollgruppen einstellen, in welchen die Massnahme nicht angewendet wurde.

Ein triviales Beispiel für ein - quasi natürlich ablaufendes - Experiment ist das Gurtentragen. Die Beobachtung und der Vergleich von Unfallfolgen bei Personen, welche Gurten trugen und bei solchen, welche sie nicht trugen, schaffte in den 70er Jahren die Grundlagen für die Einführung

der Gurtentragpflicht. Ein zweites Beispiel ist der zu Beginn der 80er Jahre durchgeführte "Versuch Tempo 50". Hierzu wurde in vier Testgebieten Tempo 50 versuchsweise eingeführt. Als Kontrollgruppe dienten vier möglichst ähnliche Gebiete, in welchen Tempo 60 weiterhin galt. Die Auswirkungen unterschiedlicher Tempolimiten wurden sorgfältig beobachtet und systematisch miteinander verglichen.⁶ Ein drittes Beispiel für den Vergleich von Situationen mit und ohne Massnahme sind die Schulversuche, die z. T. sehr eingehend auf ihre Auswirkungen (Lernerfolg, Zufriedenheit der Schüler, Motivation der Lehrkräfte, Kosten) untersucht wurden.⁷

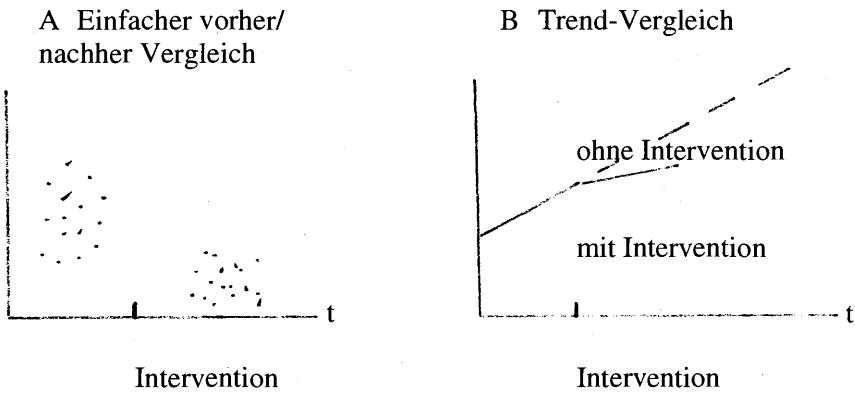


2.2 Der Vergleich von Situationen vor und nach dem Ergreifen von Massnahmen

Eine zweite grundlegende Form des Vergleichs ist der Vorher/nachher-Vergleich. Einander gegenübergestellt werden der Zustand vor und nach Inkrafttreten der Massnahme oder der Massnahmenänderung. Ein einfaches Beispiel ist die Menge des gemischten Kehrriechts und der Separatsammlungen vor und nach Einführung von Kehrriechtsackgebühren. Liegt der Entwicklung bereits ein Trend zugrunde, gilt es, diesen aus der Vorher/nachher-Analyse zu eliminieren. Beispiel sind Modelle von Luftschadstoffen; darin wird zu ermitteln versucht, wie die Entwicklung mit und ohne Luftreinemassnahmen verlaufen wäre. Die Durchführung von Vorher/nachher-Analysen bedingt, dass sehr viel statistisches Wissen zur Verfügung steht.

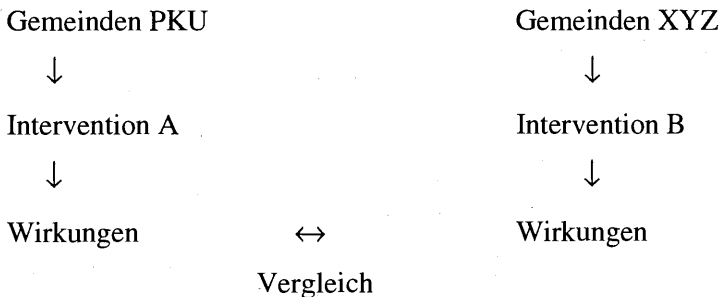
⁶ Arbeitsgruppe Verkehrssicherheit 1983.

⁷ Vgl. z.B. Jenzer 1981.



2.3 Interkommunale, -kantonale, -nationale Vergleiche

Die dritte Form des Vergleichs sind Quervergleiche über unterschiedliche staatliche Gebilde. Nationalstaatliche Eigenständigkeit und Föderalismus schaffen zusammen jene Varietät staatlicher Massnahmen, welche zwar nicht die Reinheit des Laborexperiments aufweisen, aber für die Ermittlung der Wirkungen staatlicher Massnahmen ebenfalls genutzt werden (sollten). Beispiele solcher Evaluationen sind die OECD-Länderberichte (z.B. in der Wirtschaftspolitik, in der Agrarpolitik, im Schulbereich) oder interkantonal vergleichende Analysen (Vollzug Lex Furgler⁸, Energiesparmassnahmen⁹ u.a.m.).

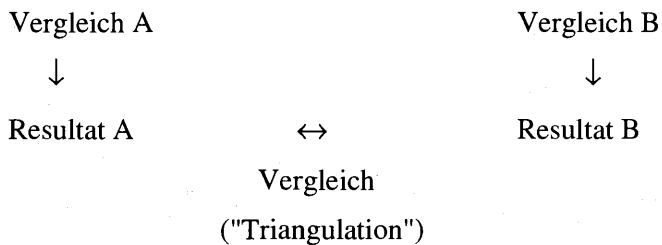


⁸ Delley et. al. 1982.

⁹ Linder u.a. 1990.

2.4 Mehrfachvergleich

Diese vierte Form basiert auf der Kombination der drei erwähnten Vergleichsmöglichkeiten, z.B. einer Vorher/nachher-Untersuchung und eines Vergleichs mit/ohne Massnahme. Als Beispiel kann die Studie zur Einführung von Tempo 50 angeführt werden, bei welcher der bereits erwähnte Vergleich der Gebiete mit Tempo 50 bzw. Tempo 60 ergänzt wurde durch eine Vorher/nachher-Untersuchung in den Untersuchungsgebieten, in welchen Tempo 50 eingeführt wurde.



2.5 Nicht vergleichende Ansätze

Auf Vergleichen beruhende Untersuchungsansätze decken bei weitem nicht die ganze Evaluationspraxis ab. Wenn eine Massnahme im ganzen Gebiet eines Landes und für die gesamte Bevölkerung gilt und sich im Zeitablauf nicht verändert oder wenn grundlegende Daten fehlen (z.B. Situation vor Einführung einer Massnahme), ist es nicht möglich, einen der oben erwähnten Vergleiche durchzuführen und die Wirkungen empirisch festzustellen. In solchen Fällen behilft man sich beispielsweise mit Experteninterviews, Befragungen von Zielgruppen oder detaillierten Fallstudien (oder Kombinationen davon). Eine Gefahr bei solchen, auf subjektiven Einschätzungen basierenden Verfahren besteht darin, dass Falscheinschätzungen und Vorurteile (unter Experten oder in der Bevölkerung) weit verbreitet sein können. Deshalb sind im Laufe der vergangenen Jahre interaktive Verfahren entwickelt worden, welche der Konfrontation von Meinungen erhebliches Gewicht zumessen. Damit soll sichergestellt werden, dass abweichende Meinungen in die Beurteilung, beispielsweise von Experten, eingespielen und einer eingehenden Überprüfung unterzogen werden.

- Die Wirkungen einer bestimmten Massnahme treten an nicht voraussehbaren Orten auf (Komplexität der Verhältnisse, Mangel an geeigneten Theorien).
- Die Wirkungen einer Massnahme treten erst nach sehr langer Zeit ein ("sleeper effects").

Beizufügen ist, dass die Evaluationszunft an und für sich die Instrumente entwickelt hat, um diese schwierigen Situationen zu meistern. Zusammen mit intuitiven und impressionistischen Formen der Wirkungsermittlungen können methodische Evaluationen zwar nicht sofort und in jedem Fall, aber mittelfristig und auf breiter Ebene zuverlässige Kenntnisse über Wirkungen staatlicher Massnahmen geben. Die erste der beiden gestellten Fragen kann somit bejaht werden. Damit kommen wir zur zweiten Frage: Sind tatsächlich auch diese Kenntnisse für die Erarbeitung neuer Massnahmen hilfreich?

4. Rückwärts in die Zukunft

Man sagt, dass der Mensch ein Wesen ist, das rückwärts in die Zukunft schreitet¹². (Ex post-) Evaluationen erlauben, die Vergangenheit und Gegenwart klarer zu sehen. Welchen Nutzen ziehen wir aus dieser Vergangenheitsschau? Wissen wir einfach, wo wir gescheitert sind und wo wir unsern nächsten - blinden - Versuch, es besser zu machen, ansetzen müssen? Können wir allenfalls lernen und Erfahrungen, die wir gemacht haben, auf neue Situationen übertragen? Können wir Entwicklungen gar antizipieren?¹³ Kurz: Erlaubt es uns das aus der Vergangenheit gewonnene Wissen, die Chancen des Irrtums bei der Gestaltung einer neuen Massnahme zu vermindern?

Meine Antwort auf die gestellten Fragen lautet: Ça dépend! Die Nutzung des aus der Vergangenheit gewonnenen Wissens für die Zukunft hängt davon ab, ob und inwieweit gewisse Hindernisse überwunden werden

¹² Dies ist auch explizit die Perspektive der Theorien über das "Organizational learning", vgl. Levitt/March 1988, S. 319.

¹³ Die möglichen Formen von Steuerungs-Wissen werden behandelt in Linder/Peters 1990, S. 72-75.

können, die einer Übernahme der in der Vergangenheit gewonnenen Erfahrungen entgegenstehen. Diese Hindernisse sind:

- Zielwechsel der "Politik",
- die mangelnde Übertragbarkeit von Erfahrungen auf neue Situationen,
- die unberechenbare Natur sozialer Veränderungen,
- die "Rentabilität" blinder Innovation.

4.1 Ziele und Zielwechsel¹⁴

Die in der Verwaltung tätigen Personen wissen, dass sie die Ärmel hochkrempeln müssen, wenn die politische Spitze Zielwechsel durchgibt und es beispielsweise gilt, zu sparen anstatt eine bestimmte Massnahme zu perfektionieren, den Umweltschutz zu beachten anstatt die landwirtschaftliche Produktion zu steigern oder Wettbewerbsbeschränkungen abzubauen statt einheimisches Gewerbe zu fördern. Zielwechsel sind - euphemistisch ausgedrückt - immer eine Chance zum Lernen. Weniger euphemistisch ausgedrückt: Zielwechsel machen immer auch Lernerfahrungen zunichte. Warum?

Lernen bedeutet immer: etwas zu lernen (ebenso wie scheitern bedeutet, an etwas zu scheitern). Das "Lernfeld", d.h. der Bereich, in dem sich Lernen abspielt, wird durch Ziele vorgegeben. Ziele geben dem Lernen somit die Richtung. Wenn Ziele ändern, ändert sich auch das Lernfeld. Das bisher erkundete Territorium ist nicht mehr dasjenige, welches man begehen möchte. Aus diesem Grund werden die entsprechenden Lernerfahrungen zu einem gewissen Teil¹⁵ wertlos (ausser es fände ein erneuter Zielwechsel in die alte Richtung statt, aber bis dahin hat sich vielleicht die Landschaft bereits wieder verändert¹⁶).

¹⁴ Zur Zielbestimmung und Mittelwahl vgl. Bussmann 1989, S. 23 - 27.

¹⁵ In welchem Umfang Lernerfahrungen obsolet werden, hängt von der "Übertragbarkeit" (oder auch "Generalisierbarkeit") von Evaluationserkenntnissen ab, ein theoretisch und empirisch noch sehr schwach erforschtes Gebiet, siehe weiter unten.

¹⁶ Vgl. die Ausführungen zum Faktor "Zeit".

Sowohl von Seiten der Verwaltung wie auch von Seiten der Evaluationszunft ergeht die Mahnung an die Politik, die Ziele endlich einmal explizit zu formulieren und dann während gewisser Zeit stabil zu halten. Die Hoffnung, dass dies je einmal eintreten wird, muss meines Erachtens jedoch frommer Wunsch bleiben. Ich möchte begründen, weshalb¹⁷. Gewisse Ziele (wie Gerechtigkeit, Freiheit, sozialer Ausgleich) haben den Charakter von Fixsternen. Sie sind aber derart allgemeinen Inhalts, dass sie konkretes Handeln kaum anleiten können. Organisationen müssen sich deshalb an Unterzielen orientieren, ähnlich wie sich ein Wanderer nicht nach dem Polarstern, sondern nach der nächsten Hügelkuppe richtet. Diese Unterziele lassen sich aus den Oberzielen nicht deduktiv ableiten, sondern werden beeinflusst durch die Möglichkeiten und Mittel, die vorhanden sind. Ein Wanderer visiert ja eine Hügelkuppe an, weil ihm der Weg dorthin begehbar scheint, und er wird seine Richtung ändern, wenn sich plötzlich eine Schlucht auftut oder sich an einer anderen Bergflanke ein bequemerer Aufstieg eröffnet. Unterziele sind, weil von der gegenwärtigen Situation beeinflusst, sehr variabel.

Nun gibt es ja noch einen neuen Hoffnungsträger, nämlich das "Lernen zu lernen"¹⁸. Solches Lernen kann beispielsweise darauf gerichtet sein, mit Zielwechseln auf geeignete Art und Weise umzugehen. Es handelt sich um eine Form höheren Lernens, bei dem die Rahmenbedingungen bisherigen Lernens zum eigentlichen Gegenstand des Lernens gemacht werden. Solches Lernen ist sehr wertvoll und hilfreich, kann aber das Lernen aufgrund von Erfahrungen mit konkreten Massnahmen nicht ersetzen. Auch wenn wir gelernt haben, mit Zielwechseln besser umzugehen, werden wir uns gegebenenfalls mit den neuen Rahmenbedingungen vertraut machen müssen. "Lernen zu lernen" kann deshalb die negativen Folgen von Zielwechseln nicht ungeschehen machen (ausser es gelänge der Verwaltung damit, sämtliche Zieländerungen zu verhindern, eine demokratiethoretisch nicht sehr angenehme Vorstellung).

¹⁷ Vgl. Bussmann 1989.

¹⁸ In der US-amerikanischen Literatur als sog. "double-loop learning" bezeichnet, vgl. Argyris 1982.

3. Wie zuverlässig ist die Wirkungsermittlung?

Erlauben Evaluationen eine genügend zuverlässige Wirkungsbeurteilung? Oft wird auch gefragt, ob eine Evaluation genügend "wissenschaftlich" ist, und in der Tat ist ein Vergleich von Evaluationen mit wissenschaftlichem Vorgehen hilfreich. Ziel der (Sozial-)Wissenschaft ist das Finden und Überprüfen von Theorien, welche uns erlauben sollen, die Wirklichkeit besser zu verstehen, sie zu "entschlüsseln". Pointiert formuliert ist für die Wissenschaft die Wirklichkeit in erster Linie Material zum Gewinnen und Überprüfen von Theorien. Diese theoriezentrierte und -kritische Dimension fehlt den Evaluationen, zumindest bis heute, völlig. Evaluationen bedienen sich allenfalls sozialwissenschaftlicher Theorien¹⁰ als Hilfsmittel zur Wirkungsermittlung. Idealtypisch ausgedrückt: Wissenschaft benutzt die Wirklichkeit, um Theorien zu gewinnen und zu überprüfen; Evaluationen benutzen Theorien (über Wirkungsverläufe), um Kenntnisse über die Wirklichkeit zu gewinnen¹¹. Evaluationsresultaten geht deshalb der Grad von Sicherheit ab, den sozialwissenschaftliche Erkenntnisse zu haben zumindest beanspruchen.

Erhebliche Probleme, die Wirkungen staatlichen Handelns zu ermitteln, bestehen für Evaluationen wie auch für impressionistische Formen der Wirkungsermittlung (Alltagsbeobachtung, "investigative journalism", politische Debatten) vor allem in folgenden Situationen:

- Die Wirkungen von Massnahmen und Massnahmenänderungen machen sich nur zögernd bemerkbar.
- Die Wirkungen einer bestimmten Massnahme sind nur schwach spürbar, treten aber bei einer grossen Zahl von Personen auf.

¹⁰ Dies entspricht auch dem Selbstverständnis der prominenten US-amerikanischen Evaluationszeitschrift "Evaluation Review", welche den Zusatz "Journal of Applied Social Research" führt.

¹¹ Metaevaluationen (somit Evaluationen von Evaluationen) werden namentlich in den USA in zunehmendem Masse eingesetzt, um die Eignung des Evaluationsdesigns (unter Einschluss der Wirkungshypothesen) und der Datengewinnung von Evaluationen zu überprüfen. Sie dienen mehr und mehr der Qualitätskontrolle von Evaluationen.

4.2 Übertragbarkeit von Erfahrungen

Das zweite Hindernis ist die mangelnde Übertragbarkeit von Erfahrungen mit bestehenden Massnahmen auf neue Massnahmen. Wäre eine solche Übertragbarkeit gegeben, würden wir zwar immer noch rückwärtsgewandert, aber nicht vollständig blind in die Zukunft hineingehen.

Leider weiss die Wissenschaft sehr wenig über die Übertragbarkeit¹⁹ von Evaluationsresultaten²⁰, denn sie beschäftigt sich in erster Linie mit Theorien und nicht mit Instrumenten staatlichen Handelns. Die Verwaltungspraktiker wissen darüber wahrscheinlich mehr als die Wissenschaftler, nur ist ihnen dieses Wissen oft kaum bewusst, und sie haben es nicht systematisch aufbereitet. Ich kann über das noch kaum ausgelotete Feld der Übertragbarkeit von Evaluationsresultaten denn auch kaum mit Ergebnissen aufwarten, aber ich möchte einige "Jalons" stecken.

Auf der einen Seite ist festzuhalten, dass bei der Vorbereitung neuer staatlicher Massnahmen von der Übertragbarkeit bisheriger Erfahrungen ausgegangen wird. Herkömmliche Instrumente (Gebote/Verbote, Beiträge, Verfahrensvorschriften, soft law u.a.m.) werden auf neue Tatbestände angewendet. Die mit der Rechtsetzung beauftragten Behörden bedienen sich dabei Analogieschlüssen ("Wenn das Instrument X in der Situation Y gewirkt hat, wird es auch in der Situation Z wirken") und Extrapolationen ("Wenn eine Verschärfung der Strafen um P Prozent einen Rückgang von Straftatbeständen um Q Prozent bewirkt hat, wird eine Verschärfung der Strafen um X Prozent einen Rückgang von Y Prozent bewirken"). Sagen wir es ganz klar: Ohne die Prämisse der Übertragbarkeit von Erfahrungen und Evaluationsresultaten wäre Rechtsetzung eine reine Lotterie.

Auf der andern Seite sind der Übertragbarkeit von Erfahrungen oft recht enge Grenzen gesetzt. Analogieschlüsse und Extrapolationen können, wie folgende Beispiele zeigen, auch in die Irre führen:

¹⁹ Die Übertragbarkeit von Evaluationsresultaten wird in der Evaluationsforschung thematisiert (aber nicht eingehend behandelt) unter den Stichworten: "external validity", "proximal validity", "generalizability". Der Begriff "Übertragbarkeit" scheint mir wesentlich aussagekräftiger als der Begriff "Generalisierbarkeit".

²⁰ Ausnahmen sind namentlich Mayntz 1983, O'Toole 1986 und Linder/Peters 1989.

- Die Verdoppelung der Subventionssätze an Gewässerschutzanlagen 1972 führte nicht zu einer Verdoppelung, sondern zu einer Verachtfachung der Beitragssumme (Extrapolation).
- Die Erhöhung der Unterschriftenzahlen für Initiative und Referendum hatte nicht die erwartete Verminderung des Gebrauchs dieser Volksrechte zur Folge (Extrapolation).
- Die Aufhebung der Prohibition für Alkohol war letztlich ein Erfolg, die Aufhebung des Verbots von Verkauf und Konsum von Heroin muss es nicht notwendigerweise sein (Analogieschluss).

Befürworter von methodischen Evaluationen gehen davon aus, dass mit diesem Instrument präzisere Analogieschlüsse und Extrapolationen möglich sind. Diese Hoffnung hat sich aber in verschiedenen Fällen nicht bewahrheitet. Ich möchte dies an drei Beispielen zeigen:

- In einer Untersuchung über den "Versuch Tempo 50"²¹ wurde festgestellt, dass diese Tempolimite im Vergleich zu Tempo 60 zu einer Verminderung der Unfälle um knapp 10 Prozent geführt habe. Daraus konnte aber nicht ermittelt werden, welche Folgen eine Verringerung der Tempolimite auf 40 Km/Std. hätte (Extrapolation).
- Die Untersuchung des in den Jahren 1970 - 1977 durchgeführten Gesamtschulversuchs Dulliken im Kanton Solothurn erbrachte äusserst positive Resultate²². Es blieb indessen sehr umstritten, ob dieses Ergebnis nicht dem Pilotcharakter und der intensiven Begleitung des Projekts zuzuschreiben war und ob es sich auf andere Schulen übertragen liesse (Analogie).
- In der Untersuchung über "Die Wirkungen der beruflichen Vorsorge auf den Arbeitsmarkt"²³ wurden die negativen Nebenwirkungen (Mobilität auf dem Arbeitsmarkt, Benachteiligung älterer und behinderter ArbeitnehmerInnen) der 2. Säule angesichts der herrschenden Konjunkturlage als gering eingestuft. Es konnte nicht ermittelt werden, wie die Situation bei wesentlich ungünstigerer Arbeitsmarktlage sein würde (Analogie/Umkehrschluss).

²¹ Arbeitsgruppe Verkehrssicherheit 1983, 247/254.

²² Jenzer 1981.

²³ Gerheuser 1991.

Was für Analogieschlüsse und Extrapolationen gesagt wurde, gilt in noch höherem Masse für Modelle. Diese bauen z. T. auf Analogieschlüssen und Extrapolationen auf und sind das wohl potenteste Instrument zur Übertragung von Erkenntnissen von einer Situation auf eine andere. Modelle sind mentale Abbildungen der Wirklichkeit. In Wirkungsmodellen werden staatlich beeinflussbare Variablen (z.B. Umfang der Geldmenge, Höhe der Beitragssätze) mit gesellschaftlichen Zielvariablen (Teuerungsrates, Niveau der Bautätigkeit) verbunden. Modelle können einerseits verwendet werden, um Vergangenheitsentwicklungen nachzuzeichnen und zu erklären, lassen sich andererseits mit ihnen die Auswirkungen alternativer staatlicher Massnahmen für Vergangenheit und/oder Zukunft simulieren. Erfolge bei der Modellanwendung in den Wirtschaftswissenschaften dürfen nicht dazu verleiten anzunehmen, Evaluationen ermöglichen in vielen Fällen die **Erklärung** und **Simulation** staatlicher Massnahmen. In den weitaus meisten Fällen verfügen wir nicht über ausgereifte Modelle, so dass Evaluationen nur Momentaufnahmen des Wirkungsgefüges bleiben und kaum Hinweise auf Entwicklungsmöglichkeiten erlauben.

4.3 Unberechenbarkeit sozialer Veränderungen

Ein drittes Hindernis für eine experimentierende Politik kann der Faktor Zeit bilden. Zwischen dem Untersuchungszeitpunkt und dem Abschluss sowie der Umsetzung einer Evaluation kann sehr viel Zeit vergehen. In diesem Zeitraum kann sich die Situation - z.B. auf dem Arbeitsmarkt oder bezüglich der Flüchtlingssituation - teilweise oder stark verändert haben. Kenntnisse, welche auf Evaluationen beruhen, können deshalb eine sehr kurze Halbwertszeit haben. Sie teilen dieses Schicksal mit sozialwissenschaftlichen Erkenntnissen, welche oft auf zeitgebundenen gesellschaftlichen Verhältnissen beruhen. Ohne hier in Details zu gehen²⁴, möchte ich trotzdem festhalten, dass eine experimentelle Politik eine wenn nicht stabile, so doch nachvollziehbare und einigermaßen voraussehbare Wirklichkeit voraussetzt; dass es daneben auch sprunghafte und chaotisch ablaufende Prozesse gibt, hat uns die jüngste Geschichte wieder in Erinnerung gerufen.

²⁴ Vgl. Bussmann 1992.

4.4 Die Rolle von Innovationen

Als vierten und letzten Punkt möchte ich die Rolle von Innovationen innerhalb einer experimentellen Politik behandeln. Experimentelle Politik muss zwei Elemente enthalten: einerseits das Innovative (Experiment), andererseits das Auswerten der damit gemachten Erfahrungen (Lernen). Experimente und darauf beruhendes Lernen können in sehr verschiedenen Kombinationen vorkommen: Auf der einen Seite gibt es die Strategie, welche in erster Linie auf Innovation und Änderung setzt und der Auswertung der damit gewonnenen Erfahrungen sehr wenig Gewicht beimisst. Verschiedene Drittweltländer haben im verzweifelten Versuch, sozialen Fortschritt zu erreichen, ständig Kurswechsel vorgenommen und Neuerungen eingeführt, ohne dem Auswerten und Verfeinern der Massnahmen das nötige Gewicht beizumessen. Auf der andern Seite gibt es die Strategie inkrementeller Veränderungen, welche dem Wagnis und Experiment nurmehr einen geringen Platz einräumt. Dies ist meist der Weg reicher Gesellschaften, die wenig zu gewinnen, aber viel zu verlieren haben. Auch dieser Weg hat seine Tücken. Der Versuch zu lernen, ohne Innovationen vorzunehmen, kann sehr leicht in eine Sackgasse führen. Die Gefahr besteht, dass "Lernen auf dem falschen Geleise" stattfindet. Somit: Perfektionierung von Massnahmen, deren Potential für Verbesserungen nur sehr beschränkt ist. Ein Beispiels solchen Lernens scheinen mir die unzähligen US-amerikanischen Versuche und Untersuchungen zur Straf-, Gefängnis- und Justizreform zu sein, die (so scheint es) an der im internationalen Vergleich sehr hohen Kriminalitäts- und Rückfälligkeitsrate nichts ändern vermochten.

Die Unergiebigkeit des "Lernens auf dem falschen Geleise" besteht darin, dass verzweifelt versucht wird, mit einer Massnahme ein bestimmtes Ziel zu erreichen, anstatt einen völlig anderen Politikansatz zu wählen. In einem gewissen Sinn können Evaluationen deshalb Feinde von Innovation sein, wenn sie allzusehr auf bestimmte Ziele fixiert sind und sich die Verwaltung und die Normadressaten zu sehr beobachtet fühlen, um Neues zu wagen. Aus der Organisationssoziologie ist bekannt, dass die Risiko- und Innovationsneigung bei "souplesse organisationelle" und einer nicht allzu genauen Erfolgsmessung höher ist als ohne diese beiden Voraussetzungen²⁵. Lernen setzt somit die Freiheit voraus, Fehler zu ma-

²⁵ Levitt/March 1988, S. 334.

chen, eine Freiheit, die sich die biologische Evolution ja auch herausnimmt.

5. Im Wald einen Weg suchen

Descartes empfiehlt dem Reisenden, das Dunkel des Waldes auf einem möglichst geraden Weg zu verlassen. Das Paradigma experimentierender Politik geht demgegenüber davon aus, dass wir den Wald erkunden und uns darin zurechtfinden können. Als Hindernisse gegenüber einer derartigen Walderkundung habe ich erwähnt:

- Richtungsänderungen, welche uns in stets neue Waldpartien führen (Zielwechsel),
- die fehlende Übertragbarkeit von Wissen über bisher bekannte Waldpartien auf unbekanntere Waldpartien (Übertragbarkeit von Evaluationsresultaten),
- ständige Vegetationsveränderungen im Wald, welche unsere bisherigen Kenntnisse zunichte machen (Unberechenbarkeit von Veränderungen),
- die Chance, anderswo, als dort wo wir suchen, einen neuen guten Weg zu finden (die Chance "blinder" Innovation).

Glücklicherweise sind wir nicht all diesen Hindernissen völlig ausgeliefert:

- Zielwechsel können zwar nicht vermieden, aber durch kluge Vorausschau und anpassungsfähige politische Institutionen geschmeidiger gemacht werden.
- Die Übertragbarkeit von Erfahrungen kann durch Transparenz über Erfolge und Misserfolge mit Handlungsinstrumenten einerseits und durch eine bessere Abstimmung von Evaluations- und übriger Sozialforschung andererseits verbessert werden.
- Der Unberechenbarkeit von Veränderungen kann durch rasche Evaluationsresultate und durch eine Beschleunigung der Entscheidungsprozesse begegnet werden.

- Die Suche nach neuen Lösungen kann erleichtert werden durch Evaluationen, welche nicht nur innerhalb eines bestimmten Massnahmenbündels ("analytische Evaluationen"), sondern über breite Politikfelder ("synthetische Evaluationen") angelegt sind.

Sollen wir Descartes oder Campbell folgen? Was Descartes angeht, so geht mir schlicht der Glauben ab, dass das Dunkel des Waldes je zu Ende ist. Was Campbell angeht, so habe ich zu zeigen versucht, dass die informatorischen Voraussetzungen²⁶ einer auf systematischem Lernen mit staatlichen Massnahmen beruhenden Strategie recht brüchig sind; wir werden uns weiterhin mit sehr unbeholfenen Mitteln im Wald zurechtfinden müssen, aber können dank dem Instrument von Evaluationen immerhin zwischen unterschiedlichen Schattierungen des Dunkels wählen.

Eine Konsequenz dieser Erkenntnis möchte ich am Schluss noch erwähnen. Campbells Vision einer experimentierenden Politik könnte zur Forderung führen, dass die PolitikerInnen die Ziele vorgeben ("multiple imperfect criteria") und die Verwaltung zusammen mit EvaluatorInnen und WissenschaftlerInnen das Gelände abtasten und wie ein "cruise missile" ins Ziel zu steuern versuchen²⁷. Neben allen andern Einwänden gegenüber einer solchen Arbeitsteilung zwischen Politik und Verwaltung²⁸ fehlen dafür die informatorischen Voraussetzungen. Zum Glück für ideenreiche PolitikerInnen und BeamtInnen besteht nämlich neben selbst-reflexiver Politik, welche auf systematisches Lernen ausgerichtet ist, noch erheblicher Raum für Wagnisse und Innovationen, welche auf dem Zufall und/oder dem Spürsinn für Zusammenhänge beruhen, welche im Rahmen von methodischen Evaluationen (noch?) nicht ermittelt werden können.

²⁶ Bezüglich der politischen und systemtheoretischen Aspekte verweise ich auf das Referat Freiburghaus in diesem Band, bezüglich der organisationellen Voraussetzungen auf Wildavsky 1987.

²⁷ Vgl. auch die Forderung nach einer Beschränkung der Führung des Bundesrates auf strategische Entscheide.

²⁸ Vgl. Habermas 1971.

Literatur

- Arbeitsgruppe Verkehrssicherheit 1983, Versuch Tempo 50, Schlussbericht, Bern, EDMT A 22315.
- ARGYRIS Chris 1982, Reasoning, Learning and Action, San Francisco/Washington/London, Jossey Bass.
- BUSSMANN Werner 1989, Von der Doppelbödigkeit des Verhältnisses zwischen Wissenschaft und Politik, *Schweizerisches Jahrbuch für Politische Wissenschaft*, 29/1989, S. 17-30.
- BUSSMANN Werner 1992, Rechtsetzung unter Unsicherheit, *Gesetzgebung heute*, 2, S. 39-52.
- CAMPBELL Donald T. 1972, Reforms as Experiments, in: Weiss Carol H., *Evaluating Action Programs, Readings in Social Action and Education*, Boston, Allyn and Bacon, p.186-223
- DELLEY Jean-Daniel et al. 1982, Le droit en action, St.-Saphorin.
- DESCARTES René 1976, Discours de la méthode. Texte et commentaire par Etienne Gilson, Paris, Librairie Philosophes J. Vrin (5^e édition).
- ELSTER Jon 1979, Ulysses and the Sirens. Studies in rationality and irrationality, Cambridge, Cambridge University Press.
- GERHEUSER Frohmüt W. 1991, Die Wirkungen der beruflichen Vorsorge auf den Arbeitsmarkt, Chur/Zürich, Rüegger Verlag.
- HABERMAS Jürgen 1971, Verwissenschaftlichte Politik und öffentliche Meinung, in: Ders., *Technik und Wissenschaft als Ideologie*, Frankfurt. a.M., Suhrkamp Verlag (5. Auflage).
- JENZER Carlo 1981, Die Evaluation des Schulversuchs "Gesamtschule Dulliken", in: Egger Eugen (Hrsg.) 1981, *Innovation und Evaluation von Schulversuchen*, Bern/Stuttgart, Haupt, S.61 - 78.
- KANT Immanuel 1964, Werke XII (hrsg. von Wilhelm Weischedel), Wiesbaden, Insel Verlag.
- LEVITT Barbara, MARCH James 1988, Organizational Learning, *Annual Review of Sociology*, 14, p. 319-340.
- LINDER Stephen H., PETERS Guy B. 1989, Instruments of Government: Perceptions and Contexts, *Journal of Public Policy*, 9, p. 35-58.
- LINDER Stephen H., PETERS Guy B. 1990, An Institutional Approach to the Theory of Policy-Making: The Role of Guidance Mechanisms in Policy Formulation, *Journal of Theoretical Politics*, 2(1), p. 59-83.

- LINDER Wolf u.a. 1990, Verbrauchsabhängige Heizkostenabrechnung, Förderung erneuerbarer Energien - Evaluation von Prozess und Wirkungen in drei Kantonen, Bern, Schriftenreihe des Forschungszentrums für schweizerische Politik.
- MAYNTZ Renate (Hrsg.) 1983, Implementation politischer Programme. Ansätze zur Theoriebildung, Opladen, Westdeutscher Verlag.
- O'TOOLE Laurence Jr. 1986, Policy Recommendations for Multi-Actor Implementation: An Assessment of the Field, *Journal of Public Policy*, 6, p. 181-210.
- WILDAVSKY Aaron 1987, The Self-Evaluating Organization, in: Speaking Truth to Power. The Art and Craft of Policy Analysis, New Brunswick, Transaction Publishers (Second Edition), p. 212-237.