

Lorenz Lassnigg

Evaluation: Aufdecken, Zudecken, oder was sonst ...

Manuskript für Beitrag in: Zilian H.G. / Flecker J.. Hg. (1997): Pathologien und Paradoxien der Arbeitswelt. Schriftenreihe Bd 2 Forum Sozialforschung. Wien: Forum Sozialforschung, 227-262.

Im vorliegenden Beitrag wird versucht, die Thematik der "Evaluation" mit dem Tagungsthema "Pathologien der Arbeitswelt und Paradoxien der Intervention" zu verknüpfen. Es gibt vielfältigste Möglichkeiten, diese Verknüpfung anzugehen, der Ansatz wird sich daher auf bestimmte Aspekte begrenzen müssen. In einem Motivenpapier zum Thema gibt Hans-Georg Zilian einen Rahmen vor, aus dem bestimmte Ansatzpunkte aufgegriffen werden können: Arbeitslosigkeit wird -- und kaum jemand wird hier widersprechen -- als eine fundamentale Pathologie der industrialisierten Länder diagnostiziert. Auch besteht, zumindest was die Rhetorik angeht, heute wieder große Einigkeit über Bedarf nach Intervention. Das Weißbuch der Europäischen Kommission über "Wachstum, Wettbewerbsfähigkeit und Beschäftigung", wie auch die Erklärungen des Gipfels von Essen und die "Jobs Study" der OECD umschreiben diesen Bedarf in eindringlicher Weise. Wenn es jedoch um die Realisierung von Interventionen und um deren Wirksamkeit geht, so zerbricht diese Einigkeit offensichtlich sehr bald.

Manche führen diesen Bruch auf Interessenskonflikte zwischen sozialen und gesellschaftlichen Gruppen zurück, letztlich auf den Konflikt zwischen Profit und Menschlichkeit oder zwischen Arbeit und Kapital. Das bereits zitierte Motivenpapier nimmt hier einen differenzierteren Standpunkt ein: Gesellschaftliche Pathologien wie Arbeitslosigkeit werden auf pathologische Mechanismen des gesellschaftlichen Lebens zurückgeführt, die das ureigenste Feld sozialwissenschaftlicher Expertise darstellen. Beispiele dafür sind unintendierte Folgen von systematischer Fehlallokation, oder Widersprüche zwischen individueller und kollektiver Rationalität, oder unkoordiniertes Verhalten selbstinteressierter Akteure, oder Widersprüche zwischen Markt und Regulation. Interventionen produzieren dann Paradoxien, wenn diese Mechanismen nicht ausreichend erkannt sind oder verstanden werden, wobei die verschiedenen Arten von pathologischen Mechanismen in rekursiver Weise auch hier wiederum zur Wirkung kommen können. Als wesentliche Problemschwerpunkte werden in diesem Zusammenhang die folgenden angeführt:

- Grundlegend differierende Perspektiven zur Bedeutung der Arbeit in der heutigen Gesellschaft, und zur Art und Weise wie die Individuen zur Arbeit motiviert werden sollen (Stichworte: Arbeitsgesellschaft, Grundeinkommen);
- Ausdehnung der Verteilungskonflikte zwischen den beiden Arbeitsmarktparteien auf Verteilungskonflikte unter den Arbeitnehmern im Zuge der zunehmenden Verbreitung von flexiblen Arbeitsverhältnissen (Stichwort: insider - outsider, Zweidrittel-Gesellschaft);
- Verschärfung des Widerspruchs zwischen steigendem Bedarf nach Ausbildung und Qualifizierung einerseits und abnehmender Neigung in den Unternehmen, zur Produktion des gesellschaftlichen "Qualifikationspools" beizutragen (Stichworte: Selektionsdilemma, Externalisierung).

Auf diesem Hintergrund ergeben sich große Aufgaben für die sozialwissenschaftliche Forschung, die mit einem Wort umschrieben werden: *Antizipation* der Probleme, die auftauchen können. Das Hauptproblem wird nicht darin gesehen, daß wir zuwenig über die Funktionsweise von Arbeitsmarktpolitik wissen, sondern darin, daß das Wenige, was wir wissen nicht genutzt wird. An diesem Punkt kommt Evaluation als *eine Form* dieses "Wissens" ins Spiel. Die Aufgabenstellung, die dieser Beitrag verfolgt, ist eine Explorierung des gegenwärtigen Standes der Evaluationsforschung im Bereich der Arbeitsmarktpolitik im Hinblick auf die angesprochenen Fragen und Thesen. Folgende Aspekte werden behandelt: (1) Grundlegende Entwicklungen im Bereich der Evaluation von Arbeitsmarktpolitik als spezifische Form des Wissens; (2) Zusammenfassung von Hauptergebnissen und deren Wertung im Hinblick auf die angesprochenen Pathologien und Paradoxien; (3) selbstreflexive Diskussion von Pathologien und Paradoxien innerhalb des Feldes.

1. Einige grundlegende Entwicklungsstränge

Nachdem sich die Evaluationsforschung seit mehreren Jahrzehnten im angelsächsischen Sprachraum, vor allem in den USA, zu einer eigenen professionellen Gruppierung entwickelt hat, beginnt diese Art von Aktivität nun insbesondere im Zusammenhang mit der europäischen Integration zu boomen. Es ist hier nicht der Ort einer umfassenden Darstellung,¹ vielmehr werden eher auf einer Metaebene einige grundlegende Unterscheidungen und Entwicklungsstränge skizziert, die folgende allgemeine Sicht unterstützen sollen: Evaluation ist eine

¹ Auch in Österreich gibt es mittlerweile einige zusammenfassende Darstellungen der wichtigsten Grundkonzeptionen von Evaluation, vgl. z. B. E. Holzinger, *Programm-Evaluation. Theoretische Grundlagen und Anwendungsmöglichkeiten in Raumordnung und Regionalpolitik*, Wien: BKS-ÖIR, 1993; L. Lassnigg et al., *Möglichkeiten der Evaluation und Effizienzbewertung des Arbeitsmarktservice*, Forschungsbericht, Wien: IHS, 1994.

spezifische Form der Wissensgenerierung, bei der der Prozeß der Wissensproduktion größeres Gewicht hat, als die Resultate. Das heißt, es geht eher darum, Prozesse der Reflexion unter den jeweils betroffenen Akteuren in Gang zu setzen, als allgemein gültige Wahrheiten zu produzieren. Entsprechend kann auch im Fortgang der Theoriebildung über Evaluation vor allem eine sehr starke Diversifizierung der verwendeten und theoretisch legitimierten Ansätze und Methoden festgestellt werden.

Ansätze, Gurus, Theorie und Praxis der Evaluation

Eine anschauliche Illustration der Vielfalt der Konzeptionen kann aufgrund von metatheoretischen empirischen Analysen der Theorie und Praxis der Evaluation in den USA vorgenommen werden. Eine Befragung von 14 der bekanntesten Figuren aus dem Bereich der Evaluationsforschung -- denen man durchaus den Status von "Gurus" zubilligen kann -- hat auf der theoretischen Ebene vier polare Dimensionen ergeben:²

- Qualitative (EISNER, STAKE, HOUSE, PATTON) vs. quantitative (FREEMAN, COOK, POPHAM) Orientierung hinsichtlich der Informationserfassung und -verarbeitung;
- Aktivität gerichtet auf Rechenschaftslegung ("Accountability") (SCRIVEN, POPHAM, STUFFLEBEAM) vs. Aktivität gerichtet auf Politikinformation (CRONBACH);
- Teilnahme der Klienten angestrebt (PATTON, WEISS) vs. keine Teilnahme der Klienten (ABT, EISNER);
- Ergebnisse zielen auf allgemeine Nutzung (WHOLEY, COOK) vs. Ergebnisse sind unmittelbar auf Entscheidungsprozesse gerichtet (FREEMAN, ALKIN).

Die Positionen der angesprochenen schulbildenden Zentralfiguren der US-amerikanischen Evaluationsszene sind gleichzeitig sehr vielfältig verflochten, indem keine der erfaßten Positionen auf mehr als zwei Dimensionen extreme Werte annimmt. Darüber hinaus hat die Gegenüberstellung der theoretischen mit den praktischen Positionen ergeben, daß die Unterschiede auf der theoretischen

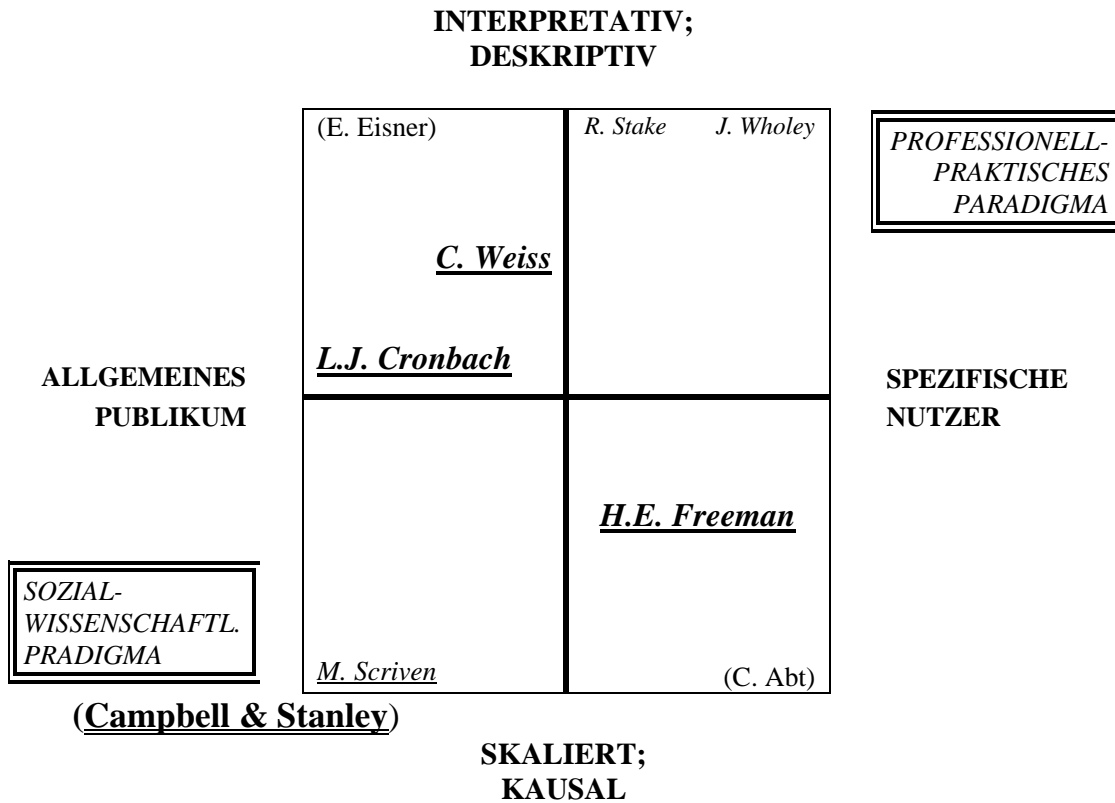
² J. E. Williams, "A numerically developed taxonomy of evaluation theory and practice", in: *Evaluation Review*, Vol.13 No.1 (February), 1989, 18-31; S.24ff)

Ebene ausgeprägter sind, als auf der praktischen Ebene. Auf der praktischen Ebene reduziert sich das Feld auf zwei Dimensionen, die der beiliegenden schematischen Darstellung zugrunde liegen:

- eine methodische Dimension (Interpretativ & deskriptiv vs. skaliert & kausal) und eine
- Zielgruppen-Dimension (Allgemeines Publikum vs. spezifische End-Nutzer).

Dieses Schema kann als eine Art Landkarte der Evaluationspraxis gesehen werden, die auch das Aufsuchen und Heranziehen von "Vorbildern" nach bestimmten Interessenskonstellationen erleichtert. Die Zuordnung unserer Gurus ergibt kein sehr ausgeprägtes Muster, es können aber tendenziell vier Extrempositionen identifiziert werden: Joseph Wholey (interpretativ-deskriptive Methodik für spezifische End-Nutzer), Clark Abt (skaliert-kausale Methodik für spezifische End-Nutzer), Elliot Eisner (interpretativ-deskriptive Methodik für das allgemeine Publikum), und Michael Scriven (skaliert-kausale Methodik für das allgemeine Publikum). Vor allem zwei Eckpunkte dieses Schemas haben in der theoretischen Diskussion große Bedeutung erhalten, die skaliert-kausale Methodologie für das allgemeine Publikum in der linken unteren Ecke kann als das sozialwissenschaftliche Paradigma bezeichnet werden, das durch das außerordentlich einflußreiche Werk von Campbell und Stanley repräsentiert wird.³ Die interpretativ-deskriptive Methodologie für spezifische Endnutzer in der rechten oberen Ecke bildet den Gegentyp des professionell-praktischen Paradigmas, das sich vor allem im rasch wachsenden Markt von Beratungsdienstleistungen manifestiert.

³ D. T. Campbell & J. C. Stanley, "Experimental and Quasi-Experimental Designs for Research", in: N. L. Gage (Ed.), Handbook of Research on Teaching, Chicago: Rand McNally, 1963, 171-246; Campbell & Stanley waren in die Untersuchung von Williams (1989) nicht einbezogen.

Tabelle 1: **Alternative Ansätze der Evaluation** (Schematisch)

Quelle: Eigene Konstruktion aufgrund der Ergebnisse von Williams 1989, sowie von Shadish & Epstein 1987

Legende: Die Zuordnung zu den konzeptionellen Feldern erfolgte aufgrund von Williams 1989 und ist aus der Lage der Namen ersichtlich; die Bekanntheit und Bedeutung für die US-Evaluationsszene ist durch die Größe und Stärke der Namen angezeigt, Namen in Klammer innerhalb der Tabelle scheinen in der zweiten Untersuchung nicht auf, Campbell & Stanley scheinen in der ersten Untersuchung nicht auf, können jedoch trotzdem eindeutig zugeordnet werden.

Vervielfachung und Diversifizierung der Ansätze und Konzepte: Der "Ballon-Effekt"

Aufgrund der schematischen Landkarte kann man auch die grundlegende Entwicklungstendenz der Evaluationsforschung skizzieren. Es handelt sich dabei nicht um eine lineare Entwicklung von A nach B, und die Entwicklung muß auch nicht unbedingt zweidimensional beschrieben werden. Vielmehr kann die bereits angesprochene Tendenz in Richtung einer Vervielfachung und Diversifizierung vielleicht am besten mit dem Bild eines Ballons illustriert werden, in den links unten beim sozialwissenschaftlichen Paradigma hineingeblasen wird, und der sich langsam ausbreitet, ohne daß die vorherigen Ansätze deshalb weniger Bedeutung bekommen würden. Auch wenn es keine lineare Entwicklung gibt, so kann man entlang der Diagonale von links unten nach rechts oben im Schema doch auch die Herausbildung zweier extremer Gegenpole von Evaluationsstrategien und -designs festmachen: Das *klassische Experiment* als Idealbild im sozialwissenschaftlichen Paradigma und die *Selbstevaluation der Agenten* im professionell-praktischen Paradigma.

Begleitet war und ist dieser Prozeß immer wieder von heftigen Auseinandersetzungen zwischen bestimmten Konzepten, wobei die jeweiligen Vertreter um ihre Hegemonie kämpfen, wie das auch immer wieder in den wissenschaftstheoretischen Diskussionen der Fall war und ist. Insgesamt scheint sich aber eine pragmatische Sicht zu verbreiten, in der die verschiedenen Ansätze und Zugangsweisen jeweils mit ihren spezifischen Vor- und Nachteilen anerkannt werden. Adäquat und sorgfältig angewandt führt jedes vernünftig ausgearbeitete Konzept zum Ziel. Diese Feststellung ist wichtig, da in der Praxis uneinlösbare Erwartungen oder aber auch Verzerrungstendenzen bestehen, die die positiven Potentiale von Evaluationsaktivitäten schwer beeinträchtigen können. Generell führt der "Ballon-Effekt" zum Eindruck von Beliebigkeit und Unüberschaubarkeit bei den Nutzern und potentiellen Nutzern von Evaluation. Es entsteht der Eindruck, man könne ohnehin "alles beweisen", Evaluation würde vor allem zur Legitimation von unangenehmen Entscheidungen oder als Ersatzhandlung zur Beruhigung von Unzufriedenen und zum Hinausschieben von Entscheidungen benutzt. Deutlich wurde dieses Problem von zwei sehr bekannten Evaluationsforschern ausgedrückt: "...if you advocate a particular policy reform or innovation, do not press to have it tested"⁴, oder noch schärfer durch *Wilensky's*

⁴ G. Burtless & R. Haveman, zit nach C. Ridell, "Evaluation of Manpower and Training Programmes: The North American Experience", in: OECD, *Evaluating Labour Market and Social Programmes. The State of a Complex Art*, Paris: OECD, 1991, 43-72, S. 49.

law: "the more evaluation, the less program development, the more demonstration projects, the less follow-through".⁵

Evaluation im Bereich der Arbeitsmarktpolitik

Im Bereich der Evaluation von Arbeitsmarktpolitik läßt sich der skizzierte Entwicklungsprozeß bis zu einem gewissen Grad ebenfalls feststellen, jedoch ist der Ballon noch viel kleiner. Gegenüber dem skizzierten Feld der Evaluationsforschung insgesamt sind im Bereich der Arbeitsmarktpolitik beträchtliche Einschränkungen zu verzeichnen und die Vielfalt der Konzeptionen kommt bisher nur sehr begrenzt zum Tragen. Verallgemeinert und in stilisierter Form kann man sagen, daß sich die Evaluation von Arbeitsmarktpolitik bis vor wenigen Jahren im wesentlichen auf die linke untere Ecke unseres Schemas konzentriert hat:

- * Sozialwissenschaftliches Paradigma, konzentriert auf die Ökonomie;
- * skaliert-kausal (quantitativ);
- * für das allgemeine Publikum, v.a. die spezialisierte wissenschaftliche Fachöffentlichkeit.

Meilensteine in der Evaluation von Arbeitsmarktpolitik waren mehrere Aktivitäten der OECD seit etwa der Mitte der achtziger Jahre. Der erste Schritt war eine Bestandsaufnahme der Aktivitäten in den Mitgliedsländern entlang einer Strukturierung in die wichtigsten Politikfelder (siehe beiliegendes Schema), deren Ergebnisse im Dymond-Bericht⁶ veröffentlicht wurden. Methodisch wurde nach dem Review-Verfahren vorgegangen, d.h. aus den Ländern wurden ExpertInnen zu den einzelnen Politikbereichen nominiert, die das verfügbare Material zusammengetragen und aufgearbeitet haben.

⁵ Wilensky, H.L., Nothing Fails Like Success: the Evaluation-Research Industry and Labour Market Policy, in: *Industrial Relations*, 24 (Winter), 1985, 1-19, S. 9.

⁶ OECD-Dymond-Report, W.R. Dymond, *Manpower measures: Main findings of the evaluation programme*, Paris: OECD (mimeo), 1989.

Tabelle 2: Strukturierung der Arbeitsmarktpolitik

I. Nachfragebezogene Maßnahmen:	- Schaffung von Arbeitsplätzen im öffentlichen Sektor - Lohnkostenzuschüsse im privaten Sektor - Regionale Anreize
II. Angebotsbezogene Maßnahmen:	- Frühpensionierung - Arbeitszeitverkürzung - Aus- und Weiterbildung, Umschulung - Maßnahmen gegen geschlechtsspezifische Diskriminierung - Regionale Mobilitätsförderung
III. Überbrückende Maßnahmen:	- Anpassungsmaßnahmen - Maßnahmen für Langzeit-Arbeitslose - Förderung des Überganges von der Schule in den Beruf
<hr/>	
IV. Überbrückende Institutionen	- Öffentliches Arbeitsmarktservice - Dezentralisierte Implementation von Maßnahmen

Quelle: OECD-Dymond-Bericht

Die Bestandsaufnahme der Aktivitäten in den verschiedenen Ländern ergab zu diesem Zeitpunkt eine sehr unterschiedliche Intensität. Zwischen einigen Ländern, die alle wichtigen Aktivitäten evaluierten (Australien, Kanada, U.K., USA) und einigen Ländern ohne Evaluationsaktivitäten (immerhin sieben Länder, darunter Japan und Italien), wurden zwei Zwischenkategorien identifiziert: Deutschland und Schweden mit einigermaßen entwickelter Praxis und eine Gruppe von sechs Ländern, darunter Frankreich, die Niederlande und auch Österreich mit "evaluations of varying degrees of sophistication on a sporadic basis"⁷. Insgesamt führte diese Studie zu einer ziemlich pessimistischen Einschätzung des Entwicklungsstandes. Kurz zusammengefaßt werden folgende Probleme konstatiert:

- * Die Rolle von Evaluations- und Monitoring-Aktivitäten bei der Erreichung von nützlichen Einsichten ist gering
- * Es bestehen große Schwierigkeiten bei internationalen Vergleichen von Maßnahmenenerfolg und bei Übertragungsversuchen (aufgrund unterschiedlicher Arbeitsmarktsituation und ökonomischer Entwicklung; aufgrund großer Unterschiede im Maßnahmeneinsatz auch bei ähnlichen Maßnahmetypen)
- * Die Effekte unterschiedlicher Arbeitsmarktsituationen und unterschiedlicher ökonomischer Kontexte auf den Erfolg von speziellen Maßnahmetypen sind schwierig zu erfassen

⁷ Ebd. S. 49.

- * Die Verwendung von synthetischen Vergleichsgruppen in Relation zu experimentellen Kontrollgruppen bei Wirkungsanalysen ist problematisch
- * Die Erfassung von Verdrängungsprozessen durch die Maßnahmen ist nicht ausreichend gelöst.
- * Die Ziel- und Aufgabenformulierung durch die politischen Autoritäten ist meist unzureichend: Die Ziele sind zu allgemein und uneindeutig, vielfältig und widersprüchlich gegenüber dem Erfordernis von objektivierbaren und quantifizierbaren Zielen
- * Eine Balance zwischen quantitativen und qualitativen Methodologien ist nicht gegeben, es besteht ein Übergewicht der quantitativen Seite
- * Die Erfassung der Kosten/Effektivität auf der Ebene der Erreichung der politischen Zielsetzungen erfordert eine simultane Evaluation aller wichtigsten Maßnahmen, was aber praktisch unmöglich ist
- * Die Erfassung von (nicht-intendierten) Effekten von zielgruppenbezogenen Maßnahmen auf andere Gruppen wird vernachlässigt.⁸

Der nächste Schritt bestand im Vergleich und der Herausarbeitung von unterschiedlichen nationalen Strategien und der Analyse von wichtigen positiven Beispielen. Ergebnis war eine Publikation über den "state of the art" der Evaluation von Arbeitsmarktpolitik, die bereits wichtige Erweiterungen der Evaluationspraxis vorzeichnet.⁹ Insbesondere werden nun drei Kategorien von Evaluationsaktivitäten unterschieden: Wirkungsanalysen, Politik-Evaluation (Policy-Analysen), Implementationsanalysen.

(1) Wirkungsanalysen. Ergebnisorientierte Wirkungsanalysen von Programmen der aktiven Arbeitsmarktpolitik bildeten zunächst deutlich den Schwerpunkt. Der Großteil der einschlägigen Literatur konzentriert sich auf derartige Studien und es ist nicht lange her, daß dieser Typ von Studien mehr oder weniger synonym mit Evaluation von Arbeitsmarktpolitik gehandelt wurde -- teilweise ist das immer noch der Fall. In der Wirkungsanalyse werden die Ziele von Maßnahmen als gegeben genommen, es geht um die Erfassung der Wirkungen bzw. Ergebnisse, die quantitativ erfolgen soll. In diesem Feld wurde ein breites methodisches Instrumentarium entwickelt, das auf die Erfassung von Netto-Effekten von Programmen abzielt (im Idealfall Kosten-Nutzen-Analysen). Es gibt eine ausführliche Auseinandersetzung um die Frage, inwiefern die klassische experimentelle Methode, die mit einer Menge von Voraussetzungen,

⁸ Ebd. S. 45-50.

⁹ OECD, Evaluating Labour Market and Social Programmes, a.a.O.

Implikationen, Beschränkungen, Nebenbedingungen und Problemen verbunden ist, durch andere Verfahren ersetzbar ist. Insgesamt ist es aufgrund der methodischen Probleme nach wie vor schwierig, klare, eindeutige und einigermaßen konkrete Aussagen zur Wirkung von arbeitsmarktpolitischen Maßnahmen zu machen.

(2) Politik-Evaluation. Die Politik-Evaluation umfaßt vor allem die Analyse der breiteren Zielsetzungen, die mit Arbeitsmarktprogrammen in den verschiedenen Ländern abgedeckt werden sollen. Damit soll ein klareres Bild des Referenzrahmens abgesteckt werden, in dem sich arbeitsmarktpolitische Maßnahmen bewegen. Es geht darum, die Angemessenheit der Zielsetzungen gegenüber der Problemlage zu beurteilen und auf diesem Hintergrund Mechanismen für die Identifikation von Problemen und Hindernissen für die Wirkung von Maßnahmen zu entwickeln. In diesem Zusammenhang wird auch ein Rahmen entwickelt, der als Beurteilungsmaßstab für die Fähigkeit der administrativen Systeme zur Entwicklung angemessener Programme und ihrer Implementation dienen kann. Es geht also vor allem um den Bereich der Ziele und angestrebten Ergebnisse von Arbeitsmarktpolitik, der erfaßt werden soll.

(3) Implementationsanalyse. Hier geht es um die näheren Umstände der praktischen Durchführung von Programmen. Es hat sich gezeigt, daß in den traditionellen ergebnisorientierten Wirkungsanalysen oft der Aspekt der Durchführung der Maßnahmen vernachlässigt wurde: Ein Programm kann wirkungslos sein, weil es von seiner Theorie her eine Fehlkonzeption darstellt, es kann aber auch ein grundsätzlich geeignetes Programm wirkungslos sein, weil der Stimulus zu gering war oder es schlecht -- oder im Extremfall (fast) gar nicht -- durchgeführt wurde. Es wird auch immer wieder betont, daß für eine Weiterentwicklung von bestimmten Programmen Implementationsanalysen erforderlich sind, da die klassischen Wirkungsanalysen nur ja/nein-Aussagen ergeben, jedoch die Umstände der Verbesserung und Weiterentwicklung nur sehr bedingt berühren. Das Feld der Implementationsanalyse ist jedoch im Bereich der Arbeitsmarktpolitik sehr wenig entwickelt. Es gibt kein Instrumentarium, dessen Anwendbarkeit breit diskutiert wäre und wo über bestimmte "Standard"-Methoden Einigkeit bestehen würde. Es werden drei unterschiedliche Aspekte der Implementationsanalyse hervorgehoben:

- Monitoring. Darunter versteht man den Aufbau eines Informationssystems, das die Überprüfung der Durchführung von Aktivitäten in Begriffen von Output und von Kosten ermöglicht. Es handelt sich darum, eine Basis von Informationen aufzubauen, die

regelmäßig und flächendeckend, dafür aber mehr oder weniger oberflächlich, über wichtige Aspekte der Durchführung von Maßnahmen Auskunft gibt.

- Performance Indicators. Darunter versteht man eine Erweiterung des Monitoring in Form der Entwicklung von Kennzahlen, die auf die Erreichung von Zielen bezogen sind. Hier gibt es eine Überschneidung mit der Wirkungsanalyse, indem dieser Zugang auch bereits in begrenztem Maß über Wirkungen Auskunft gibt.
- Implementationsanalysen im engeren Sinne. Dabei geht es darum, die Mechanismen der Durchführung von Programmen zu untersuchen. Es wird betont, daß diese Implementationsanalysen als eigenständiger Aufgabenbereich zu sehen sind und auch unabhängig von Wirkungsanalysen durchgeführt werden können (bzw. sollen). Dabei handelt es sich um mehr oder weniger detaillierte prozeß-orientierte Analysen der in Frage stehenden Arbeitsbereiche oder Programme.

In der Einschätzung des OECD-Sekretariats wird ausdrücklich hervorgehoben, daß alle drei erwähnten Kategorien -- Wirkungsanalysen, Politik-Evaluation, Implementationsanalysen -- als gleich bedeutend einzuschätzen sind, und daß insbesondere im Bereich der Implementationsanalysen der größte Entwicklungsbedarf besteht. "The three elements need to be equally well developed and co-ordinated. Deficiencies in any of the three would limit the overall contribution to policy development. (...) However, it appears that implementation evaluation is a weak link in most countries, as is the coverage of issues within each of the three components (particularly economic and social effects in impact studies)."¹⁰

In der ersten Hälfte der neunziger Jahre erfolgte im Rahmen der OECD schließlich der Aufbau einer international vergleichbaren Datenbasis über arbeitsmarktpolitische Interventionen und eine Ausarbeitung von Modellen der "Makroevaluation", die näheren Aufschluß über breitere ökonomische und soziale Wirkungen von Arbeitsmarktpolitik zu geben versuchen; parallel dazu erfolgte auch eine Aufarbeitung der Vielfalt von Detailergebnissen aus der Evaluation von spezifischen Maßnahmen in verschiedenen Ländern. Darüber hinaus hat sich die Analyse auch von der Begrenzung auf Programme und Maßnahmen befreit und den Aspekt der *Institutionen der Arbeitsmarktpolitik*, v.a. den Arbeitsmarktservice, und damit die Aktivitäten im Bereich der Beratung und

¹⁰ Ebd. S. 10

Vermittlung, sowie die Systeme der Arbeitslosenversicherung in die Evaluation einbezogen.

2. Hauptergebnisse aus der Evaluation von Arbeitsmarktpolitik

Während sich die Diskussion über lange Zeit auf spezifische methodische Probleme konzentriert hat, gibt es nun auch zusammenfassende Darstellungen von Ergebnissen. Der folgende Überblick beruht auf den Aktivitäten der OECD in den neunziger Jahren und enthält daher auch die Einschätzungen dieser Organisation, ist also sicherlich entsprechend gefiltert. Dennoch kommen darin wesentliche Fortschritte im vorhandenen Wissen über Wirkungen der aktiven Arbeitsmarktpolitik zum Ausdruck, die die zitierten kritischen und pessimistischen Einschätzungen des Dymond-Berichtes von 1989 zumindest teilweise aufheben. Diese Ergebnisse sollen und können sinnvollerweise als Ausgangspunkt für weitere Aktivitäten genommen werden -- wobei sie selbstverständlich auch den Ausgangspunkt für kritische Einwände bilden.

Ergebnisse zum Einsatz von Arbeitsmarktpolitik

Zur Erfassung des Einsatzes von Arbeitsmarktpolitik wurde eine offizielle Taxonomie entwickelt, die nun die Grundlage für die regelmäßige Erfassung von Ausgaben und Aktivitäten der verschiedenen Länder darstellt (siehe Schema). In vielen Punkten spiegelt sich darin auch die Erweiterung der Perspektive gegenüber der ersten Bestandsaufnahme in den achtziger Jahren, insbesondere die Reihung der Institutionen des Arbeitsmarktservice an die erste Stelle aktiver Maßnahmen.

Folgende Hauptergebnisse können zusammengefaßt werden:

- Die gesamten Ausgaben für Arbeitsmarktpolitik streuen in den OECD-Ländern zwischen 0,5% - 5% des Volkseinkommens, die Ausgabenquote für aktive Maßnahmen ist wesentlich geringer: zwischen 0,1% - 1,6%;
- die Höhe der Ausgaben für Arbeitsmarktpolitik ist nicht korreliert mit der Höhe der Arbeitslosenrate;
- die Höhe der aktiven Ausgaben ist aber korreliert mit der Höhe der passiven Ausgaben, und dieser Zusammenhang wird tendenziell enger (Korrelationskoeffizienten 1985: 0,5 und 1990: 0,62);
- der Anteil der aktiven Ausgaben steigt in vielen Ländern, er ist aber auch in einigen Ländern mit ursprünglich besonders hohem Anteil aktiver Ausgaben (wie

Schweden, Neuseeland, Norwegen, Australien) zurückgegangen. Es gibt also eine gewisse Tendenz in Richtung von Konvergenz des Anteils der aktiven Ausgaben;

- die Verteilung der Ausgaben auf die verschiedenen Kategorien von aktiven Ausgaben zeigt gewisse charakteristische Merkmale: Der Anteil der Ausgaben für den öffentlichen Arbeitsmarktservice am Volkseinkommen liegt in den verschiedenen Ländern in vergleichbarer Größenordnung, Ausbildungsmaßnahmen machen den größten Anteil aus und haben steigende Tendenz, bei den Maßnahmen zur Förderung von Beschäftigung gibt es ein Muster fluktuierender Steigerung und Schrumpfung;
- es gibt große Unterschiede der Beteiligung an Maßnahmen aktiver Arbeitsmarktpolitik gemessen am Arbeitskräftepotential, bei einer Spannweite zwischen Null und acht bis neun Prozent, liegt die Mehrzahl der Länder bei 2% - 4%;
- der Zusammenhang zwischen den Ausgaben für aktive Maßnahmen und dem Niveau der Arbeitslosigkeit wird tendenziell enger, und die aktive Arbeitsmarktpolitik scheint auch auf Änderungen der Arbeitslosenrate mehr zu reagieren.

Tabelle 3: OECD-Taxonomie der Maßnahmen der Arbeitsmarktpolitik

Aktive Programme

1. Öffentlicher Arbeitsmarktservice und Administration
2. Arbeitsmarktausbildung
 - a) Ausbildung für arbeitslose Erwachsene und Risikogruppen
 - b) Ausbildung für beschäftigte Erwachsene
3. Maßnahmen für Jugendliche
 - a) Maßnahmen für arbeitslose oder benachteiligte Jugendliche
 - b) Unterstützung für Lehrlinge oder ähnliche Formen allgemeiner Ausbildung
4. Förderung von Beschäftigung
 - a) Förderung von regulärer Beschäftigung im privaten Sektor
 - b) Unterstützung von Arbeitslosen für Unternehmensgründung und
 - c) Direkte Schaffung von Beschäftigung (öffentlich oder non-profit)
5. Maßnahmen für Behinderte
 - a) Berufliche Rehabilitation
 - b) Beschäftigung

"Passive" Programme (Einkommenssicherung)

6. Arbeitslosenunterstützung
7. Frühpensionierung aus arbeitsmarktpolitischen Gründen

Quelle: OECD, Employment Outlook 1993

Makroevaluation

Im Zusammenhang mit dem Aufbau der erwähnten Datenbasis wird es -- trotz der immer noch großen Probleme -- zunehmend möglich, die theoretischen Annahmen und Modelle über Makrowirkungen von Arbeitsmarktpolitik auch empirisch zu testen. Die größten Fortschritte wurden bislang jedoch eher auf konzeptioneller Ebene gemacht, indem die möglichen Wirkungen klarer herausgearbeitet wurden, und Strategien zur empirischen Prüfung entwickelt worden sind. Lars Calmfors hat in einem umfassenden Überblick die verschiedenen Wirkungen zusammengefaßt.¹¹ Makroökonomische Effekte werden vor allem dadurch erwartet, daß eine effizientere Abstimmung von Angebot und Nachfrage am Arbeitsmarkt erreicht werden könnte, bzw. daß durch eine Überbrückung der Kluft zwischen *Insidern* und *Outsidern* am Arbeitsmarkt dämpfende Effekte auf die Lohnentwicklung ausgehen könnten. Weiters könnte durch Ausbildungsmaßnahmen die Produktivität der Beschäftigten und damit die Wettbewerbsfähigkeit erhöht werden, was wiederum Beschäftigung und Einkommen positiv beeinflussen könnte. Durch Maßnahmen gegen Langzeitarbeitslosigkeit könnte die Qualität des Arbeitsangebotes aufrechterhalten werden. Andererseits könnten problematische Effekte darin bestehen, daß -- insbesondere durch Förderung von Beschäftigung -- lediglich

¹¹ L. Calmfors, "Active Labour Market Policy and Unemployment - A Framework for the Analysis of Crucial Design Features", in: *OECD Economic Studies*, No. 22, 1994, 7-47.

Substitution oder Verdrängung von regulärer Beschäftigung ausgelöst wird, oder daß aktive Maßnahmen nur "kosmetische Wirkungen" der Umdefinition von Arbeitslosen in Maßnahmenteilnehmer bewirken. Die Übersicht zeigt die breite Vielfalt möglicher Effekte, die teilweise durchaus auch in gegenläufige Richtungen weisen.

Gegenüber dem Stand der Diskussion in den achtziger Jahren besteht ein wesentlicher Fortschritt zumindest darin, daß mögliche makroökonomische Wirkungen so weit ausgearbeitet wurden, daß sie im Vergleich zu den allgemeinen wirtschaftspolitischen Konzepten überhaupt "konkurrenzfähig" geworden sind. Weiterhin sind die Zusammenhänge zu Fragen der Politik-Evaluation wie auch zur Implementation von Arbeitsmarktpolitik klarer und allgemeiner hergestellt worden. Richard Jackman hat beispielsweise die *Versicherungsfunktion* der aktiven Maßnahmen formuliert, indem über die intendierten direkten Effekte hinaus auch Probleme des Mißbrauches der Arbeitslosenversicherung gemildert werden könnten: Wenn nach einer bestimmten Periode der Arbeitslosigkeit die verpflichtende Teilnahme an einer Maßnahme erfordert würde, so würden Personen mit der Möglichkeit einen Job zu finden, dies vielleicht der Teilnahme an einer Maßnahme vorziehen -- diese Verpflichtung "can replace administrative selection by self-selection ... The disincentive effects of the benefit system would be countered, while the insurance objective would be safeguarded".¹² Die Konkretisierung von möglichen Wirkungen im Bereich der Makroevaluation stellt auch Zusammenhänge zur Implementation her. Beispielsweise wurde die Frage nach Wirkungen des Umfangs von aktiver Arbeitsmarktpolitik, gemessen an den Ausgaben insgesamt oder an den Ausgaben pro Teilnehmer, auf die Lohnentwicklung untersucht, mit dem Ergebnis von möglicherweise dämpfenden Effekten; weitere Beispiele sind Zusammenhänge zwischen Wirkungen spezieller Maßnahmen, z.B. Ausbildung, und den allgemeinen Funktionen des Arbeitsmarktservice, oder Auswirkungen des Umfangs von Langzeitarbeitslosigkeit auf Wirkungen arbeitsmarktpolitischer Interventionen.

Die empirischen Untersuchungen im Bereich der Makroevaluation haben bisher keine klaren und eindeutigen Ergebnisse geliefert, es gibt jedoch ermutigende Resultate in Richtung der erwarteten Wirkungen der aktiven Arbeitsmarktpolitik: In den untersuchten Ländern könnte das Beschäftigungswachstum verstärkt worden sein, und der Umfang aktiver Maßnahmen könnte auch tatsächlich

¹² R. Jackman, "What Can Active Labour Market Policy Do?" *Centre for Economic Performance, Discussion Paper No.226*, London: LSE, 1995, S. 12)

dämpfende Wirkungen auf die Lohnentwicklung ausgeübt haben.¹³ Nicht zuletzt als Konsequenz aus diesen Untersuchungen wird eine Verstärkung der aktiven Arbeitsmarktpolitik in den Politikempfehlungen der OECD in den letzten Jahren sehr stark betont, was früher nicht unbedingt der Fall war.

Wirkungen von Maßnahmen

Die umfassende Aufarbeitung von Evaluationsprojekten über Maßnahmewirkungen hat bisher nur wenig allgemeine Resultate erbracht, gegenüber dem pessimistischen und noch ziemlich zersplitterten Bild des Dymond-Berichtes gibt es im Employment Outlook 1995 aber doch einige grundlegende verallgemeinernde Aussagen:

-- Für gezielte Maßnahmen der direkten Unterstützung von Aktivitäten der Arbeitssuche konnten positive Effekte verhältnismäßig gut nachgewiesen werden. Diese Maßnahmen werden überdies als vergleichsweise kostengünstig hervorgehoben.

-- Für Maßnahmen im Bereich der Ausbildung und der Arbeitsbeschaffung gibt es keine eindeutigen Ergebnisse, was auch bei der breiten Vielfalt von Maßnahmen in sehr unterschiedlichen Kontexten nicht unbedingt überraschend ist. Es gibt jedoch einige gut bestätigte allgemeine Hinweise:

- * Die sorgfältige Abstimmung auf Zielgruppen fördert die Wirksamkeit;
- * ausreichende Ressourcen pro Förderfall sind ein weiteres Merkmal erfolgreicher Maßnahmen;
- * Maßnahmen zur Förderung von Beschäftigung scheinen hohe Effekte der Mitnahme, Substitution und Verdrängung zu haben.

Insgesamt folgt aus diesen Ergebnissen, daß jedenfalls breit gestreute, und unspezifische Maßnahmen mit geringer Kostenintensität ("Gießkannen-Prinzip") im Bereich der Ausbildung und der Förderung von Beschäftigung wirkungslos sind. Dies ist nicht ohne Bedeutung, weil gerade derartige Maßnahmen politisch attraktiv erscheinen -- man kann auf insgesamt hohe Mittel für viele Empfänger hinweisen -- und auch verhältnismäßig leicht implementierbar sind -- es gibt wenig und einfache Kriterien, eine rasche Ausschöpfung ist angestrebt, die Förderzahlen selbst zählen als Erfolg, etc.

Diese Ergebnisse unterstreichen die Bedeutung der Tendenz, die auch in Österreich zu beobachten ist, nämlich die institutionellen Strukturen der

¹³ OECD, "Active Labour Market Policies: Assessing Macroeconomic and Microeconomic Effects", in: *Employment Outlook*, 1993, 39-80.

Arbeitsmarktverwaltung und deren alltägliche Funktionen der Beratung und Vermittlung als wesentliche und integrale Elemente der Arbeitsmarktpolitik in die Evaluationsaktivitäten einzubeziehen, und auch die Gewichtung der verschiedenen Aktivitäten im Rahmen der gesamten Politik zu überprüfen.

"Crucial Design Features"

Als Ergebnis bisheriger Evaluation hat Lars Calmfors mehrere allgemeine Merkmale zusammengefaßt, die für die Gestaltung der Arbeitsmarktpolitik wesentliche Bedeutung haben:¹⁴

- Das Niveau der finanziellen Kompensation sollte im Bereich der aktiven Maßnahmen nicht wesentlich höher sein als im Bereich der Arbeitslosenversicherung bzw. anderer Systeme der Einkommenssicherung, um die Struktur der Anreize nicht zu verzerren;
- die Maßnahmen sollten gezielt auf die angestrebten Personengruppen und Wirkungen zugeschnitten werden;
- die Arbeitsmarktpolitik insgesamt sollte auf ein ausgewogenes Portofolio von Maßnahmen zurückgreifen, da der überproportionale Einsatz bestimmter Maßnahmetypen tendenziell zu sinkenden Skalenerträgen führt;
- die Beziehung von passiven Maßnahmen der Einkommenssicherung und von aktiven Maßnahmen der Integration in und Schaffung von Beschäftigung sollte als zusammenhängendes System gesehen und konzipiert werden, um insgesamt Anreize für Beschäftigung zu bieten. Insbesondere die Entstehung von zyklischen Wechsels zwischen diesen beiden Systemen, wo durch die aktiven Maßnahmen vorwiegend neue Ansprüche für passive Maßnahmen erzeugt werden, sollte vermieden werden.

3. Pathologien und Paradoxien

Wenn man nun versucht, den Bogen zur Frage der Beziehung zwischen Evaluation und den angesprochenen gesellschaftlichen Pathologien und Paradoxien im Bereich von Beschäftigung und Arbeitslosigkeit zu schließen, so kann als erstes gefragt werden, inwieweit die Entwicklung und die Ergebnisse von Evaluation im Bereich arbeitsmarktpolitischer Interventionen zur Aufdeckung derartiger Mechanismen beigetragen haben. Zweitens kann gefragt werden,

¹⁴ Calmfors, Active Labour Market Policy and Unemployment, a.a.O.

inwieweit innerhalb der Evaluationsaktivitäten selbst bestimmte Pathologien und Paradoxien auftreten.

"Erzeugung von Realität" durch Evaluation -- Aufdeckung der "Paradoxien der Intervention"?

Verallgemeinernd kann man aus der beschriebenen Entwicklung des *"Evaluationsbusiness"* schließen, daß offensichtlich viele Wege nach Rom führen, bzw. daß der Wert von Ergebnissen immer nur in Relation zu bestimmten Gegebenheiten bestimmt werden kann, zu den Aufgaben und Zielsetzungen, zur Rationalität in den evaluierten Systemen oder Aktivitäten, zu den gegebenen Interessenslagen und -konflikten, etc. Wenn man Evaluation als professionelle Praxis der Gewinnung und Übermittlung von anwendungsorientiertem Wissen ansieht, mit vielen Möglichkeiten der konkreten Vorgangsweise, so steht diese Praxis zwischen der "wirklichen" Praxis in den evaluierten Feldern einerseits und der Praxis der Produktion von Wissen in den etablierten Systemen der "Wissensproduktion", also v.a. der akademischen Wissenschaft oder Grundlagenwissenschaft, andererseits. Evaluation ist weder das eine noch das andere, aber je nach Konzept "borgt" sie mehr oder weniger, in verschiedenem Mischungsverhältnis, davon.¹⁵ Wesentlich dabei ist aber jedenfalls die Nähe zur Praxis, und auch ein hoher Grad der Abhängigkeit von der Praxis: Erstens gibt es ohne Praxis keine Evaluation, zweitens gibt es ohne Auftrag aus der Praxis keine Evaluation, drittens gibt es ohne Zugang zur Praxis keine Evaluation. Aus dieser Konstellation resultieren einerseits Gefahren, andererseits Möglichkeiten.

Die Gefahren ergeben sich daraus, daß bei den Evaluatoren ein hoher Grad an Verantwortung besteht, ihre Stellung als "Wahrheitsinstanz" und ihren privilegierten Zugang zur Praxis nicht zu mißbrauchen, ohne daß gleichzeitig ein Anreiz besteht, dies nicht zu tun -- eher im Gegenteil. Man kann sagen, daß jede Evaluation, in welcher Form sie auch immer durchgeführt wird, in der Erzeugung von Realität besteht. Irgend eine Praxis wird formell in irgend eine Form der Berichterstattung übersetzt, die dann irgendwie weiterwirkt (und sei es, daß sie nicht verwendet wird). Dadurch bekommt eine bestimmte Form der Darstellung "offizielle" Bedeutung. Zumindest in diesem Vorgang der formellen Berichterstattung besteht die "Macht" der Evaluatoren, indem der Status ihrer

¹⁵ Neuere wissenschaftssoziologische Untersuchungen zeigen, daß im Bereich der Wissensproduktion vielleicht grundlegende insgesamt Veränderungen vor sich gehen, die den Status der formellen "Wissenschaft" stark relativieren. Man könnte Evaluationsaktivitäten in diese Entwicklung einordnen; vgl. M. Gibbons et al., *The New Production of Knowledge. The Dynamics of Science and Research in Contemporary Societies*, London: Sage, 1994.

Aussagen jedenfalls gegenüber dem Alltagswissen erhöht ist, seien diese auch noch so unzutreffend. Auf der anderen Seite bestehen die *Möglichkeiten* darin, daß die formelle Berichterstattung einen gewissen Grad an Objektivierung haben muß, und dadurch der Überprüfung und Kritik unterzogen werden kann. Auch wenn die Ergebnisse noch so verzerrend sind, so repräsentieren sie meist doch eine real -- vielleicht unterschwellig -- vorhandene Position die in der Kritik möglicherweise aufgedeckt und widerlegt werden kann. Carol Weiss hat gegenüber unmittelbar anwendbaren Resultaten stark die allgemeinere Aufklärungsfunktion von Evaluation hervorgehoben; andere Theoretiker wie z.B. Michael Scriven oder Ernest House haben stark betont, daß im Feld der Praxis quasi "immer schon" evaluative Praktiken vorhanden sind, die durch Evaluation lediglich professionalisiert werden, wobei aufgrund des privilegierten Zuganges zur Praxis der Blick auf Phänomene oder Aspekte geworfen wird, die vielleicht von der akademischen Forschung sonst nicht aufgegriffen werden -- insofern sehen sie auch die traditionelle Unterscheidung von Grundlagenwissenschaft und angewandter Technologie im Schwinden begriffen, und Evaluation als wesentliche Quelle für den Erkenntnisfortschritt insgesamt.¹⁶

Ein Lehrstück dafür, wie tatsächlich aus dem Umkreis der Evaluationsforschung Realität erzeugt werden kann, und auch für die hohe Gefahr des Mißbrauches ist *Losing Ground*, ein Buch, das in der US-amerikanischen Sozialpolitik der achtziger Jahre eine große Rolle gespielt hat, und dessen Ausläufer auch in der einen oder anderen Form die europäische Debatte "befruchtet" haben.¹⁷ Ernest House, einer der bekanntesten Theoretiker der Evaluationsforschung, hat mit Kollegen die präsentierten Daten reanalysiert, wobei sie davon ausgegangen sind, daß der große Erfolg dieses Buches auch in seriösen Bereichen der Wissenschaft, Politik und Gesellschaft, es zu einem "social phenomenon in its own right"¹⁸ gemacht hat. Das Buch startet einen Generalangriff auf die US-amerikanischen Sozialprogramme des "Krieges gegen die Armut", in dessen Zentrum die "paradoxe" These steht, daß diese Politik nicht nur unwirksam gewesen sei, sondern "the social programs and policies of the past 20 years have caused a decline in the well-being of the poor".¹⁹ Diese These wird anhand der verschiedenen von den Sozialprogrammen betroffenen Bereiche (Armut, Arbeitslosigkeit, Einkommen und Berufe, Ausbildung, Kriminalität, "unvollständige" Familien) durch die Präsentation von empirisch quantitativem

¹⁶ Vgl. z.B. C. H. Weiss, *Evaluation Research*, Englewood Cliffs: Prentice Hall, 1972; M. Scriven, New Frontiers of Evaluation, in: *Evaluation Studies Review Annual*, Vol. 11, 1986, 93-130.

¹⁷ C. Murray, *Losing Ground*, New York: Basic Books, 1985.

¹⁸ E. House & W. Madura, "Race, Gender, and Jobs. Losing Ground on Employment", in: *Policy Sciences*, 21, 1988, 351-382; S. 351.

¹⁹ Ebd. S. 351.

Material untermauert, das aus der öffentlich verfügbaren Evaluationsforschung stammt. So weit, so gut. Die Reanalysen dieser Daten von House und seinen Mitarbeitern erbrachten für zwei Bereiche -- Ausbildung und Beschäftigung -- (für mehr fühlten sie sich als seriöse Sozialwissenschaftler nicht kompetent) den direkten Nachweis der Manipulation (oder Fälschung?), -- im einen Fall, indem nicht vergleichbare Statistiken gegenübergestellt wurden ("... in examining the educational data previously we discovered that Murray had used the wrong statistic in the main part of his analysis and that when the correct statistic was used, his argument was invalidated")²⁰, -- und im anderen Fall, indem bestimmte Entwicklungen unvollständig oder verzerrt dargestellt wurden ("How does Murray deal with this data that seem to conflict with his explanation that ... young black males ... didn't want to work ...? Mostly, by avoiding them").²¹ Auch wenn die Kritik für alle behandelten Bereiche gilt, so hat das Buch durch seine Publizität dennoch Realität erzeugt, und die Reanalysen sind auch ein Lehrstück dafür, wie schwierig es z.T. ist, und welchen Aufwand es erfordert, mittels empirischen Befunden einfache Vorurteile zu widerlegen.

Die paradoxe Gesamtaussage von Charles Murray, daß die wohlfahrtstaatliche Sozialpolitik den Armen schadet, scheint also näherer Überprüfung nicht standzuhalten. Wie der historische Essay von Albert O. Hirschman über "*the rhetoric of reaction*" zeigt, ist aus einer reformerischen Perspektive gegenüber der intellektuellen Faszination für Paradoxien überhaupt eher eine gewisse Vorsicht angebracht.²² Er zeigt in seiner Analyse der Argumentationsfiguren der Reformgegner ("Reaktionäre") in drei historischen Perioden seit dem 18. Jahrhundert (Französische Revolution, Kampf um das allgemeine Wahlrecht, Entstehung des Wohlfahrtsstaates), daß drei bevorzugte Argumentationen verwendet werden (*Perversity, Futility, Jeopardy*), deren erste der versuchte Nachweis von *perversen Effekten* ist. Er zeigt auch die bestechende Logik dieser Argumentation, da sie die Werte der Reform nicht abzulehnen braucht, sie sogar unterstützen kann, jedoch zeigt, daß aus den gutgemeinten Absichten, das gerade Gegenteil, nämlich die Verletzung eben dieser Werte resultiert.

Nachdem wir nun die Murray'sche angebliche fundamentale "Paradoxie" der Sozialpolitik angesprochen haben, können wir fragen, inwieweit die Aktivitäten und Ergebnisse von Evaluation der Arbeitsmarktpolitik zur Aufdeckung (oder

²⁰ Ebd. S. 353.

²¹ Ebd. S. 363.

²² A. O. Hirschman, *The Rhetoric of Reaction. Perversity, Futility, Jeopardy*, Cambridge Mass.: Belknap, 1991.

Zudeckung) der im Programm der Denkwerkstätte angesprochenen Pathologien und Paradoxien beitragen. Der Einfachheit halber sollen direkt die drei Problemschwerpunkte angesprochen werden:

-- Paradoxie 1 im Bereich der Bedeutung der Arbeit und Motivation für die Arbeit (Produktivismus): Der Zwang zu Beschäftigung soll gesteigert werden, obwohl keine ausreichenden Beschäftigungschancen bestehen, "die Arbeit ausgeht";

-- Paradoxie 2 im Bereich der Flexibilisierung und Ausdehnung der Verteilungskonflikte auf die Arbeitnehmer (Zweidrittelgesellschaft, Insider-Outsider): Die Strategie zur Schaffung von zusätzlicher Beschäftigung schafft einen Armutssektor innerhalb des Beschäftigungssystems;

-- Paradoxie 3 im Bereich der Qualifikationsentwicklung (Selektionsdilemma, Externalisierung): Die Strategie der Verbesserung der Qualifikationsstruktur durch Selektionsprozesse im einzelnen Unternehmen führt zur Verschlechterung des verfügbaren Qualifikationspools für alle Unternehmen.

Es zeigt sich, daß die angesprochenen Paradoxien über die Ergebnisse und auch Fragestellungen der Evaluation hinausreichen. Teilweise werden sie aber angesprochen und man kann folgende Grundlinien zusammenfassen:

-- Grundsätzlich ist der Bereich der Politik-Evaluation prädestiniert, diese Probleme anzusprechen. Es ist jedoch festzustellen, daß dieser Bereich nicht sehr entwickelt ist, und auch in die einschlägigen Diskurse eher nicht in expliziter Form eingeht. Ursprünglich, und implizit teilweise immer noch, gehen die grundsätzlichen Fragestellungen zur Arbeitsmarktpolitik eher in eine "defensive" Richtung, ob davon überhaupt positive wirtschaftliche Effekte erwartet werden können (was bis vor wenigen Jahren eher noch in Abrede gestellt wurde), in welchem Ausmaß aktive Maßnahmen eingesetzt werden sollen, und wie passive und aktive Maßnahmen in ein zusammenhängendes kohärentes System gebracht werden können. Erst damit, daß das Beschäftigungsproblem als persistentes gesellschaftspolitisches Problem nicht mehr zu leugnen war, und die Probleme mit den traditionellen Mechanismen der wirtschaftspolitischen Steuerung offenbar wurden, hat sich die Aufmerksamkeit stärker auf die Möglichkeiten der Arbeitsmarktpolitik gerichtet. Die grundsätzliche Pathologie der Arbeitslosigkeit, daß einem Teil der Menschen der Zugang zu nützlicher Betätigung verwehrt wird, wird insofern anerkannt, als die Ausweitung von aktiver Arbeitsmarktpolitik als Strategie der Eingliederung in den Arbeitsmarkt empfohlen und befürwortet wird, insbesondere gegenüber passiven Maßnahmen der Ausgliederung aus dem Arbeitsmarkt und "bloßen" Einkommenssicherung. Wie andeutungsweise gezeigt wurde, ist die theoretische und insbesondere die empirische Basis im Vergleich zu

anderen Bereichen der Ökonomie trotz aller Fortschritte nach wie vor schwach, und insbesondere sind die Beiträge nicht-ökonomischer Sozialwissenschaften geradezu verschwindend, und in den meisten Fällen kaum mehr als Randbemerkungen zu den ökonomischen Ansätzen. Kein Wunder, daß die angesprochenen Paradoxien noch weitgehend außerhalb der Reichweite der Politik-Evaluation liegen.

-- Die angesprochenen Problemschwerpunkte werden am ehesten im Bereich der Makroevaluation berührt, wobei die Fragerichtung sich aber eher nicht auf die Aufdeckung von Paradoxien bezieht, sondern durch bestimmte Prämissen vordeterminiert ist. Die vorhandene Forschung, wie sie in den OECD-Reviews zusammengefaßt wird, ist klar dem Paradigma des Produktivismus verpflichtet, d.h. Arbeitsmarktpolitik soll sowohl zur Ausweitung von Beschäftigung beitragen, als auch für die Akteure möglichst große Anreize zur Beschäftigung erzeugen (auf der Arbeitgeberseite durch Dämpfung der Arbeitskosten, auf der Arbeitnehmerseite durch entsprechende Konstruktion der Systeme der Einkommenssicherung). Die These, daß "die Arbeit ausgehe", wird bereits in den Prämissen ausgeschlossen. Die Insider-Outsider Problematik wird durch die Wirkungen auf die Einkommensverteilung angesprochen, wobei eine Dämpfung des Einkommensniveaus der Insider als wünschenswert gesehen wird. Teilweise wird dies als Gegenstrategie gegen die Spaltung der Gesellschaft konzipiert. Die Ausrichtung auf die Verbesserung des gesellschaftlichen Qualifikationspools bildet den wichtigsten Bestandteil der Arbeitsmarktpolitik, sowohl von der Konzeption her, als auch in der Realisierung. Hier wird davon ausgegangen, daß die Verschlechterung des Qualifikationspools durch Beiträge von außen (Öffentlichkeit, Arbeitnehmer) kompensiert werden soll.

-- Im Bereich der Wirkungsanalyse von arbeitsmarktpolitischen Programmen sind die zentralen Indikatoren (Einkommens- und/oder Beschäftigungszuwächse für die TeilnehmerInnen im Vergleich zu Nicht-TeilnehmerInnen) von vornherein nur begrenzt geeignet, zu den Problemschwerpunkten Aussagen zu ermöglichen. Auch wenn sich die Situation für MaßnahmenteilnehmerInnen verbessert, muß sich dadurch nicht das Beschäftigungsniveau insgesamt erhöhen. Es kann vielmehr zu Verdrängungs- und Substitutionseffekten kommen, d.h., es können die Einkommens- und Erwerbschancen von anderen beeinträchtigt werden. Dies wird zwar als Problem und als Fragestellung gesehen, es gibt jedoch nur unzureichende Ergebnisse dazu. Zur Insider-Outsider-Problematik gibt es Befunde, daß Maßnahmen in der Gefahr stehen, lediglich zur temporären Umverteilung innerhalb der "Outsider" beizutragen. "Training, when and if it is effective, makes economically disadvantaged persons less poor, but the gains are

not sufficiently large to lift many out of poverty".²³ Schließlich zeigt ein zusammenfassender Review über die Wirkungen von drei Jahrzehnten arbeitsmarktpolitischen Ausbildungsprogrammen in den USA, daß "... the results of these evaluations suggest, that it is unlikely that a substantial increase - even a doubling - in government-funded training services will significantly improve the workforce's skills."²⁴

-- Die Betonung von *Implementationsevaluation* bewegt sich zunächst im Rahmen der gegebenen Politik, soll eine ständige Beobachtung und kurzfristige Rückmeldungen über die laufenden Aktivitäten ermöglichen (Monitoring, Performance Indicators), und erscheint daher nicht besonders geeignet, dieses Geschehen zu transzendieren. Insbesondere die Verwendung von Performance Indicators als Management Instrument hat das Ziel Realität zu erzeugen, wobei es sich gleichzeitig auch um einen klassischen Fall des möglichen Entstehens unerwünschter Effekte handelt: An die Stelle der Erzielung sinnvoller Resultate kann in der Praxis die Absicht der Erzielung guter Indikator-Werte treten, was meistens nicht in einer 1:1 Entsprechung steht. Andererseits ergibt sich aber gerade aus den qualitativen, interaktiven und offenen Verfahren der Implementationsanalyse die Möglichkeit, neue Einsichten zu gewinnen. Tatsächlich haben sich "kritische" Positionen stark im Bereich der Implementationsforschung etabliert, die -- im Unterschied zur ökonomischen Dominanz in der Wirkungsforschung -- von anfang an wesentlich von der Politikwissenschaft dominiert war. Das Interesse war dabei stark an der Analyse des politischen und administrativen Prozesses -- von der Formulierung von Prioritäten über die Festschreibung in Gesetzen oder Programmen bis zur praktischen Realisierung -- orientiert. Der Untertitel des "klassischen" Werkes der Implementationsforschung von Pressman & Wildavsky gibt die Stoßrichtung gut wieder: "How great expectations in Washington are dashed in Oakland; or why it's amazing that federal programs work at all. This being a saga of the economic development administration as told by two sympathetic observers who seek to build morals on a foundation of ruined hopes".²⁵ Dieser Zweig hat sich jedoch nicht entsprechend entwickelt, und es ist auch bisher nicht zu einer Integration dieses Wissens mit dem Mainstream-Wissen der Evaluation von Arbeitsmarktpolitik gekommen. Beispielsweise kann das "naive Politikverständnis", das Harold Wilensky an der US-Wirkungsforschung kritisiert, in den neueren zentralen Publikationen der OECD -- z.B. im Dymond-Bericht --

²³ R. J. LaLonde, "The Promise of Public Sector-Sponsored Training Programs", in: *Journal of Economic Perspectives*, Vol 9 (Spring), 1995, 149-68; S. 165.

²⁴ Ebd. S. 165.

²⁵ J. L. Pressman & A. Wildavsky, *Implementation*, Berkeley: University of California Press, 1973ff.

nach wie vor gefunden werden.²⁶ Es wird gefordert, die Politik sollte sich in ihrem Umgang mit Zieldefinitionen an den Notwendigkeiten der Wirkungsanalyse orientieren; diese traditionelle Orientierung der Programmplanung bestimmt auch wesentlich die Politik-Konzeption der Europäischen Strukturfonds, obwohl in der Gemeinschaft der Evaluatoren eher die Ansicht verbreitet ist, diese Konzeption sei in der politischen Praxis gescheitert.²⁷

Insgesamt kann man trotz aller Widersprüchlichkeiten sagen, daß unter dem Strich eine Realität geschaffen wird, die in der Richtung eines bestimmten Paradigmas vorstrukturiert ist, und zu kurz greift, um direkte Beiträge zur Aufdeckung der angesprochenen Paradoxien zu erbringen. Ohne sagen zu wollen, daß diese zutreffend wären, wird jedenfalls nicht an ihrer Aufdeckung gearbeitet - - im Sinne der Ausgangsfrage ist die Antwort also hier tendenziell: Zudecken! Mit der Schaffung von Wahrheit oder Wahrheiten -- was immer man darunter versteht -- sollte diese Schaffung von Realität nicht verwechselt werden, was wiederum nicht ausschließt, daß wichtige Beiträge zur Schaffung von Wahrheit(en) geliefert werden.

Evaluation der Evaluation -- Widersprüche und Paradoxien innerhalb der Evaluation

In diesem Abschnitt wird nun unter Heranziehung von empirischem Material der Frage nachgegangen, inwieweit innerhalb der Praxis von Evaluation selbst Probleme und Widersprüchlichkeiten auftreten, die sich bei weiterer Analyse als Paradoxien erweisen könnten. Den Hintergrund für diese Überlegungen bildet die Beschreibung der grundlegenden Entwicklung, die durch die schematische Landkarte und das Bild des Ballons zusammengefaßt wurde. Diese Beschreibung suggeriert noch, daß die verschiedenen Konzeptionen und Ansätze mehr oder weniger harmonisch nebeneinander koexistieren, und je nach Bedarf, Informationsstand und verfügbarem Angebot -- wie im Supermarkt ausgewählt und gekauft werden können. Nun interessiert die Beziehung zwischen den verschiedenen Elementen, also was im Ballon vor sich geht, und auch die

²⁶ Wilensky, Nothing Fails Like Success, a.a.O. S. 9)

²⁷ Die Beschreibung der Ergebnisse zur Evaluation der Strukturfonds-Aktivitäten erscheint nicht besonders ergiebig, auch die Hervorhebung, daß bestimmte Länder trotz mehrmaliger Aufforderung bestimmte einfache und unverzichtbare Basisdaten nicht geliefert haben, wirft vielleicht ein Licht auf diese Probleme, vgl. E. Schoneweg, "Die Evaluierung im Rahmen der Strukturpolitik der europäischen Gemeinschaft", in: E. Holzinger (Hg.), *Programmevaluation - Patentrezept für Effizienz*, Wien: ÖIR, 1994, 21-42.

Interaktion mit den Akteuren außerhalb des Feldes, also den Auftraggebern und den Objekten.

Als Ausgangspunkt für die Beschreibung dieser Probleme und Widersprüchlichkeiten wird die Metapher der Prüfung gewählt. In der Tat scheint das Beispiel, welches vielleicht am klarsten auf den Punkt bringt, was mit Evaluation gemeint ist, die *Prüfung* -- oder mehr technokratisch ausgedrückt: *Leistungsbeurteilung* -- in der Schule zu sein, die auch eine der wichtigsten Quellen für die Entwicklung der Evaluationsforschung dargestellt hat. Die meisten Probleme und auch die sozialen Beziehungen, die im Zusammenhang mit der schulischen Leistungsbeurteilung auftreten, treten in irgend einer Form auch im Zusammenhang mit allgemeineren Aktivitäten der Evaluation auf. Es gibt die Rollen der Geprüften und der Prüfer, diese stehen in Beziehung zu den Administratoren und auch zu anderen mehr indirekt beteiligten Gruppierungen; es ist nicht von vornherein klar, wer oder was geprüft werden soll (die Schüler, die Lehrer, die Institution, das "System"); es gibt Diskussionen um die Frage, wer prüfen soll (interne oder externe Instanzen, Vorgesetzte, peers); es gibt breite Auseinandersetzungen um die Methodik, die Kriterien und die Probleme der Objektivierung (objektivierte Tests, qualitative Methoden, schriftlich oder mündlich), sowie um die Ziele der Prüfung und die Wirkungen, die sie im Unterricht auslöst.

Guba & Lincoln beschreiben den Entwicklungsprozeß von Aktivitäten und Konzeptionen der Evaluation in den USA als Abfolge von verallgemeinerten Rollenbildern der Evaluatoren im Zuge von vier "Generationen", wobei sie stark auf den Bereich des Erziehungswesens rekurreren:²⁸

(I) der *messende Techniker* nach dem ersten Weltkrieg, der sich vor allem mit standardisierten Tests beschäftigte, Evaluation wird mit Messung identifiziert;
 (II) der *Beobachter und Beschreiber* von Prozessen der Zielerreichung in den dreißiger Jahren, der positive und negative Aspekte von bestimmten Interventionen herausarbeitet und die technische Rolle beibehält, aber nur als eines unter vielen Werkzeugen;
 (III) der *beurteilende Richter* trat in den sechziger Jahren auf den Plan, der seine Urteile auf dem Hintergrund des Messens und Beschreibens fällt -- die Aufmerksamkeit konzentrierte sich auf den Prozeß der Wahrheitsfindung und auf die Methoden und Instrumente der Sicherung der Gültigkeit der Aussagen;

²⁸ E. G. Guba & Y.S. Lincoln, "The countenances of fourth-generation evaluation: Description, judgment, and negotiation", in: *Evaluation Studies Review Annual*, Vol.11, 1987, 70-88; S. 70-74.

(IV) der *Verhandler und "change agent"*, der zwischen den Interessen und Positionen der beteiligten Gruppierungen vermittelt, wird schließlich für die achtziger Jahre als Rollenbild vorgeschlagen.

Obwohl die Autoren diese Entwicklung als folgerichtigen Prozeß beschreiben, ist es in der Tat plausibler, zwischen der dritten und der vierten Generation einen Bruch zu sehen. Es findet eine wesentliche Veränderung der Rollenstruktur statt, die die Positionen der Beteiligten verändert: Die ursprüngliche Aufgabe der Evaluatoren wird in gewissem Sinne an die Objekte der Evaluation zurückgespielt, wobei diese jedoch die Mittel und Instrumente in die Hand bekommen sollen, um zu einer möglichst zutreffenden und ausgewogenen Beurteilung zu kommen. Gleichzeitig wird eventuell die Position der Auftraggeber wesentlich geschwächt, indem durch die Zusammenarbeit der Evaluatoren mit den Objekten die Evaluation nicht mehr als Machtmittel gegen diese eingesetzt werden kann. Man kann diesen Bruch auch durch die Gegenüberstellung der beiden groben Muster der ergebnisorientierten vs. der prozeßorientierten Evaluation, oder im Bereich der Arbeitsmarktpolitik mit den Kategorien der Wirkungsevaluation vs. der Implementationsevaluation beschreiben. Auf Basis dieser Überlegungen werden nun bestimmte Widersprüche klarer, die anhand der folgenden Stichworte diskutiert werden.

* Implementation und Wirkung: Konfligierende Zielsetzungen?

Wir haben gesehen, daß die Kategorie der Implementationsevaluation erst auftauchte und näher ausgearbeitet wurde, als die Wirkungsforschung bereits ziemlich weit entwickelt war. Nach wie vor gibt es eine hohe Konzentration auf den Aspekt der Wirkung, wie z.B. die dargestellten Review-Ergebnisse der OECD zur Arbeitsmarktpolitik zeigen, und wie auch die "experimentelle Wende" der achtziger Jahre in den USA zeigt, und in der Tat ist die ergebnisorientierte Evaluation immer noch die "Königsdisziplin", wenn sie auch nicht mehr synonym mit Evaluation überhaupt gesehen wird. Betrachtet man die Beziehung zwischen den beiden Kategorien von Evaluation näher, so kann man bestimmte Widersprüche feststellen, die vielleicht an eine Paradoxie heranreichen. Zunächst könnte man davon ausgehen, und das war auch die ursprüngliche "naive" Zugangsweise zur Evaluation, daß man bei neuen Programmen oder Politikbereichen zuerst einmal feststellt, *ob* sie wirken (Ergebnis), und dann in der weiteren Folge untersucht, *warum bzw. wie* sie wirken, um Möglichkeiten für Verbesserungen zu finden (Implementation). Die ergebnisorientierte Evaluation

hat dementsprechend ihre Aufmerksamkeit lediglich auf die Definition von Wirkungsindikatoren, und den methodisch einwandfreien Vergleich von Behandlung vs. Nicht-Behandlung konzentriert, die untersuchten Aktivitäten selbst wurden jedoch als "black box" behandelt in die man nicht hineinschaut. In dieser Sicht wäre Implementationsevaluation dann eine folgerichtige Erweiterung und Ergänzung zur Wirkungsforschung.

Bereits bald hat man jedoch festgestellt, daß die *ob*-Frage nach der Wirkung von Aktivitäten oft nur sinnvoll beantwortbar ist, wenn man ziemlich viel über die Implementation weiß. Wichtige Beispiele in dieser Hinsicht waren Ergebnisse zur regionalen Variation der Durchführung von Aktivitäten, manchmal auch die schlechte oder von den abstrakten Vorgaben abweichende Durchführung. Es hat sich also herausgestellt, daß die Beziehung eigentlich umgedreht werden sollte, daß die Fragen der Implementation eigentlich *vor* der Frage nach der Wirkung kommen sollten. Richard Nathan hat diese Beziehungen sehr klar ausgedrückt: "We need to think *sequentially*. Implementation/ institutional research needs to precede the study of programme impact on individuals in evaluation research. We need to know what it is that may or may not cause the individual impacts we now want to measure."²⁹ Verschärft wird diese Beziehung dadurch, daß die experimentelle Methode, die für die Erfassung von "wirklichen" Wirkungen immer wieder als einzig geeignet apostrophiert wird, aufgrund des Diktates der kontrollierten Bedingungen die Weiterentwicklung der evaluierten Aktivitäten ausschließt. Dieser Widerspruch wird vor allem bei großen, länger dauernden Projekten schlagend: Entweder man muß mit Weiterentwicklungen warten, bis die Ergebnisse vorliegen, oder die Aktivitäten werden anhand der Erfahrungen weiterentwickelt, dann gelten die Ergebnisse nicht mehr für die aktuelle Praxis.

Im Anschluß nähern wir uns nun der angesprochenen Paradoxie, indem wir die Frage stellen, warum diese einleuchtenden Zusammenhänge erst so spät aufgetaucht sind, bzw. warum diese auch dann, wenn sie einmal erkannt sind, nicht unbedingt ernstgenommen werden. Harold Wilensky (1985) hat seinem vergleichenden Kommentar über die Evaluation der US-amerikanischen und der europäischen Arbeitsmarktpolitik den schillernden Titel gegeben "*Nothing fails like success*", und er argumentiert mit der Möglichkeit einer inversen Beziehung zwischen Politik und ergebnisorientierter Evaluation: Die einen geben viel für Politik aus, die anderen geben viel für Evaluation aus. Der Schluß liegt nahe: *Evaluation als Ersatz, oder als Verhinderung von Politik*. Untermauert wird diese

²⁹ R. Nathan, R., "Evaluation Strategies With Particular Emphasis on Demonstration Studies", in: OECD, Evaluating Labour Market and Social Programmes, a.a.O. 21-42; S. 25.

Argumentation durch Beispiele, die einerseits zeigen, daß oft Wirkungen bereits in einem Stadium untersucht werden, wo die Programme sich noch gar nicht ausreichend entwickeln konnten, bzw. daß andererseits -- als Ausnahme von der Regel -- Programme gerade auch dann reduziert werden, wenn positive Wirkungen erfaßt werden. Die Paradoxie löst sich auf in "Interessen". Die oben zusammengefaßte weitere Entwicklung im letzten Jahrzehnt widerlegt eigentlich diese Argumentation, indem aus der Wirkungsforschung machtvolle Argumente für Arbeitsmarktpolitik erwachsen sind, daß aber gerade die Erfahrungen, die von Wilensky noch emphatisch als anscheinend schlagende positive Beispiele hervorgehoben werden -- wie etwa der "schwedische Weg" -- heute viel differenzierter gesehen werden.

* Die unendliche Geschichte über "wirkliche" Wirkung: Experiment oder Nicht-Experiment

Seit den sechziger Jahren gibt es im Bereich der ergebnisorientierten Evaluation die Auseinandersetzung um die Notwendigkeit der experimentellen Methode für die Erfassung der "wirklichen" Wirkungen. Die Frage dabei ist, wie man die Vergleichssituation zum Einsatz einer Maßnahme konstruiert: In Form einer experimentellen Kontrollgruppe, wo die Zuordnung zur Versuchs- oder Kontrollgruppe allein nach dem Zufallsprinzip erfolgen muß -- oder in Form von quasi-experimentellen Vergleichsgruppen, die aufgrund von verfügbaren Datensätzen ausgewählt und nach statistischen Verfahren mit der jeweiligen Versuchsgruppe verglichen werden (dieser Ansatz wird seit einiger Zeit auch in Österreich mittels Daten über Versicherungsverläufe angewendet). Das Problem zwischen experimenteller und quasi-experimenteller Methode besteht darin, daß beim Vergleichsgruppenansatz die "wirklichen" Unterschiede zwischen den verglichenen Gruppen, die im Experiment durch die zufällige Aufteilung "automatisch" kontrolliert werden, nicht bekannt sind. Im allgemeinen kennt man nur relativ wenige abstrakte Merkmale der einbezogenen Fälle, und man kann nicht ausschließen, daß die Versuchsgruppe eben indem sie in die Maßnahme einbezogen wurde, über unbeobachtete -- vielleicht sogar unbeobachtbare -- Merkmale verfügt, die die Erfolgswahrscheinlichkeit verändern. Dieses Problem wird unter dem Stichwort "selection bias" diskutiert, und seine Lösung ist die eigentliche Herausforderung der quasi-experimentellen Methode. Die Lösung wird darin gesehen, daß die Entscheidung zur Teilnahme, und damit die Zuordnung zur Versuchs- oder Vergleichsgruppe, als eigene Funktion gesondert analysiert und zur Kontrolle der Wirkungen eingesetzt wird -- und tatsächlich

werden die Ergebnisse dadurch gegenüber dem rohen Vergleich zwischen den Gruppen oft wesentlich modifiziert.

Die Auseinandersetzung zwischen den Vertretern der beiden "Lager" hat die methodischen Probleme des Nachweises von Wirkungen gegenüber den inhaltlichen Ergebnissen von Maßnahmen in den Vordergrund gerückt. Tatsächlich demonstriert diese Auseinandersetzung die Unsicherheit, die auch bei sehr sophistizierten Evaluationen über die "wirklichen" Wirkungen von Maßnahmen bleibt. Ein Meilenstein dieser Auseinandersetzung war eine Untersuchung, die an einem experimentellen Datensatz mit bekannten Ergebnissen für Versuchs- und Kontrollgruppe verschiedene Versionen von Vergleichsgruppen simulierte.³⁰ Verglichen wurden Einkommenszuwächse bei Programmen für Frauen und Jugendliche, und es ergaben sich beträchtliche Unterschiede: Von sechs unterschiedlichen Versionen von Vergleichsgruppen der Frauen erbrachte nur eine ein Ergebnis, das dem der Kontrollgruppe entspricht, drei brachten geringere Effekte, zwei erbrachten stärkere Effekte. Diese Untersuchung hat dazu beigetragen, daß in den USA die Glaubwürdigkeit quasi-experimenteller Methoden stark erschüttert wurde, so daß in den achtziger Jahren eine "experimentelle Wende" stattgefunden hat. James Heckman, der Hauptvertreter des quasi-experimentellen Lagers hat mit seinen Mitarbeitern diese Ergebnisse damit kommentiert, daß "their studies simply illustrate that false models produce poor estimates and that different false models produce different false estimates."³¹ Aufgrund von Reanalysen der Daten der jüngsten größeren Experimente werden wiederum deren Ergebnisse grundlegend in Frage gestellt: Heckman & Smith zeigen unter anderem, daß in den experimentellen Daten die angenommene Zuordnung zu Versuchs- und Kontrollgruppe empirisch nicht gedeckt ist, indem nur etwa die Hälfte der Versuchsgruppe nach Selbstauskunft tatsächlich an der Maßnahme teilgenommen hat, während etwa ein Drittel der Kontrollgruppe ähnliche Maßnahmen konsumiert hat³² -- die Versuchsanordnung unterstellt aber, daß erstere zu 100% und letztere gar nicht teilgenommen hätten...

Offensichtlich geht die unendliche Geschichte weiter, und die beschwichtigende Hoffnung eines Reviewers aus den achtziger Jahren, daß "... significant results should be available in several years. As we have waited 25 years already, this

³⁰ R. J. LaLonde & R. Maynard, "How Precise are Evaluations of Employment and Training Programs? Evidence from a Field Experiment", in: *Evaluation Review*, Vol.11, No 4, 1987, 428-51.

³¹ J. J. Heckman, V. J. Hotz & M. Dabos, Do We Need Experimental Data to Evaluate the Impact of Manpower Training on Earnings?, in: *Evaluation Review*, Vol.11, No 4, 1987, 395-427; S. 424.

³² J. J. Heckman & J. A. Smith, Assessing the Case for Social Experiments, in: *Journal of Economic Perspectives*, Vol 9 (Spring), 1995, 85-110; S. 106.

does not seem to be too great a wait"³³ hat sich nun, nach weiteren 10 Jahren -- das entspricht immerhin einer Verlängerung der Frist um 40% -- nur sehr bedingt erfüllt, was die methodische Untermauerung betrifft. Aber immerhin wurden in der Zwischenzeit Anstrengungen zu einer Synthese der Ergebnisse gesetzt, und wir verfügen über gewisse verallgemeinerte Aussagen zur Wirkung von Arbeitsmarktpolitik, wenn auch auf einer unsicheren Basis -- und der Zweifel an der "Wirklichkeit" von Wirkungen wird wohl nicht ausgeräumt werden können. Obwohl sich die OECD in der Theorie ziemlich klar der Position des experimentellen Paradigmas angeschlossen hat, wird in der Praxis wohl die eher resignative Stellungnahme des Dymond-Berichts noch länger aufrecht bleiben "... *it is better to have a somewhat imperfect control group than none at all.*"³⁴ Ein wenig drängt sich dabei die Assoziation zum Wilden mit seiner Maschin' auf: egal wohin, hauptsach' ich bin schneller dort -- auf der anderen Seite haben wir es hier wohl auch mit dem paradoxen Streben nach dem Unmöglichen zu tun, das vieles menschliche Bemühen antreibt.

* Prüfer und Geprüfte

Das vielleicht deutlichste Beispiel für die prekäre Beziehung zwischen Prüfern (Evaluatoren) und Geprüften (Programm-Personal) sind die zitierten Befunde von Heckman und seinen Kollegen. Die rigorose experimentelle Methode scheitert an der Implementation, letztlich am Widerspruch zwischen praktischen Überlegungen und Erfahrungen zum wirksamen Programmeinsatz seitens der Menschen, die mit dem Programm arbeiten einerseits, und den Notwendigkeiten einer möglichst "wirklichen" Erfassung der Wirkungen andererseits. Es ist bekannt, daß das Programm-Personal sich im Zweifelsfall ziemlich sicher gegen die Notwendigkeiten der Evaluation entscheiden wird, und außerdem auch allgemein eine gewisse Tendenz hat, die Bemühungen zur Evaluation zu unterlaufen, insbesondere wenn diese für irgendwelche vorgesetzten Instanzen durchgeführt wird. Im Unterschied zu den Schülern in der Schule hat das Programm-Personal meistens auch mehr Möglichkeiten in der Hand, dies zu tun. In diesen Widersprüchen liegen wichtige Argumente dafür, sich dem Paradigma der "vierten Generation" zuzuwenden, das wesentlich auf der Zusammenarbeit von "Prüfern" und "Geprüften" aufgebaut ist.

³³ E. W. Stromsdorfer, Economic evaluation of the Comprehensive Employment and Training Act: an overview of recent findings and advances in evaluation methods, in: *Evaluation Review*, Vol.11, 1987, 387-394; S. 393.

³⁴ OECD-Dymond-Report, a.a.O., S. 46.

Wesentlich an dieser Konzeption ist jedoch, daß nicht nur Prüfer und Geprüfte, sondern *alle* Beteiligten in den Prozeß eingebunden werden sollen, d.h. insbesondere auch die Auftraggeber. Wenn der Prozeß ernstgenommen wird, so besteht ein wesentliches Element der Evaluation darin, unterschiedliche Positionen, Wahrnehmungen und Bewertungen zwischen den Beteiligten zu erfassen, zu kommunizieren, und zu praktischen Schlußfolgerungen und Lösungen zu bringen. Wenn also bei den Auftraggebern der Verdacht besteht, daß die Geprüften die geforderten Leistungen nicht entsprechend erfüllen, so müßte dies im Prozeß der Evaluation offengelegt werden. Für und wider die verschiedenen Darstellungen müßte Material gesammelt und präsentiert werden, mit dem Ziel, eine tragfähige Wirklichkeitskonstruktion für weitere Aktionen zu erarbeiten. Diese Konzeption transzendiert mögliche problematische Konstellationen wie z.B. eine Koalition zwischen Auftraggeber und Prüfer gegen die Geprüften, aber auch eine Koalition zwischen Prüfern und Geprüften gegenüber den Auftraggebern. Wenn konstruktive Ergebnisse in einem konfliktreichen Feld angestrebt werden, erscheint dieser Weg der Evaluation erfolgversprechend zu sein -- nicht anwendbar ist diese Konzeption jedoch dann, wenn das Ziel der Evaluation lediglich in symbolischen Handlungen, oder der Legitimation von bereits mehr oder weniger gefallenen Entscheidungen besteht.

Ähnlich wie im Fall der Leistungsbeurteilung in der Schule besteht im Bereich der Evaluation von Arbeitsmarktpolitik die Tendenz, in der Praxis den Übergang zur vierten Generation nicht zu suchen, oder zumindest ihn nicht konsequent zu suchen, und anstelle dessen geradezu heroisch an der Verbesserung der traditionellen ergebnisorientierten Evaluationspraxis festzuhalten -- ob dies ein Paradox ist, oder durch "ganz normale" Mechanismen wie Macht- und Interessenskonstellationen, Kosten, Anwendungsprobleme, Logik administrativen Handelns, Faszination durch methodische Zauberei, etc. erklärt werden kann, sei hier dahingestellt.

Tabelle 4: Prinzipien der Vierten Generation

Axiome	Prinzipien
I. Werte-Pluralismus	1. Eher Konflikt als Konsens bei den Werten der Beteiligten 2. Neue Methoden der Annäherung finden (trad. Evaluation reicht nicht aus) 3. Einbeziehung der Parteien mit unterschiedlichen Werten
II. Beteiligte konstruieren ihre Wirklichkeiten	4. Evaluator muß diese Konstruktionen erlernen 5. Evaluator klärt und vermittelt die Konstruktionen wechselweise
III. Fairness	6. Evaluator muß sich um alle Positionen bemühen 7. Evaluator muß alle relevante Information sammeln und zugänglich machen 8. Evaluator muß Information aufbereiten und "Fakten" für Gegner und Befürworter liefern 9. Befunde müssen laufend und offen zur Verfügung gestellt werden
IV. Leistung und Wert(schätzung)	10. Wert abhängig von Bedürfnissen und Kontext 11. Einschätzung von Bedürfnissen und Wert nur aufgrund von lokaler Feldarbeit 12. Übertragbarkeit dieser Einschätzung nur aufgrund empirischer Ähnlichkeit der Kontexte
V. Aushandlung von Prozeß und Ergebnis	13. Beteiligte müssen auf Ergebnisse reagieren können 14. Beteiligte müssen an allen Punkten der Evaluation gehört und ernstgenommen werden 15. Beteiligte müssen in die Formulierung der Schlußfolgerungen eingebunden werden 16. Zulassen von dauernden Meinungs-verschiedenheiten (Verstehen statt Konsens) 17. Evaluator ist Mittler und <i>change agent</i> in den Verhandlungen, Monitor von Aktionen

Quelle: Guba & Lincoln 1987, 74-79

Die Prüfung der Prüfer I: Theorie und Praxis

Bei der Frage nach der Begriffsbestimmung und Abgrenzung von Tätigkeiten der Evaluation hat Ernest House vorgeschlagen, Evaluation als professionelle Praxis zu definieren, die zumindest im Mutterland der Evaluation -- den USA -- auch bereits in der Entstehung einer entsprechenden Berufsgruppe, organisiert in professionellen Vereinigungen, resultierte (die zwei wichtigsten Vereinigungen hatten in der Mitte der achtziger Jahre, vor ihrem Zusammenschluß zur American Evaluation Association, zusammen etwa 2000-3000 Mitglieder). In den verschiedensten Bereichen, insbesondere auch im Bereich der Evaluation von Sozialprogrammen und von Arbeitsmarktpolitik existiert in den USA ein verhältnismäßig breiter Markt von Evaluatoren, die teilweise in öffentlichen Institutionen (zwei Fünftel), teilweise an Universitäten (etwa ein Drittel) angesiedelt sind, oder auch in Form von privatwirtschaftlich agierenden Firmen (etwa ein Fünftel) arbeiten. Für die siebziger Jahre wird bereits eine Zahl von 500-600 privaten Firmen angegeben, die angewandte sozialwissenschaftliche Forschung anbieten. Die Fachliteratur ist dominiert von US-amerikanischen Beiträgen, kurz, man kann begründet davon ausgehen daß es in diesem Land am ehesten eine entwickelte *community* von EvaluatorInnen gibt.

Wie geht nun diese community in ihrem eigenen Gebiet mit den vorhandenen theoretischen Entwicklungen um? Idealerweise würde man annehmen, daß es eine gewisse Ausbildung in diesem Bereich gibt, und daß auch ein gewisser Grundstock an gemeinsamem Wissen geteilt wird. Tatsächlich geben mehr als die Hälfte der Evaluatoren an, an einer graduate school in Evaluation ausgebildet worden zu sein, von den akademischen Abschlüssen dominieren Pädagogik und Psychologie, die sozialwissenschaftlichen Disziplinen machen zusammen etwa ein Viertel aus. Untersuchungen über die professionelle Praxis der Evaluatoren und Evaluatorinnen deuten jedoch an, daß der vorhandene Stock an Theorie nur in geringem Maße geteilt wird, und sich daher auch nur sehr bedingt im Feld der Praxis niederschlagen kann. Shadish & Epstein haben die Bekanntheit und den Einfluß einer Reihe von 21 grundlegenden theoretischen Werken, die zwischen 1963 und 1983 (größtenteils 1975-80) erschienen sind, unter den EvaluatorInnen erhoben.³⁵ Die Hälfte dieser Werke (11) war unter maximal 20-30% der PraktikerInnen überhaupt bekannt, 5 Werke waren etwa der Hälfte bekannt, 4 Werke erfreuten sich einer breiten Bekanntheit unter etwa 70% der EvaluatorInnen, und nur ein Werk -- das älteste (Experimental and Quasi-Experimental Designs for Research von Cambell & Stanley aus dem Jahr 1963) kann als allgemein bekannt angesehen werden. Die Bekanntheit korrespondiert nicht mit dem Erscheinungsjahr. Das bedeutet, daß eine theoretische Fundierung -- wenn überhaupt -- nur maximal in bestimmten Spielarten von Evaluation vorhanden ist. Die Praxis muß daher auch im Wunderland der Evaluation jeweils eher ziemlich eng begrenzt auf bestimmte Ansätze, oder aber eher autodidaktisch "handgestrickt" sein. Im Markt der Auftragsforschung bestimmen daher mit hoher Wahrscheinlichkeit die Auftraggeber durch ihre Auswahl, welcher Ansatz für welches Problem zum Tragen kommt -- leider kann man nicht annehmen, daß diese besser informiert sind. Wenn man nun berücksichtigt, daß in den meisten anderen Ländern nicht einmal ein Mindestmaß an professioneller Öffentlichkeit vorhanden ist, so kann man im Vergleich zur "entwickelten" US-Struktur spekulieren, daß auch das professionelle Niveau noch begrenzter sein dürfte. Im Bereich der Arbeitsmarktpolitik ist in den Ländern, wo viele Evaluationsaktivitäten durchgeführt werden -- etwa Kanada, Australien, oder neuerdings die Niederlande -- festzustellen, daß diese Aktivitäten sich oft in bestimmten Institutionen, die meist auch Bestandteil der Verwaltung sind, konzentrieren. Da überdies bestimmte standardisierte Vorgaben für Evaluation

³⁵ W. R. Shadish Jr. & R. Epstein, Patterns of program evaluation practice among members of the Evaluation Research Society and Evaluation Network, in: *Evaluation Review*, Vol. 11 No.5 (October), 1987, 555-590.

bestehen, ist anzunehmen, daß das Spektrum der Methoden und Ansätze ebenfalls eng begrenzt ist.

Noch viel geringer als die Bekanntheit ist der Einfluß der genannten Werke. Wiederum sind es Campbell & Stanley mit ihrem Werk von 1963, die die US-EvaluatorInnenszene in überragendem Maß beeinflusst haben (75%); fast die Hälfte der angeführten Werke (9) wurden nur jeweils etwa von 10% der Befragten und ein weiteres Drittel von 15-20% der Befragten als für sie einflußreich bewertet. Etwas mehr Einfluß hatten die Werke von Carol Weiss (35%), Rossi & Freeman (30%) und Cronbach (zwei Werke mit 20-25%).

Als Schlußfolgerung ergibt sich, daß auch in einer viel breiteren Szene von Evaluatoren, die über eine professionelle Infrastruktur verfügt (Ausbildung, Organisation, Publikationen. etc.) die Professionalität nicht unbedingt sehr weit entwickelt sein muß -- etwas, was die Evaluatoren vermutlich nicht selten von ihren "Objekten" verlangen, scheint also bei ihnen selbst nicht unbedingt vorhanden zu sein.

Die Prüfung der Prüfer II: Standards und deren Verletzung

Weitere Untersuchungen über die Wahrnehmung von ethischen Herausforderungen und die Einhaltung bzw. Verletzung von Evaluationsstandards -- wiederum unter den EvaluatorInnen der USA -- beleuchtet die angedeuteten Defizite noch etwas schärfer. Die Aufstellung der Einschätzungen über die Verletzung von Standards der Evaluation verweist beim Merkmal, das als gravierendstes und häufigstes vorkommt, auf die mangelnde Professionalität: Es werden Evaluationen durchgeführt, obwohl die Kenntnisse und die Erfahrung dazu fehlen. An dritter Stelle steht der Einsatz von Testverfahren nicht aufgrund ihrer inhaltlichen Angemessenheit, sondern weil die Evaluatoren damit vertraut sind. Auch die übrigen Merkmale lesen sich wie eine Liste von Argumenten der Kritiker von Evaluation.

Tabelle 5 Verletzung von Evaluations-Standards

	HÄUFIG	NICHT HÄUFIG
GRAVIEREND	<ul style="list-style-type: none"> * <u><i>Durchführung einer Evaluation trotz fehlender Kenntnisse oder Erfahrung</i></u> * Opportunistischer Bericht * Auswahl von Testverfahren wegen Vertrautheit damit * Ungleiche Berücksichtigung von Interessensgruppen * "Verlangte" Evaluation trotz absehbarer Nutzlosigkeit 	<ul style="list-style-type: none"> * Änderung der Fragestellung als Anpassung an die Datenanalyse * Versprechen von Vertraulichkeit ohne das garantieren zu können * Entscheidungen ohne Rücksprache trotz entsprechender Vereinbarung
NICHT GRAVIEREND	<ul style="list-style-type: none"> * Fehlende Kenntnis der Werte von wichtigen Gruppierungen * Verlieren von Interesse nach Abgabe des Endberichtes 	

Quelle: Newman & Brown 1992

Auch die Wahrnehmung der ethischen Herausforderungen, die die Evaluatoren persönlich erlebt haben, verweist auf die Probleme, die in der Kritik an Evaluation häufig geäußert werden. Hier verweisen die häufigsten und gravierendsten Merkmale auf die Beziehungen zu den Auftraggebern, bzw. zu wichtigen Beteiligten im Evaluationsprozeß, und es geht immer um die Ergebnisse, die in verschiedensten Formen verzerrt werden sollen.

Ethische Herausforderungen für EvaluatorInnen

	HÄUFIGSTE	GRAVIERENDSTE
1. Präsentation der Ergebnisse (klar, vollständig, fair) (Druck von "Stakeholdern" für Änderung der Präsentation von Ergebnissen)	59%	45%
2. Fehlinterpretation und Mißbrauch von Ergebnissen (Unterdrückung oder Ignorierung der Ergebnisse)	28%	19%
3. Beziehungen zu "Stakeholdern" ("Vorgefaßte" Ergebnisse seitens "Stakeholder" oder beabsichtigter Mißbrauch)	24%	15%
4. Vertraulichkeit (Steit um Rechte oder Verteilung des Berichtes; Verletzung von Vertraulichkeit)	22%	11%

Quelle: Morris & Cohn 1993

Paradoxien und Pathologien?

Die dargestellten Bruckstücke zur "Evaluation der Evaluation" zeigen einige erklärungswürdige Widersprüche und Probleme im Feld der Evaluationsaktivitäten selbst, die uns jedenfalls vor einer Idealisierung dieser

Tätigkeiten und ihrer Repräsentanten warnen sollten. Wir haben gesehen, daß Evaluation eine komplexe Tätigkeit ist, die durchaus auch der Reflexion und Selbstreflexion unterzogen wird. Ein wichtiger Gedanke, den man in diesem Zusammenhang vielleicht in den Mittelpunkt rücken sollte, ist der, daß der Grundvorgang der Evaluation, nämlich die Bewertung von irgendetwas im Vergleich zu Bedürfnissen oder Zielen, bzw. im Vergleich zu Alternativen, ein in allen unseren Aktivitäten ständig vor sich gehender Vorgang ist, und daß auch gesellschaftliche Institutionen und Organisationen viele solche Bewertungspraktiken eingebaut haben. Die Evaluationsforschung, oder überhaupt professionelle Aktivitäten der Evaluation, sind in diesem Lichte als Reflexion dieser alltäglichen Praxis zu sehen, und als Versuche, diese alltägliche Praxis vielleicht ein wenig zu verbessern und zu objektivieren. Es geht also um die Weiterentwicklung unserer Bewertungsmethoden. Es ist klar, daß die Probleme und Widersprüche, die wir in unseren alltäglichen Bewertungen haben, auch in den professionellen Praktiken auftauchen. Und das allgemeine Paradox in diesem Feld ist vielleicht, daß die meisten dieser Probleme nicht durch die Verbesserung der Bewertungsmethoden zu lösen sind, obwohl eine naive Sicht dazu neigt, gerade dies zu erwarten. Wie die Erfahrungen im Bereich der Arbeitsmarktpolitik zeigen, besteht auch die Tendenz, in die ergebnisorientierten Ansätze der Evaluation höhere Erwartungen zu setzen als in die prozessorientierten Ansätze, obwohl viel dafür spricht, daß letztere eher in der Lage sein können, zur Weiterentwicklung der Praxis beizutragen.

4. Epilog: Evaluation von Arbeitsmarktpolitik in Österreich

Die Evaluation von Arbeitsmarktpolitik in Österreich wurde im Dymond-Bericht als sporadische Aktivität unterschiedlicher Qualität charakterisiert. Seitdem sind an verschiedenen Aspekten wichtige Fortschritte gemacht worden. Vor allem drei Gesichtspunkte sind zu nennen:

- Seit einigen Jahren wird daran gearbeitet, ein Monitoring System aufzubauen, das regelmäßige Informationsflüsse und Rückkoppelungsprozesse über die Arbeit im AMS ermöglichen soll, im Bereich der Beratungs- und Vermittlungstätigkeit gibt es darüber hinaus Ansätze für die Verwendung von Performance-Indicators.
- Im Bereich der Evaluation von Maßnahmen wurden wichtige Fortschritte bei der Entwicklung einer Datenbasis und der Anwendung der Methodologie von Vergleichsgruppen gemacht.

-- Schließlich ist die "externe" Organisationsanalyse des Institute for Policy Research zu nennen, die -- bei all ihren Problemen und Schwächen -- eine eindrucksvolle Gesamtdarstellung der Arbeitsmarktverwaltung gegeben hat, die als Referenz für weitere Fragen genommen werden kann. Gleichzeitig gibt es eine größere Menge an Evaluationen von bestimmten Maßnahmen mit unterschiedlicher Methodik und Reichweite.

Bisher ist es jedoch so, daß die Aktivitäten der Evaluation in die Politik nur wenig einfließen. Ein OECD-Review-Team über lokale Beschäftigungsinitiativen hat in seinem Bericht festgestellt, das AMS "does not seem to have a consistent programme of evaluation and only limited outcomes-testing policy to check the cost-effectiveness of its varied programmes".³⁶ Es wird auch eine Konzentration auf Vermittlungsquoten kritisiert, die zwar gewisse unmittelbare Aufschlüsse geben könnten, jedoch über längerfristige Wirkungen keine Auskunft geben. Im Lichte unserer vorstehenden Überlegungen ist dieser Darstellung zwar zuzustimmen, gleichzeitig ist jedoch an der Position der Reviewer eine starke Begrenzung auf ergebnisorientierte Evaluation zu erkennen.

Mit der europäischen Integration ist für die österreichische Arbeitsmarktpolitik auch die Möglichkeit der Förderung aus den Mitteln der Strukturfonds, insbesondere dem Europäischen Sozialfonds (ESF), eröffnet worden. Das Förderungskonzept folgt im Prinzip dem Ansatz der Programmbudgetierung und schließt einen ziemlich umfassenden Ansatz der Evaluation ein, der die österreichische EvaluatorInnenszene vor große Aufgaben stellen wird. Die Mittel werden auf der Grundlage von explizit formulierten und begründeten Programmdokumenten von der österreichischen Seite vergeben, die in einem ersten Schritt einer *ex-ante-Einschätzung* unterzogen werden, und dann sowohl begleitend im Hinblick auf die Durchführung, als auch *ex-post* auf die Wirkungen evaluiert werden müssen.

Es gibt jedoch verschiedene Möglichkeiten des Umganges mit diesen Prozeduren. Man kann sie als Chance der Weiterentwicklung und Rationalisierung der politischen Kultur sehen, aber man kann sie aber auch als lästige Pflichtübung, und als Spielwiese für die altbekannten politischen Intrigen und Krämereien sehen. Das zitierte Review-Team hat für den Bereich der Regionalpolitik sehr zielsicher eine bestimmte Gefahr hervorgehoben, die keinesfalls von der Hand zu weisen ist: "... the Structural Funds can be simply a redistributive measure unless they are allocated attentively. The principles that call for detailed regional planning

³⁶ OECD, Local Responses to Industrial Restructuring in Austria, Paris: OECD, 1995; S. 104.

concepts should not mask a simple international fiscal transfer mechanism."³⁷ Im Interesse der Weiterentwicklung der Arbeitsmarktpolitik sollten die nationalen Förderprogramme von der interessierten Öffentlichkeit ernstgenommen und im Hinblick auf die Durchführung auch im Auge behalten werden. Bei entsprechender Handhabung können sie auch als Ansatzpunkt dienen für den Aufbau eines zusammenhängenden Evaluationskonzeptes, für die Heranbildung einer EvaluatorInnenszene und für die Entwicklung von Mechanismen, die eine praktische Nutzung der Evaluationsaktivitäten befördern. Die Entwicklung von personellen Ressourcen für die Durchführung von Evaluationsaktivitäten wurde bereits vor einigen Jahren im Bericht über den *state of the art* als wesentlicher Faktor erkannt, als vielleicht noch wichtiger wurde jedoch eine entsprechende Einstellung in der Öffentlichkeit hervorgehoben, die Schlüsselfrage sei, daß "*evaluation gives rise to an adequate and informed public debate*"³⁸, etwas was für die öffentliche Diskussion von Arbeitsmarktpolitik in Österreich bisher nicht unbedingt erfüllt ist.³⁹ Michael Quinn Patton, einer der wichtigsten Theoretiker und Praktiker im Bereich qualitativer Methoden, plädiert jedenfalls stark gegen den Aufbau von engen Systemen standardisierter und vorgeschriebener Evaluationsprozeduren: "The establishment of an orthodox evaluation methodology is no different from the establishment of a state religion. Officially telling you what methods to use is only one step removed from officially telling you what results to find. At that point utilization of findings will cease to be an issue -- for there will be nothing to use, only orders to follow."⁴⁰

³⁷ Ebd. S. 105.

³⁸ OECD, *Evaluating Labour Market and Social Programmes*, a.a.O., S. 15)

³⁹ Zilian, H.-G. & B. Malle, *Spreu und Weizen. Das Verhalten der Arbeitskräftenachfrage*, Forschungsbericht, Graz: Büro für Sozialforschung, 1993.

⁴⁰ Michael Quinn Patton, *Qualitative Evaluation Methods*, Beverly Hills: Sage, 1980; S. 346.