

Evaluation als modernes Ritual¹

„Anstatt die Aufklärung, in welche die Religion übergegangen war, mit vollem Bewusstsein durchzuführen, die trügerische Freiheit der Revolution zur Gerechtigkeit voranzutreiben, hat die westliche Gesellschaft resigniert. Die Rückkehr zur Religion meint nicht, dass sie wieder an den Himmel glaubt, sondern dass es ihr zur besseren Einrichtung der Erde an Glauben gebricht, dass sie nichts mehr will als sich selbst.“

Max Horkheimer 1959 [1991]: 330).

„Wieso muss jetzt eigentlich immer alles evaluiert werden?“ so lautet die Frage, die mich vor drei Jahren zu meiner Doktorarbeit bewog. Und die Antwort darauf ist eigentlich ganz einfach. Sie steckt im Titel meines Vortrags: Evaluation fungiert heute als modernes Ritual inmitten einer durch und durch ambivalenten Rationalisierung². Wie komme ich nun dazu, ausgerechnet in einer Begutachtungsform wie der Evaluation etwas Religiöses zu sehen?

Um das etwas plausibel zu machen, werde ich in einem ersten Schritt von den Fallstudien erzählen, die ich über die inflationäre Ausbreitung von Evaluation angefertigt habe. Mittels Interviews, Dokument- und Diskursanalysen, teilnehmender Beobachtung und anderen ethnographischer Erkundungen in Hochschulen, Ministerien, Beiräten, Kompetenzzentren und Unternehmen habe ich mir die Wirkungsweisen von Evaluation genauer angesehen. Den Kreis der rund 40 Personen, die ich interviewt und beobachtet habe, verband, dass sie mit der Evaluation eines der beiden Hochschul-Projekte zu tun hatten, die ich für Fallstudien ausgewählt hatte.³ Bevor ich im zweiten Schritt zu den möglichen gesellschaftstheoretischen Erklärungsansätzen zunehmender Evaluationsrituale komme (2.), zunächst zu den empirischen Anhaltspunkten:

Was wird erkennbar über Nutzenbewertungen, wenn jemand, wie ich dies getan habe, Projekt-Zusammenhänge und deren Evaluationsgeschichten rekonstruiert und somit noch einmal aufrollt (1.)?⁴

¹ Vortrags-Manuskript. Es gilt die gesprochene Sprache.

² Rationalisierung ist hier im dreifachen Sinne gemeint: der technischen und gesellschaftlichen, aber auch der psychischen Verarbeitung von Veränderungsprozessen.

³ Dabei handelte es sich um staatlich geförderte e-learning-Projekte.

⁴ Analysiert wurden die Berichte und Meinungsbilder von mehr als 35 Personen verschiedener Akteursgruppen: Mitarbeiterinnen aus Ministerien, Kompetenzzentren, Beiräten sowie Gutachter und an dem zu evaluierenden Projekt beteiligten Hochschullehrende – Personen die heute im e-learning Jargon auch „Inhaltslieferanten“ oder „content provider“ genannt werden. In diesen Fallrekonstruktionen kam zwischen 2001 und 2002 ein Mix von Methoden zur Anwendung: Dokumentenanalyse, Leitfadeninterviews, Visualisierung zur Arbeitssituation der Befragten, teilnehmende Beobachtung sowie Forschungstagebuch (vgl. Schwarz 2004).

1. Quantität und Qualität von Evaluation

Schockierendes Ergebnis meiner Fallstudien war, dass sich die Evaluationsaktivitäten quantitativ so weit ausgedehnt hatten, dass zum Teil mehr Aufwand betrieben wurde, um etwas auf seine Wirksamkeit zu überprüfen, als das zu Überprüfende zu realisieren (z. B. eine Mitarbeiterinnen-Stelle über 3 Jahre zur Evaluation lediglich eines Seminars oder vier Personen, die ein Evaluations-Konzept für einen Seminarleiter erarbeiteten. Die beiden Hochschul-Projekte wurden von jeweils mindestens sechs Institutionen oder eigens für die Evaluation begründeten Teams ‚evaluiert‘ – und dies ohne dass diese Summe an Evaluationen auch nur einer einzigen Person im Projekt bekannt war.

➔ Folie: „Who evaluates whom?“: Dieses Projekt wurde nicht nur durch die Beteiligten selbst evaluiert („Selbstevaluation“), sondern auch durch das Projektmanagement sowie durch ein externes Drittmittel-Projekt (ein Habilitationsprojekt), ein Kompetenzzentrum, externe Gutachter und schließlich durch ein Hochschulentwicklungszentrum.⁵ Diese – sich uns erst im gemeinsam Forschungsprozess erschließende – Totalität von Evaluation sowie die Umkehrung ihres Kosten-Nutzen-Verhältnisses entfachte bei den Beteiligten zunehmend Entsetzen.⁶

Leben wir also, wie es der britische Wissenschaftssoziologe Michael Power (Power 1994, 1997) zugespitzt hat, in Gesellschaften, in denen immer mehr beobachtet und immer weniger gehandelt wird?

Die verunsichernde Qualität von Evaluationen

Nach Aufdeckung der wuchernden Quantität von Evaluation zeigt ein Blick auf deren Qualität, was sich hinter diesem eigentlich Legitimität zu schaffenden Verwaltungsinstrument⁷ verbirgt. Zur Vorwarnung: Evaluation kann interessanterweise genau das bewirken, was sie beseitigen sollte: Verunsicherung über die Legitimität öffentlicher Ausgaben.

Ich würde Ihnen jetzt gerne ausführliche anekdotische Beweise erbringen für meine These über Evaluation als wichtiges Sinn stiftendes Ritual in fragmentierten sozialen (Projekt-)Situationen.

⁵ Noch extremer waren die Evaluationsaktivitäten im zweiten Projekt, denn es beinhaltete ein eigens für Evaluationszwecke vorgesehenes „Teilprojekt“ mit vier verschiedenen fachlichen Abteilungen. Darüber hinaus wurde eine Gastwissenschaftlerin mit der Evaluationsaufgabe betraut, zusätzlich zu ihrer Dissertation, wegen der sie eigentlich ins Projekt gekommen war. Diese Summe zunächst als eher intern wahrgenommenen Evaluationen adiiert sich noch um Evaluationen, die als extern gedacht waren: ein Kompetenzzentrum, einen vierköpfigen Beirat sowie wiederum selbiges Hochschulentwicklungszentrum (in völlig anderer Personen-Zusammensetzung).

⁶ Spätestens jetzt könnte man sich fragen, was fällt denn hier alles überhaupt unter „Evaluation“? Gute Frage! Messe ich hier nicht vor allem die enorme rhetorische Ausdehnung des Labels „Evaluation“? Denn Evaluation ist ja offenkundig zu einer Hieroglyphe geworden für fast jede mehr oder minder systematische Betrachtung, die Legitimation schaffen soll. Indem ich quasi sinnverstehend das nachvollziehe, was die Akteure unter Evaluation fassen, untersuche ich also beides: die real beobachtbare Praxis von Evaluation sowie die sich aufblähende Rhetorik (zur Plastiksprache vgl. Pörksen 1986)

⁷ Evaluation wird hier vornehmlich als Form sozialer Kontrolle, als Verwaltungsinstrument betrachtet. Die beobachteten Mikropolitiken von Evaluation könnten allerdings auch als neue Kommunikationsformen untersucht werden. So würde die Systemtheorie das Phänomen vielleicht nur als Kommunikation sehen, die lediglich an Evaluation parasitiert. So interessant diese analytischen Zugänge auch sein mögen, hier wird Evaluation einmal vornehmlich als Element bürokratischer Herrschaft analysiert.

Doch muss ich mich begrenzen auf sechs Charakteristika der von mir rekonstruierten Fälle:

- 1.1 *Es bestehen große Interpretations- und damit Handlungsspielräume, wer sich überhaupt „Evaluator“ nennen darf und warum.* Unter der beschriebenen Vielzahl der Evaluierenden bekamen die vom Ministerium beauftragten Gutachter erst im Laufe der (kritisch verlaufenden) Projekte den Beinamen „Zwischen-Evaluation“. Das führte wiederum dazu, dass die qua Projektantrag ernannte „Evaluation“ sich in Opposition stellte zu den angeblich externen Evaluatoren, die aber mittlerweile begonnen hatten, mit einzelnen der zu Begutachtenden in anderen Projekten zu kooperieren. Wiederum andere, wie das Kompetenzzentrum, das sich als „Beratung“ darstellte und betontermaßen nicht als „Evaluation“, wurde aber von vielen Projektbeteiligten als „Evaluation“ (miss)verstanden. Damit erweist sich als dynamischer sozialer Konstruktionsprozess, wer als Evaluatorin gilt, aber auch, wer welche Rechte und Pflichten mit der (Selbst-)Ernennung als „Evaluation“ einhergehen.
- 1.2 *Welche Ziele mit Evaluation selbst verfolgt werden, entwickelte sich aus den Auseinandersetzungen der Beteiligten sowie aus den Dynamiken im Projekt und im gesellschaftlichen Verlauf.* In beiden Fallstudien konnte ein Changieren zwischen – teilweise unvereinbaren – Evaluationsfunktionen beobachtet werden: Evaluation lediglich als eine formalisierte Form der Selbstreflexion, eine Spielart von Forschung oder Aufklärung, als Beratung oder Moderation, Konsens- und Zielfindung der Projekte selbst, zum erneuten Austragen alter Konflikte oder schlicht als Selbstzweck. Damit befanden sich die Evaluierenden auf einem Terrain, das ideale Voraussetzungen für Rollenkonflikte bietet. Die meisten Evaluierenden versuchten diesem Dilemma zu entkommen, indem sie sich zu möglichst vielen Seiten als Berater gaben, so dass schon bald kaum mehr zu entschlüsseln war, wer eigentlich wen zu beraten imstande und willens war. Als kleinster gemeinsamer Nenner galt schließlich als Aufgabe von Evaluation, lediglich „Dialogfähigkeit“ über den Status quo herzustellen.
- 1.3 *Wie Wissenschaft findet auch Evaluation nicht im machtleeren Raum statt.* Das eher junge Team evaluierender sozialwissenschaftlicher Mitarbeiter wurde – trotz der ihrer Ansicht „sauberen Messmethoden“ wie Fragebogen und Experiment – für die Evaluierten entweder zum verschmähten Serviceangebot oder zur Rechtsanwältin, die Argumentationsketten liefert. Die eingeflogenen Gutachter urteilten eher auf Grundlage persönlicher Projekterfahrungen sowie des Meinungsaustauschs. Weniger als auf Grundlage von Daten oder systematischen Untersuchungen denn inszeniert als eine Art des Rates der Weisen (in der Evaluationsforschung gibt es dafür den ironischen Begriff „connaissance-Evaluation“), formierten sich die externen Gutachter – von der ratlosen politischen Administration zu Lichtgestalten des Evaluierens inthronisiert – zum mächtigen aber ungeliebten Tribunal, das sich mal behauptete wie Richter, Vorgesetzte oder gar Eltern.
- 1.4 *Nicht alles, was unter „Evaluation“ firmiert, bewährt sich also als nützliche Nutzenbewertung – zumindest nicht, was das unmittelbare Handeln angeht.* Im Vergleich zum Aufwand der um Evaluation betrieben wird, hatte sie relativ geringen Einfluss auf das praktische Geschehen. Neben einigen kleinen Entdeckungen und hilfreichen Tipps sahen viele den eigentlichen Nutzen der vielen Evaluationsarbeit vor allem in der Legitimation der politischen Administ-

ration, also dem Ministerium. Zwar wurde den Evaluierenden, die sich als dialogische Politik- oder Projektberater gaben, zunehmend Skepsis entgegen gebracht, doch hatte ihr Berateranspruch auch Hoffnungen auf ein weniger hierarchisches Wissensmanagement geweckt.

1.5 *Evaluationen wecken die Erwartung, dass bald – von oben – etwas passieren wird. Und diese geweckten Erwartungen auf Veränderung sind vielleicht der Hauptnutzen von Evaluation.*⁸

1.6 *Evaluationen können einen Teufelskreis gegenseitiger Versicherung und Verunsicherung etablieren statt Transparenz zu schaffen.* Im Zuge der Vervielfachung von Evaluation verflüchtigte sich auch deren institutionelle Zugehörigkeit. Evaluationsgremien oder -institute wurden schneller geschlossen als es dauerte sie aufzubauen, sei es, weil sie als zu sehr oder zu wenig mächtig und kompetent beurteilt wurden. Bürokratische Subsysteme formierten sich stets neu zu einem Dschungel von – weder intern noch extern – überschaubaren bürokratischen Entscheidungs-Vertagungs-Prozessen (EVP). So verwischt auch die Spur bürokratischer Zuständigkeit überhaupt und damit von Rechten und Pflichten oder auch nur Spielregeln.

Sie könnten jetzt sagen, das sei ja nur eines unter vielen Beispielen. Das stimmt zwar. Aber es bleibt ein real existierendes Beispiel! Der Zusammenhang zum Religiösen wurde hier immer deutlicher: Zeremoniell wird Evaluation, weil man sich fragen muss, was der Nutzen ist, der den vielen Aufwand rechtfertigen könnte. Evaluationsrituale ermöglichen zeitweise die Illusion eines Konsenses, indem die Einzelnen Entscheidendes verschieben: auf bürokratische Parallel-Universen, auf das Votum aller, auf später – vielleicht nicht ganz bis ins Jenseits wie in der Religion – aber doch auf: immer später.

Diese entschleunigende, sich Hintertüren aufhaltende Funktion von Evaluation stimmten fast alle Befragten zu („Wenn keiner mehr weiß wo es lang geht, dann evaluiert man heute erstmal ausführlich.“) Ist also Evaluation zum allseits geduldeten Ritual geworden, an das aber gar niemand glaubt? Zur Klärung der Ritual-Frage werde ich nun die soziologische Literatur konsultieren. Glasklare Antwort hat sie mal wieder nicht zu bieten. Aber eine Fülle von Hinweisen zum Zusammenhang von Bürokratie und Alltagsreligionen. Die von mir im Folgenden konstruierte Konstellation von Autorinnen ergibt eine – gewissermaßen akkumulierte – Kritik instrumenteller Vernunft.

⁸ Allein die Beobachtung vom Menschen bewirkt Reaktionen (Stichwort „reaktive Effekte“ oder „Hawthorneffekt“). Jedoch nicht zwingend als Verbesserung der Arbeitsleistung, sondern in der *Erwartung* einer Veränderung von oben. Entgegen dem evaluativen Berater-Selbstverständnis wirkten Evaluierende, als Repräsentanten dieses oben, auf die Evaluierten eher desinteressiert bis ignorant, autoritär bis manchmal sogar selbstüchtig, weil sie für die einzelnen Projektteile nie genügend Aufmerksamkeit zeigten. Die mangelnde Handlungsrelevanz von Evaluationen wird dagegen im Ministerium mit der Langsamkeit des Evaluationsvorgehens rückblickend entschuldigt. Häufig müsse man „republikanischer Effizienz“ den Vorzug geben gegenüber „demokratischer Legitimation durch Evaluation“, so ein Mitarbeiter des die Evaluation in Auftrag gebenden Ministeriums. Diese von vielen vollzogene psychische Rationalisierung täuschte aber nicht darüber hinweg, dass den meisten gleichzeitig die Macht des mikropolitischen Spielballs Evaluation bewusst war.

2. Wie ist der Evaluationsbooms zu fassen? Soziologische Erklärungsansätze

Die hier zu Tage geförderte Verselbständigung bürokratischer Prüfverfahren erinnert stark an Max Webers Horrorszenerario vom „stahlharten Gehäuse“ (Weber 1920: 203, 1958) der Hörigkeit unter das Kosten-Nutzen-Optimierungsprinzip. In diesem Käfig werde die gesamte Menschheit nach und nach eingekerkert von Bürokratie und Effizienzdenken. Als „stahlhartes Gehäuse“ der Hörigkeit kündigte Weber das Schicksal an, in das die kapitalistische Rationalisierung letzten Endes führe.⁹ War die materielle Sorge im abendländischen Denken nur ein „hauchdünner Mantel“ des asketischen Protestantismus, so wurde das Effizienzdenken zum „stahlharten Gehäuse“. Laut Weber führe der global alternativlos gewordene Kapitalismus in ein Gefängnis, das bewohnt ist von Kreaturen, die ihre „mechanisierte Versteinerung, mit einer Art von krampfhaftem Sich-Wichtig-Nehmen“ verbrämen. Webers stahlhartes Gehäuse der Rationalitätshörigkeit steht für den Verlust des souveränen bürgerlichen Individuums: von nun an erwarten die Individuen, von der Zweckrationalität Zwecke genannt zu bekommen, statt sie selbst zu setzen.¹⁰

Webers Prognose kapitalistischer Rationalisierung konfrontiert uns mit einer beängstigenden Apokalypse: als „mächtigen Kosmos der modernen, an die technischen und ökonomischen Voraussetzungen mechanisch-maschinelles Produktion gebundenen Wirtschaftsordnung (...), der heute den Lebensstil aller einzelnen, die in dies Triebwerk hinein geboren werden – nicht nur der direkt ökonomisch Erwerbstätigen –, mit überwältigendem Zwange bestimmt – und vielleicht bestimmen wird, bis der letzte Zentner fossilen Brennstoffs verglüht ist“ (Weber 1920 [1988]: 203).

Mit Weber könnte das Anwachsen von Evaluation also begriffen werden als lediglich eines der Symptome durchgedrehter Zweckrationalität, die sich in ihrer eigenen Kosten-Nutzen-Bilanz verrannt hat. Betrachten wir den Evaluationsboom nun als Teil expandierender bürokratischer Verwaltungssysteme, ergeben sich verblüffende Parallelen zu Webers Theorem unaufhaltsamer Ausbreitung zweckrationalen Abwägens: Evaluiert wird nicht mehr für den Menschen, sondern der Mensch ist scheinbar für das Evaluieren da. Meine Untersuchungen haben aber ergeben, dass die Evaluationitis gar kein kalter grauer Arm bürokratischer Herrschaft ist, der uns hier in Würgegriff genommen hat. Das Evaluations-Symptom bedeutet, dass Evaluation zu einem mikropolitischen Spielball wird, wenn nicht sogar zum Faustrecht in Verteilungskämpfen.¹¹

⁹ Die dramatische Zukunftsvision vom stahlharten Gehäuse geistert nicht nur im sozialwissenschaftlichen Studium herum, sondern ist auch sehr beliebt in politische Ansprachen – wenn auch mit unterschiedlichsten Intentionen und Interpretationen. Statt des „stahlharten Gehäuses“ findet sich häufig auch die Metapher des „eisernen Käfigs“. Sie stammt vermutlich von der Rückübersetzung von Talcott Parsons englischer Übersetzung des Weberschen Begriffs des „stahlharten Gehäuses“ als „iron cage“.

¹⁰ Weber sah wenig Chance, „überhaupt noch (...) irgendwelche Reste einer in irgendeinem Sinn 'individualistischen' Bewegungsfreiheit zu retten“ (Weber 1984: 465f). Die Verselbständigung zweckrationaler Vernunft könne laut Weber lediglich eingedämmt werden durch kritische Wissenschaft, die eine „Flucht nach vorn“ antreten kann statt zu „desertieren“.

¹¹ Interessant ist der Vergleich heutiger Evaluations-Symptomatik mit der Entstehungsgeschichte von Evaluation während des *New Deals* in den USA der 30er Jahre. Ging es damals darum, wo das Geld für den Wohlfahrtsstaat am besten ausgegeben werden kann, so geht es beim heutigen Abbau des Sozialstaates darum, wo das Geld am besten gespart werden könne.

Wir könnten das hier beobachtete Evaluationsphänomen aber auch mit Th. W. Adornos Worten begreifen: „Das Anwachsen der Quantität von Verwaltungsapparaten hat eine neue Qualität erzeugt“ (Adorno 1969: 122). Doch worin besteht diese neue Qualität? Nach Angaben der Kalifornisch-Frankfurter Schule ist die kapitalistische Herrschaft deshalb totalitär, weil in ihr nur noch etwas von Verwaltung freigehalten wird, wenn es dazu hilft, alles andere zu verwalten.

„Gerade weil es dubios bestellt ist um die Nützlichkeit des Nützlichen, ist es dem Apparat doppelt wichtig, sich als ein Nützliches, um der Konsumenten willen Ablaufendes zu repräsentieren. Darum wird in der Ideologie die Demarkationslinie von Nützlichem und Unnützem so streng gezogen.“ (Adorno 1969: 129) (...) „Kultur soll durchaus unnützlich, darum auch jenseits der Planungs- und Verwaltungsmethoden der materiellen Produktion sein, damit der Rechtsanspruch des Nützlichen ebenso wie der des Unnützen umso mehr Relief gewinnt“ (ibid. 130).

In Adornos Gesellschafts- und Erkenntniskritik ist es nicht nur die Waren- sondern auch die Medienvergesellschaftung, die Unterschiede produziert, vermittelt und wiederum – technisch, psychisch und gesellschaftlich – verarbeitet zu neuen Grenzziehungen: Nützlich – Unnützlich, Richtiges – Falsches, Gutes und Böses.

Die entzauberte Welt ist zwar beherrschbar und damit zerstörbar geworden. Doch nicht beherrschbarer Rest sind die Verinnerlichungsprozesse rationaler Lebensführung sowie die Ausweitung bürokratischer Verwaltung. An diesem nicht-beherrschbaren aber doch allzu leicht verführbaren Rest des bürgerlichen Individuums setzt die Weber-Rezeption kritischer Theorie an, für die Weber selbst noch viel zu rationalistisch war.¹² Die Menschen wirken zwar eifrig mit an der Ausweitung von Verwaltung. Doch irgendwie sehen sie bei der Verselbständigung instrumentellen Denkens doch nur hilflos zu – ferngesteuert und im Halb-Bewusstsein eigener Unfreiheit. Eine Geisterstimmung, ein Gruselkabinett, wie es massenhaft im Plot aller aktuellen Science-Fiction-Kassenschlager reproduziert wird – und damit eben nicht verarbeitet aber doch vorverdaut wird.¹³

Mit rationalisierungskritischem Zugang zu subversiven Wissenschaftsritualen wie z.B. Evaluation erschließt sich uns der Zusammenhang von Schein und Sein der Verwaltung: subversiv waren die Praktiken z.B. der Akteure, die sich des Labels „Evaluation“ so bedienten, dass sie damit ihre Interessen verwirklichen konnten, Jobs für den Nachwuchs ermöglichen, Qualifikationsarbeiten schreiben, persönliche Netzwerke erweitern und vieles mehr. Nicht alle unterwarfen sich der Evaluation zur Gänze, sondern verbanden sie mitunter trickreich mit eigenen Zwecken. Evaluation dient quasi als „Alltagsdieterich“ (vgl. Pörksen 1986) zwischen Experten und Laien. Das scheinbar allzu Plausible, Standardisierte und Vernünftige ist immer auch Tummelplatz auch für Ungleichnamiges, Oppositionelles, Widerspenstiges und Subversives.

¹² Nach Adorno hätte Weber in seiner Rationalitätskritik den fatalen Fehler begangen, sich selbst genau das zu Eigen zu machen, was er analysiert hat. Er ist streckenweise der Verführung erlegen zu glauben, es gehe zweckrational zu und sei demnach auch rational zu erklären. Der von Weber „unterstellte Rationalitätsbegriff“ sei selbst „beschränkt auf die Zweck-Mittel-Relation“ (Adorno 1969: 124) und deshalb sein „Urteil über die Rationalität der Zwecke selbst behindert; in Webers eigener Rationalitätstheorie mag man den Niederschlag von Verwaltungsdanken argwöhnen.“ Bezogen auf den Evaluationsboom würde dies dazu ermahnen: Nicht die Rationalitätsansprüche zu reproduzieren, die wir gerade eben in Frage gestellt haben.

¹³ Die Plots von Science-Fiction-Filmen wie „Matrix“, „Alien“ oder „12 monkeys“ haben verblüffende Ähnlichkeit mit Webers gesellschaftlicher Apokalypse: die einzelnen wissen um den Verlust ihrer Handlungsspielräume Urteilkraft, müssen diesem Schauspiel aber selbst – noch im Besitz aller Sinne – zusehen.

Ungegläubte Rituale haben eine konformatitätsstiftende Kraft und sind somit gesellschaftlicher Kitt. Wie Alltagsreligionen (vgl. Claussen 1987, 1994, 2000) balancieren sie Dissonanzen zwischen verschiedenen Bewusstseins- und Ernsthaftigkeits-Stufen. Nun mag es Alltagsreligionen *immer schon* gegeben haben. Aber von stets neuer Qualität ist das beharrliche Anwachsen eines Apparats, um die Ideologie einer gerechten Marktwirtschaft wenigstens notdürftig am Leben zu erhalten. Vom ungegläubten Aufklärungsversprechen ist nicht mehr viel übrig als sein immer wilder werdendes ideologisches Getriebe – und möglicherweise kann der Evaluationsboom als kleines Rädchen in diesem Getriebe verstanden werden, Glaubhaftigkeit in immer ungläublicheren Zeiten zu schaffen?

Wir könnten aber den Aufstieg der Evaluation auch als einen Prozess der Entpersönlichung von Vertrauen verstehen, wie dies z. B. der britische Wissenschaftssoziologe Theodor Porter in seinem Buch „Trust in Numbers“ illustriert hat. Anhand vieler Beispiele wissenschaftlich-politischer Kontroversen beschreibt Porter die zunehmende Durchdringung von Quantifizierung als einen Prozess, in dem Zahlen für Wirtschaft, Politik und Wissenschaft immer unverzichtbarer werden und zwar vor allem aus einem Grund: weil man mit Quantifizierungen äußerem Legitimationsdruck begegnen kann. Wenigstens kurzfristig wirken Quantifizierungen als entpersönlicht und damit als vertrauenswürdig, v.a. da, wo Personen sich als unsolide erwiesen haben. Evaluationen, die stark von Quantifizierungen leben, könnten demnach Indikatoren für angeschlagene Vertrauensverhältnisse sein: wo bisherige Eliten in Misskredit geraten sind oder ehemals geduldete informelle Regeln verdächtig geworden sind. In Zeiten des Misstrauens nehmen Zahlen fatalerweise eine wichtige Stellvertreter-Funktion für persönliche Glaubhaftigkeit. Was die Zahlen aber tückischerweise von Menschen unterscheidet: Zahlen sind nicht haftbar!

Diese Perspektive bringt uns noch einen Schritt weiter auf dem Wege der Entlarvung von Evaluation als modernem Ritual: Evaluation kann auch als Instrument verstanden werden, Vertrauen zu schenken und zu entziehen.¹⁴ Aber gab es diese Indienstnahme von Wissenschaft zur Legitimation von Politik und Management nicht immer schon?

Nein, sagen auch die beiden Autorinnen, die sich ähnlich intensiv mit der kritischen Analyse des Evaluationsbooms beschäftigt haben. Michael Power von der London School of Economics sowie die britische Sozialantropologin Marilyn Strathern qualifizieren den Evaluationsbooms als „checking gone wild“, als „Verwaltungsregimes¹⁵, als Evaluations- und Buchführungs-Diktaturen (Power 1997: xii). Seit Ende der 80er Jahre habe sich etwas Grundlegendes verändert im Bewertungsverhalten der Menschen, so Power und Strathern in den Büchern „The Audit Society“ sowie „The Audit Culture“.

„Die ‘audit society’ ist Symptom dieser Zeit, zufälligerweise zum fin de siècle. Es hat sich eine Kluft geöffnet zwischen wenig geschätztem Handeln und viel beachtetem Beobachten. Mit diesem Buch habe ich lediglich versucht, diese wachsende Industrie von Behaglichkeitsproduktion etwas unbehaglicher zu machen“ (Power 1997: 147, Übers. cs).

¹⁴ Vertrauen ist eine endliche und damit sparsam zu behandelnde Ressource menschlicher Interaktion. Vertrauensverlust hat aber nicht nur zerstörerische sondern auch heilsame Folgen.

¹⁵ Der Begriff „audit“ bedeutet Rechnungsprüfung bzw. Buchungsprüfung und ist damit wie Evaluation ein Instrument zur Rechenschaftslegung („accountability“). Wie bei Power und Strathern werden auch hier „audit“ und „evaluation“ synonym verwendet.

Power betrachtet Gesellschaften als Konstellationen von Vertrauen, Risiko und Rechenschaft, in denen es niemals eine komplette Sicherheit geben wird, noch eine völlige Abwesenheit von Kontrollverfahren. Lediglich die Inspektionsstile verändern sich gesellschaftshistorisch und auch die Art und Weise, wie Risiken wahrgenommen, definiert und gemindert werden (ibid. 138). Regelrecht „attraktiv“ werde Evaluation als Ritual scheinbarer Kontrolle von Risiko und Unsicherheit (ibid. 140), weil sie hilft, die Einsicht in die wahre Komplexität zu vermeiden (ibid. xvii). Denn um verwickelte Zusammenhänge wie Umwelt- oder Bildungspolitik evaluierbar zu machen, müssen sie aus ihrer gesellschaftlichen Komplexität herausgerissen werden. Absurderweise verschließe man genau damit aber die Augen vor den Risiken, in die man sich begäbe. Aus diesem Grunde sei auch die Etikettierung und das Reden über Evaluation (die Rhetorik des „quality label style“) mindestens ebenso wichtig, wie zu handeln – das hätten die Fallstudien fast nicht besser zeigen können. Als Charakteristika des *common sense* in der „audit society“ nennt Power einerseits die Abgewöhnung von Skepsis und Reflexion und ein Eingewöhnen in eine „gelehrte Ignoranz“ all der empirischen Hinweise, die nicht ins Visier des Evaluationsschemas passen. Das größte Risiko läge aber darin, Evaluation derart zu vertrauen und damit sich schon die ganz intuitive Skepsis abzugewöhnen (ibid. 144: „closed avenues of official scepticism and modesty“) – etwas was der Philosoph Horkheimer Verdunklung der Vernunft genannt hat.

Mit dem zu Ende gehenden 20. Jahrhundert stimuliere Evaluation lediglich die historisch offenbar nie zu Ende gehende Suche nach der ‚Golden Goose of effectiveness‘ (ibid. 141, Zitat Cohen 1985: 115), so Michael Power 1997. Der heutige Evaluationskult sei also lediglich ein bisher ungekannter Höhepunkt auf der Hetzjagd nach Effizienz:

Ähnlich wie bei Weber – nur ohne eine einzige Referenz – lautet Powers beunruhigende Diagnose, dass die Mess-Manie zum wichtigen Bestandteil des sozialen Gewebes (ibid. 142: „social fabric“) geworden ist. Power protestiert vor allem so vehement gegen diese Behaglichkeits- und Imageproduktion neuer Management- und Regierungs-Stile (143). Der Evaluationsboom wäre dann eine Argumentationsressource an der sich heute jeder ein bisschen bedient und nicht nur die Bürokraten. Also ganz normale Ideologieproduktion durch Wissenschaft, aber nur dezentraler und unbekümmerter als je zuvor?

Wie Michael Power übt auch Marilyn Strathern scharfe Kritik an der wachsenden Beliebtheit neo-liberaler Doktrinen inmitten der akademischen Welt. Inszeniert sich der naive Wissenschaftler inmitten des Sozialabbaus zur kurzfristigen Überlebensstrategie als Berater, wird er langfristig nur ein Körnchen des Kanonen-Futters politischer Administration.

Aus Sicht der Sozial-Anthropologin ist das kulturell Neue: die Verschmelzung von Finanzwirtschaftlichem mit dem Moralischen. Wie in einem Zaubertrick erscheint so ökonomische Effizienz als vereinbar oder geradezu identisch mit ethischem Handeln. Diese gegenseitige Legitimation von Moral und Effizienzdenken habe fatale Konsequenzen auch für die eigene intellektuelle Arbeit, denn durch ihr Mitwirken an der Image-Produktion von Entscheidungsträgern entmündigten sich Akademikerinnen selbst. Die Rituale der Verifizierung etablieren heute einen *common sense*, indem das als relevantes Wissen gilt, was den individuellen Erfahrungshintergrund mit

Einverständnis der Beteiligten abschneidet – so jedenfalls Strathern.

Paradoxerweise ist die Explosion von staatlichen Verwaltungsvorschriften zugleich ein Phänomen der Entlassung und Vereinnahmung des Individuums im Post-Sozialstaat. Evaluation verstärkt damit die politische Doktrin individueller Rentabilität: „Buchführung und Evaluation wird zu einer Form der Selbstregierung“ (Strathern 2000: 289). Dieser Rückzug des Staates werde nach Strathern mit Ritualen der Nachweisbarkeit lediglich eingeleitet, die dann nach Verinnerlichung des selbstevaluativen Imperativs selbst unnötig werden. Die sozial-anthropologische Sichtweise enthüllt die gesellschaftshistorische Brisanz des Evaluationsrituals: Konformismus höhle das Subjekt aus, weil die Einzelnen in ihre Entmündigung noch einwilligen: „Wie ein Ritual versuchen Buchführung und Evaluation die Menschen davon zu überzeugen wie die Welt sei, ohne ihre ganz persönliche Sichtweise“ (Strathern 2000: 287).

Global setze sich eine Triade durch, in der sich die drei Domänen ehemaliger Staatlichkeit in einer radikal-marktwirtschaftlichen Ideologie gegenseitig aufeinander beziehen: Politik, Rechenschaftslegung und Ethik.¹⁶ Der Evaluationsboom ist bei Strathern lediglich ein Begleit-Phänomen des sich zurückziehenden Sozial-Staates und der das Individuum entlässt in die Verstrickungen selbst-evaluativer Vernunft, also eine Verinnerlichung des Ex-Sozialstaats.

Spätestens hier könnte man mir Dämonisierung vorwerfen. Massenhaft ließen sich Gegenpositionen zur apokalyptischen Deutung des Evaluationsbooms finden, vor allem in der amerikanischen Evaluationsdebatte.¹⁷ So etwa die enorme Masse an Literatur, die das glatte Gegenteil meiner Ausführungen behauptet: als responsive oder partizipative Evaluation sei sie wichtiges Element von Emanzipation und Demokratisierung. Ein – besonders unter Linken – anerkannter Evaluationstheoretiker etwa, Michael Patton versteht unter „evaluativem Denken“ sogar „kritisches Denken“ (Patton 2002). Evaluation wird ihrem Demokratisierungsversprechen nicht gerecht, trotzdem gibt es – mit Ausnahme von Power und Strathern – zu diesem *empowerment durch Evaluation*-Diskursen bisher kaum Gegenstimmen.

Der rote Faden, der sich aber durch hier diskutierte sozialwissenschaftlicher Arbeit zieht, ist: Die Verinnerlichung bürokratischer Herrschaft als Reaktionsweise der Einzelnen auf die Widersprüche, die ambivalente gesellschaftliche Rationalisierung produziert. Aber die vom mir zitierten Autorinnen haben auch allerlei Zweifel gesät an der Härte des Weberschen Gehäusestahls. Notwendig unvollständig und überholt ist Webers Metapher, denn sie entstammt dem Industriezeitalter.¹⁸ Die Metapher des Stahls für äußeren Arbeitszwang und innere Hörigkeit ist der Fabrikar-

¹⁶ Das erinnert an die Einheit von Wirtschaftsethik und Lebensführung, wie sie Weber beschrieben hat. Lediglich in einer Fußnote stellt Strathern einen vagen Bezug zu Webers Bürokratie-Kritik her (Strathern 2000: 284). Bei Power findet Weber keine Erwähnung, auf meine persönliche Nachfrage hin, antwortete dieser, er hätte Weber lediglich "unbewusst" verarbeitet. ... let's have that audited ...

¹⁷ Eine kritische Aufarbeitung dieser Demokratisierungs-Ansätze – v.a. in Gegenüberstellung mit dem Weber-Theorem – steht aber leider noch aus.

¹⁸ Die Unterscheidung zwischen Industriegesellschaft und Medienvergesellschaftung bekräftigt hier nicht die ein-

beitswelt entlehnt, sie müsste angesichts veränderter globaler und lokaler Bedingungen heutigen Rationalisierungsdrucks ohnehin neu betrachtet werden. Die in Büro-Etagen Erwerbstätigen sind von anderen Zwängen bestimmt als die Menschen im Stahlwerk. Die dauerhaft von Erwerbsarbeit Ausgeschlossenen sind mit einer anderen Form von Herrschaft konfrontiert als die permanent – von innen oder außen – zur Arbeit Angetriebenen.¹⁹

In den hier beschriebenen mikropolitischen Auseinandersetzungen hatte Evaluation auch sehr unbürokratische und ungeordnete Seiten. Evaluation ist nicht die verlängerte Exekutive eines kalten, grauen, entpersonifizierten Kosten-Nutzen-Kalküls. Evaluation wird viel dezentraler und verstreuter praktiziert. Sie ist ein Spielball, der in eine Arena von Akteuren rollt. Verwaltung erweist sich somit als heterogener, lebendiger, trickreicher und subversiver, als Weber dachte.

Aus welchem Material, in welchem Zustand ist das Gehäuse der Verwaltungshörigkeit also heute?

3. Evaluation als modernes Ritual

Als Verwaltungsinstrument ist Evaluation keine mechanische Versteinerung, sondern eher ein flexibler Modus sozialer Kontrolle. Danach wäre Evaluation Teil einer kollektiven Strategie, gesellschaftliche Rationalisierung kollektiv zu verarbeiten – aber vielleicht unberechenbarer als dies bisher der Fall war. Rationalitätshörigkeit ist somit gar kein „stahlhartes Gehäuse“, sondern ein viel flexibleres Gefängnis, eher wie ein Irrgarten. Im Irrgarten sind es keine Stahlwände, die den Freiheitsentzug bedeuten, sondern die Verwirrung durch die vielen Wahlmöglichkeiten an Wegen, die das Individuum als „Entscheidungsträger“ bekommen hat. Wie die unzähligen Gänge in einem Irrgarten haben sich die Ausbruchs- und Einlösungsversuche der Aufklärung vervielfacht, egalitäre Chancen auf Glück herzustellen.

Evaluation als modernes Ritual zu qualifizieren, bedeutet nicht, sie sei bloß „sinnentleerte Scheinaktivität“ (Schwermer 1999). Moderne Rituale als kollektive Handlungen dienen zur Verständigung und Vergewisserung über geteilte Werte und versuchen, soziale Sinnhaftigkeit herzustellen. In Ritualen gelingt diese Sinnproduktion, indem auf etwas Höheres, außer ihr selbst liegendes, also Überindividuelles verwiesen wird: die Autorität des Wissenschaftlers, das anonyme Prüfverfahren oder die (scheinbar) interesselose Zahl, die Signifikanz, das akkumulierte Gutachterurteil, dem Konsens oder dem *common sense* oder einfach die nächst höhere (bürokratische) Instanz. Diese als eigene Züchtigung erscheinende Hörigkeit gegenüber dem *common sense* ist viel flexibler als simpler Stahl.

Damit stellt sich die Frage: Kann bürokratische Herrschaft (bzw. Organisation) weiterhin als anderen (traditionalen) Herrschaftsformen überlegen gelten, weil sie die effizienteste Form der

schlägige These von der „Wissens- oder Informationsgesellschaft“. Hingewiesen sei damit lediglich auf die vielen Unterschiede zwischen heutigen Lebens- und Arbeitsbedingungen und denen Anfang des 20. Jh.

¹⁹ Bürokratische Herrschaft befindet sich heute vielmehr in einem Stadium der Verwaltung erwerbsarbeitsloser Körper (Zygmund Baumann hat diese Produktion menschlicher Ausschussware als „human waste – wasted humans“ bezeichnet.)

Gewalt ist, wie Weber dies postuliert hatte? Ich meine, diese flexibilisierte bürokratische Herrschaft (bzw. Organisation) ist deshalb so effizient, *gerade weil* neue Formen sozialer Kontrolle wie Evaluation so unverbindlich und flüchtig sind. Evaluation ist lediglich eine heute (noch) zeitgemäße Art und Weise, Legitimität einer Ordnung zu konstruieren – auch und gerade, wenn die Ordnung wegen ihrer Widersprüchlichkeit zunehmend für illegitim gehalten wird. Flexibilität ist zum abstrakt-flüchtigen Leitbild geworden – und auch sie muss evaluiert werden. Im kontinuierlichen Verlangen nach Flexibilität drückt sich einerseits das Bedürfnis nach Veränderung aus, und zugleich macht es anschlussfähig. Gerade weil Evaluation so ein unberechenbares Verwaltungsinstrument ist, erweist sie sich heute als so wirkungsvoll: Sowohl mit der Ankündigung als auch mit dem Vollzug von Evaluation legt man sich nicht wirklich fest, sondern bleibt allumfassend flexibel für die Reaktionsweisen anderer. In demokratisch strukturierten Gesellschaften kann so ganz situativ die Legitimität bürokratischer Herrschaft hergestellt oder entzogen werden, ohne inhaltliche Auseinandersetzungen riskieren zu müssen. So bleibt Evaluation zwar ein problematisches – aber inmitten einer Ökonomie trügerischer Versprechungen – unabkömmlich gewordenen Ritual in demokratisch strukturierten Gesellschaften.

Die heutigen Mechanismen der Verantwortungsdelegation, die hier als Verwaltungsregimes beschrieben wurden, stimmen nicht viel hoffnungsvoller als Webers stahlhartes Gehäuse der Hörigkeit. Auf die eingangs gestellte Frage nach der beobachtenden statt handelnden Gesellschaft, ließe sich entgegnen: es wird zwar weiterhin auch gehandelt, nur nicht so planmäßig und effizient, wie es der allgemeine Evaluations-Imperativ unterstellt. Evaluation dient zur prophylaktischen Vorbereitung von Handeln und zugleich zu seiner Verhinderung. Evaluation schafft zwar Vertrauen, aber sie nährt das Misstrauen gleichermaßen. Im Prozess der Entpersönlichung von Vertrauen erfüllt Evaluation eine amalgamierende Funktion zwischen zweckrationalem Konformismus und wachsendem Misstrauen. In Konfliktsituation eignet sich Evaluation auch deshalb so gut als kollektiv geteilte und zugleich kritisierte Plattform, weil sie keinen Inhalt transportiert.

Mittlerweile formiert sich starkes Misstrauen²⁰ gegenüber Evaluation. Es etabliert sich schon so etwas wie eine Anti-Evaluations-Attitüde (z.B. eine Artikelserie in der Zeit im letzten Sommer), sogar unter denen, die ihr täglich Brot ‚einwerben‘ mit Evaluationsaufträgen. Bitte missverstehen Sie mich aber nicht und ordnen Sie mich bitte nicht ein in dieses Anti-Evaluierer-Lager. Es ging mir nicht darum zu klären, ob Evaluation richtig oder falsch ist, sondern etwas Sensibilität zu schaffen für die ambivalente Funktion, die sie inmitten von Rationalisierungsprozessen einnimmt, die als „Innovationen“ verkauft werden.

²⁰ So haben sich die Interviewten den Evaluationsbooms erklärt: Die einen sehen in der Evaluation einen Ausdruck von Demokratisierung und das Versprechen wissensbasierterer Politik, manche kleiden diese Hoffnung sogar in die Worte, dass jetzt endlich mit der Faulenzerei Privilegierter „Schluss gemacht wird“. Andere sehen darin simples Marketing von Entscheidungsträgern, die wissen wollen, „wo sie ihr Geld denn nun hinpumpen sollen“ ohne sich öffentlich zu blamieren. Politiker verlagern ihre normativen Entscheidungen auf Wissenschaftler, deren scheinbar analytisches Urteil sie in der politischen Praxis wie von Geisterhand wieder in ein normatives verwandeln. Manche erkennen in der Evaluation also ein „Geschäft mit der Sicherheit“, dass man gut mit verunsicherten Politikern und Managern machen kann. Wiederum andere sehen im Evaluationswahn eine notwendige Begleiterscheinung des sich selbst abbauenden Sozialstaats. Andere fassen ihre heutige Gesellschaftsdiagnose sogar mit einer Zahl zusammen: 1989. Mit dem Ende der Systemalternative von Kapitalismus und Sozialismus habe sich die Welt einer einzigen ökonomischen Totalität vereinheitlicht.

Aus vielerlei Gründen sehen sich die Gesellschaftswissenschaften heute leider eher außerstande, die Hintergründe derzeitiger „Innovationen“ zu ermitteln. Was der soziologisch inspirierten Wissenschafts-Parodie immerhin noch bleibt, ist: wenigstens die Ineffizienz und die Kosten dieser so genannten Innovationen zu illustrieren – ganz gleich welches Werturteil Sie als „Entscheidungsträgerinnen“ daraus ziehen.

Als ungeglaubtes Ritual bewirkt Evaluation meines Erachtens vor allem eines: als ökonomisch vereinnahmte Technik beschleunigt sie die gesellschaftliche Erosion von Vertrauensverhältnissen.

Ich bedanke mich für Ihre Aufmerksamkeit