

Friedhelm Neidhardt

Über die Kunst der Selbstverteidigung

Und dass sie manchmal nichts nützt

Mehr und mehr wird noch und noch evaluiert: die Kreditwürdigkeit von einzelnen Unternehmen und von ganzen Nationen (zum Beispiel mit den Ratings von S&P und Moody's), die Exzellenz von Köchen (zum Beispiel mit den Sternen von Michelin), die Reputation von Politikern (zum Beispiel mit den Prestigeskalen der Forschungsgruppe Wahlen), seit ein paar Jahren nun auch die Qualität von Universitäten (zum Beispiel mit dem Spiegel-Ranking). Moderne Gesellschaften entwickeln sich offensichtlich in Richtung Stiftung Warentest.

Diese Entwicklung mag erstaunen, denn sie vollzieht sich gegen die Komfortbedürfnisse der Bewerteten. Und deren Unbehagen lässt sich auch gut verstehen. Evaluation läuft doch darauf hinaus, dass andere einem auf die Finger schauen und das, was sie dann sehen, bewerten und allen kundtun. Da sich nicht wissen lässt, was dabei herauskommt, wird Evaluation keinem, den es trifft, von vornherein gefallen können. Dies zumal dann nicht, wenn dem Vorgang die Drohung beigefügt ist, man verlöre Geld, Rechte, jedenfalls Reputation, falls sich der Eindruck ergäbe, dass man selber nicht fleißig oder aber nicht talentiert genug sei, um etwa so Gutes zu Stande zu bringen, wie es anderen vermeintlich gelingt. Man wird also nachvollziehen können, dass sich bei den Betroffenen zuerst einmal der Abwehrreflex durchsetzt: Evaluation sollte es gar nicht geben.

Nun muss sich solch ein Abwehrreflex natürlich erklären, sobald er sich öffentlich behaupten will, und es ist interessant zu sehen, auf welche Weise dies geschieht – zum Beispiel bei den Professoren, die seit einigen Jahren zunehmend unter Evaluationsdruck geraten. Auch

ihnen will man das Menschenrecht auf Selbstverteidigung natürlich nicht verweigern. Aber, wie nehmen sie dieses Recht wahr? Darum geht es mir im Folgenden. Denn inzwischen läuft die Diskussion über die Einführung von Evaluationen in Forschung und Lehre schon so lange, dass auch die Rhetorik des Widerstandes einigermaßen durchprobiert ist. Es lässt sich deshalb versuchen, dem Thema Evaluation eine Art Ethnomethodologie der Gegengründe zu widmen. Man kann zusammenstellen, mit welchen Argumenten die Betroffenen in der Wissenschaft versucht haben, sich gegen die Zumutung von Evaluation zu immunisieren.

**>>> Einrede 1:
Das kann doch nicht wahr
sein!**

Wenn einem gegen Evaluationen spezielle Gründe ausgehen, lohnt es sich, fundamentalistisch zu werden mit der Frage, worum es denn eigentlich geht. Eigentlichkeitsfragen bringen die Gegenseite immer in Schwierigkeiten. So auch hier. Selbst wenn man sich, um die Erörterung an dieser Stelle möglichst überschaubar zu halten, ausschließlich auf den Evaluationsbereich ›Forschung‹ beschränkt, schafft man eher Probleme, anstatt sie zu lösen, wenn man sagt, natürlich ginge es um die Bewertung von Qualität. Was aber, bitte schön, ist in der Wissenschaft ›Qualität‹?

Die Wissenschaftslehre sagt uns, dass Forschungsqualität sich nicht objektiv bestimmen lässt und dass es deshalb auch dazu keine Aussagen gibt, die in einem strengen Sinne ›wahr sein könnten. Keinem Zwischen- oder Endprodukt von Forschung kann man unmittelbar

ansehen, was es für den Fortschritt der Wissenschaft oder aber für das Gedeihen von Kultur, Politik und Wirtschaft letztlich bedeutet. Es gibt bei jedem Vorstoß ins Unbekannte nicht nur unvermeidbare Sackgassen, sondern auch ausgesprochen nützliche Umwege. Sogar Irrtümer können höchst instruktiv sein. Und die Wirkung von allem wird durch vielerlei mitbestimmt, was mit einem konkreten Stück Forschung selber wenig zu tun hat, zum Beispiel von der Forschung anderer Forscher und von den Anschlusshandlungen sonstiger Leute. Auch die Verleihung von Nobelpreisen ist deshalb nur mit einem erheblichen Time-lag risikierbar. Zumindest zu den Zeitpunkten, an denen Bestehendes oder gerade Abgeschlossenes evaluiert wird, sind keine methodisch eindeutigen und universell gültigen Regeln verfügbar, um die Qualität konkreter Forschung einwandfrei bestimmen zu können.

**>>> Einrede 2:
Qualität lässt sich nicht
messen!**

Auch wenn man wüsste, was Qualität wäre, ließe sich kaum erwarten, dass sich diese für Vergleichszwecke nummerieren ließe. Das Eigentliche und Besondere, vernimmt man von den Kollegen, steckt in Eigenschaften, die sich der Gleichmacherei von Zahlenspielen prinzipiell entziehen. Und wer solche Zahlenspiele dennoch betreibt, muss den Vorwurf gewärtigen, er handele mit Fragezeichen. Was soll sich denn daraus entnehmen lassen, dass der eine sechs, der andere aber zwölf Artikel publiziert hat. Dass die Forschung des anderen zweifach besser sei als des einen? Und wäre sie zehnfach besser, wenn er 60 Artikel veröffentlicht hätte?

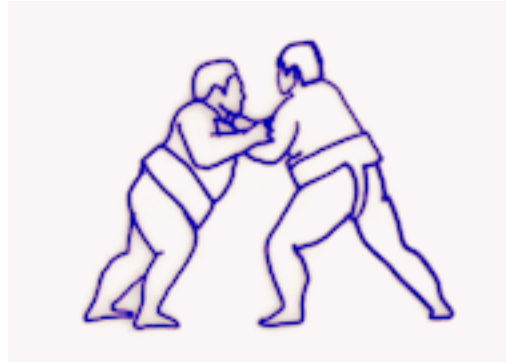
Man würde dem Gegner von Evaluationsstatistiken nur noch weiteres Material anliefern, wenn man sagen würde, natürlich könne Forschungsqualität nicht mit einem einzigen Indikator gemessen werden, man brauche mehrere. Üblich sei zum Beispiel auch das Einbeziehen von Drittmitteln bestimmter Art. Nur, wie lässt sich das gegeneinander verrechnen? Wenn etwa der eine mehr Drittmittel einwürbe als der an-

dere, könnte er damit seine Defizite beim Publizieren kompensieren? Wenn ja: nach welchen Verrechnungsregeln?

**>>> Einrede 3:
Vergleichen lässt sich sowieso
nichts!**

Man kann den Praktikern unbekümmerten Evaluierens die Arbeit mit dem Hinweis noch schwerer machen, dass es bei Evaluationen, sollen sie praktisch bedeutsam sein, nicht eigentlich um die Bewertung von Forschungsergebnissen geht, sondern um die Einschätzung von Forschern, Forschungsgruppen und Forschungsinstituten, also der Akteure, die Forschung betreiben. Will man diese beurteilen, muss man berücksichtigen, dass sich in der Qualität ihrer Ergebnisse nicht nur ihre eigene Qualität, sondern auch die Gunst der Umstände ausdrückt, unter denen sie arbeiten. Welche Umstände sind für Forschungsqualität aber wichtig? Und wie lässt sich, wüsste man dies, benoten, in welchem Maße deren Gunst eine Rolle spielt?

Zu den unfalsifizierbarsten und deshalb erfolgreichsten Einreden derer, die sich vor einem evaluativ erzeugten Tadel immunisieren wollen, gehört die Bekundung, dass sie mit außergewöhnlichen Schwierigkeiten zu tun hätten. Derlei Aussagen verschieben die Kausalattribution von Akteuren auf Kontexte, dienen also der Vermeidung von Schuldprüchen. Sie eignen sich dann, wenn man zum Rückzug gezwungen wird und wenigstens dafür sorgen will, dass der Abgang ehrenhaft bleibt. Falls das wirklich nicht gut ist, was wir machen – besser könnte es in unserer Lage gar nicht sein. Es liegt an irgendwelchen Besonderheiten des eigenen Milieus, die einen selber inkommensurabel machen: die Last der Lehre, der Mangel an Geld, die Unzuverlässigkeit der Kollegen – von privater Unbill ganz zu schweigen. Die Häufigkeit solcher Aussagen lässt zwar darauf schließen, dass sie der Klasse der Ausreden zugehören; aber in jedem Einzelfall lässt sich natürlich nicht ausschließen, dass an ihnen etwas dran ist.





>>> Rückfrage: Anything goes?

Verschwindet damit nun überhaupt die Möglichkeit, gute Forschung von schlechter Forschung zu unterscheiden? Gibt es also auch kein Recht dazu, bei Forschern und Forschergruppen tüchtige von untüchtigen zu unterscheiden?

Vor einem ›Anything goes‹ schützt dann doch wohl die uns allen wohl vertraute Routine, Urteile ständig und durchaus erfolgreich auf etwas zu beziehen, wovon wir nicht genau wissen, was es eigentlich ist. Wäre es anders, würde sich letztlich auch der Umgang mit uns selber verbieten. An Fiktionen sind wir also gewöhnt. Deshalb kann uns auch nicht überraschen, dass die Wissenschaftstheorie den Verzicht auf absolute Maßstäbe für Forschungsqualität mit der Annahme aufgefangen hat, es seien hervorragenden Kollegen zumindest »vernünftige Wetten« (Stephen Toulmin) darüber zuzutrauen, ob eine Forschung etwas taugt oder nicht. Wenn

waren, setzten wir den Computer ein. Wir fütterten ihn mit den Namens- und Themenangaben aus der Projektdatei des Bonner Informationszentrums für Sozialwissenschaften sowie mit den Veranstaltungsnennungen von Soziologen in den Vorlesungsverzeichnissen. Suchten wir nun Rezensenten für eine Neuerscheinung, dann stellte sich regelmäßig ein Resultat ein, das die Qualität unseres Suchprogramms wunderbar bestätigte, unseren Zwecken aber nicht unbedingt dienlich war: Wir stießen am Ende eines weit ausgreifenden Suchprozesses immer auf den Autor selber.

Nun kann man sicher sagen, dass ein Autor über viele Einzelheiten seiner Forschung besser Bescheid weiß als irgendein Außenstehender. Und jeder Forscher mag mit Recht reklamieren, dass er die Umstände, die seine Projekte fördern oder aber behindern, genauer kennt als ein anderer. Aber in der Rolle als Gutachter seiner selbst ist schwerlich sicherzustellen, dass sein Urteil geeignet ist, nicht nur ihn selber zu überzeugen. Es ist ja leider so, dass die Befangenheit, die bei der Betrachtung eigener Angelegenheiten entsteht, nicht selten mit der Eintrübung des gesunden Menschenverstandes einhergeht. Also muss man nach Evaluatoren suchen, die dem Betroffenen möglichst fremd sind. Das bringt mit sich, dass man für Peer review auch Kollegen aus der unmittelbaren Nachbarschaft seiner Themen meiden sollte. Von diesen muss man fürchten, dass sie dem Objekt ihrer Bewertung entweder als allzu gute Kollegen oder aber als das gerade Gegenteil davon herzlich verbunden sind.

Also muss man Experten suchen, die in Distanz zu den Evaluationsobjekten und ihrer Forschung stehen. Das aber erhöht die Wahrscheinlichkeit des Vorwurfs: Die haben von unserer Sache überhaupt keine Ahnung! In der Tat ergeben sich Probleme aus dem Umstand, dass es eine offenkundig inverse Beziehung zwischen Expertise und Unbefangenheit gibt. Je näher dran, umso mehr Einsicht und umso wahrscheinlicher Voreingenommenheit – »that's the problem«. Daraus ergibt sich der Bedarf an

mittleren Distanzmaßen bei der Selektion von Gutachtern. Sie müssen der Sache nahe genug und den Kollegen, um die es geht, fern genug sein, um verlässlich zu urteilen. Es ist klar, dass sich das meistens nicht arrangieren lässt.

>>> Zum Stand der Dinge:

Aus dem bisher Gesagten ergibt sich eine robuste Bestätigung aller Vorurteile: Die Qualität von Versuchen, Qualität zu bewerten, kann nur mangelhaft sein; also sollten Evaluationen überhaupt nicht stattfinden. Nun stellen wir aber fest, dass sie dennoch stattfinden – und dies immer häufiger und künftig noch mehr. Die Gegengründe waren offenkundig nicht imposant genug und die Widerstände nicht hinreichend mächtig, um verhindern zu können, dass Evaluationsprogramme in den 90er Jahren überall in das Repertoire von Wissenschaftspolitik und Wissenschaftsmanagement aufgenommen wurden. Dass sich dies durchsetzen ließ, wird nicht nur daran liegen, dass die Geldgeber, die die Wissenschaft verwalten, den Forschern die Sache mit einer Art Deal schmackhaft machen: Wir sagen euch künftig nicht mehr so genau wie bisher, was ihr machen sollt, wenn ihr nichts dagegen habt, dass wir regelmäßig nachschauen, was ihr tatsächlich macht. Wenn nicht wirksamer, so doch eindrucksvoller als solche Offerter sind sehr allgemeine, gewissermaßen lebenspraktische Hinweise darauf, dass nicht alles, was unvollkommen und heikel ist, nicht stattfinden dürfe. Das ganze Leben ist voll von Beispielen.

Man muss auch einstecken, dass sich solche Hinweise mit ein paar Anspielungen aufdringlich machen lassen. Man kann nämlich sicher sein, dass sich jeder, der Evaluationen als unmöglich bezeichnet, selber dabei erwischen lässt, über die Dignität von Kollegen und die Qualität

Die Evaluationsprogramme wurden den Forschern durch eine Art Deal schmackhaft gemacht: Wir sagen euch künftig nicht mehr so genau wie bisher, was ihr machen sollt, wenn ihr nichts dagegen habt, dass wir regelmäßig nachschauen, was ihr tatsächlich macht.

Inzwischen ist auch die Rhetorik des Widerstandes gegen die Evaluierung einigermaßen durchprobiert

objektive Wahrheitsgarantien nicht verfügbar seien, gäbe es doch einen bestmöglichen sozialen Ersatz, nämlich Experteneinschätzung. Die Sache entsprechender Peer review müsste und könnte dann auch sein, die Validität von Quantifizierungen zu qualifizieren und die Rolle von Kontexten für die Zwecke der Vergleichbarkeit einzuschätzen.

>>> Einrede 4: Wenn überhaupt wer, wären wir es selber, die es können könnten.

In den späten Siebzigern war ich an der Gründung der Soziologischen Revue beteiligt, der Rezensionszeitschrift für die deutschsprachige Soziologie. Um für die zwei- bis dreihundert Neuveröffentlichungen pro Jahr nicht immer dieselben ›Peers‹ als Rezensenten gewinnen zu müssen, die uns drei Herausgebern vertraut

ihrer Produkte Urteile zu fällen, für die er sich nicht ständig entschuldigt. Fast jeder von uns ist an Berufungen von Kollegen beteiligt, schreibt Gutachten über Drittmittelanträge, empfiehlt junge Leute für schöne Karrieren, entscheidet sich gegen die Veröffentlichung bestimmter Aufsätze und redet laufend und nicht immer leise über diesen und jenen. Was uns jetzt als Evaluation begegnet – könnte man uns sagen –, ist nicht viel mehr als eine Formalisierung dessen, was sowieso geschieht, samt Einbau von Kontrollen, die bisher fehlten.

>>> Vorschlag, sicher zur Güte: Wasch mir den Pelz, doch mach mich nicht nass!

Unter diesen Umständen kann man nur noch versuchen, der Sache gewissermaßen die Zähne zu ziehen. Wenn sich Evaluationen schon nicht verhindern lassen, sollte man sie wenigstens möglichst harmlos halten. Man erreicht dies, indem man ihre Ergebnisse für Entscheidungen über Mittelverteilungen nicht freigibt und ihre Funktionen darauf beschränkt, »kurativ« zu sein. Evaluation hieße also so viel wie: zur Kur schicken und wäre sicher eine schöne Sache. Dies auch für die Gutachter, die sich nur noch zu kollegialer Nachbarschaftshilfe einberufen wüssten. Man muss sich freilich vergegenwärtigen, dass unter diesen Bedingungen nicht nur der Tadel nichts kostet, sondern auch das Lob sich nicht auszahlen würde. Letzteres wird nicht allen Recht und Ersteres könnte auf Dauer nicht nur zu nett sein, sondern auch zu teuer kommen. Also muss man vielleicht hinnehmen, dass mit Evaluationen doch alles so läuft, wie es läuft.

