

Marc Torka

## Der Evaluator

Positionierungen, Strategien, Deutungs-, Handlungs- und Urteilsweisen von Wissenschaftlern in Prozessen der Wissenschaftsevaluation

### 1. Einführung: Neue Bewertungsformen, neue Begutachtungsweisen, neue Gutachter?

Seit Anfang der 1980er Jahre kann man die Durchsetzung einer neuen Form der Leistungsbewertung in der Wissenschaft beobachten. Neben die Begutachtung von einzelnen Publikationen und Forschungsanträgen im Rahmen des Peer Review treten so genannte institutionelle Evaluationen von wissenschaftlichen Einrichtungen und Fachbereichen. Damit verschiebt sich die Beurteilung von Forschungsleistungen in vier strukturellen Hinsichten (Torka 2011):

1. dem *Gegenstand der Beurteilung*: Im Zentrum steht nunmehr die Gesamtleistung von ganzen wissenschaftlichen Organisationen und Wissenschaftsgebieten nicht mehr die Forschungsleistungen und zukünftigen Forschungsideen von einzelnen Wissenschaftlern.
2. der *Initiierung des Begutachtungsprozesses*: Während zuvor die Initiative, sich einer Begutachtung zu unterziehen, bei den Wissenschaftlern lag, sind institutionelle Evaluationen (etwa alle 5 bis 7 Jahre) turnusmäßig und verbindlich vorgesehen. Die Initiative geht von wissenschaftspolitischen Instanzen oder wissenschaftsexternen Beobachtern aus.
3. der *Funktion der Begutachtung*: Institutionelle Evaluationen haben nicht nur die wissenschaftsinterne Funktion, Qualitäten und Prioritäten vergangener und zukünftiger Forschungen zu überprüfen. Hinzu tritt die Funktion, Allokationsentscheidungen der Wissenschaftspolitik auf eine gesicherte Informationsbasis zu stellen und zu legitimieren („evidenzbasierte Politik“). Evaluationen sind multifunktional und multiperspektivisch.<sup>1</sup>
4. dem *sozialen Kontext der Urteilsfindung*: Während die Leistungsbeurteilung im Peer Review ausschließlich wissenschaftsintern und im besten Fall anonym („double blind“) erfolgt, richten sich institutionelle Evaluationen an eine erweiterte (politische) Öffentlichkeit und nehmen mancherorts (z.B. Großbri-

---

<sup>1</sup> Sie haben in unterschiedlicher Gewichtung eine auf die Politik gerichtete Verteilungsfunktion, eine öffentlichkeitsbezogene Legitimationsfunktion, eine damit verbundene Leistungskontrollfunktion und schließlich haben Evaluationen den Anspruch, Informationen über die wissenschaftliche Qualität in inhaltlicher und organisatorischer Hinsicht zu erheben, damit Verbesserungen angestoßen werden können (Hornbostel 2010).

tannien) sogar den Charakter eines medialen Events an. Die Evaluatoren in den Gutachterteams interagieren miteinander und sie treffen ihr Urteil in Kenntnis der evaluierten Institution oder sogar aus der direkten Interaktion mit deren Mitgliedern. Zudem werden die Evaluationsergebnisse veröffentlicht.

Diese prinzipiellen Verschiebungen in der Begutachtungsform bilden den Anlass für vielfältige Thesen der Wissenschafts- und Professionssoziologie, deren gemeinsamer Kern darin besteht, Evaluationen als Ausdruck und Motor einer tiefgreifenden Transformation der Wissenschaft im Allgemeinen und der Gutachterrolle im Besonderen zu sehen. Die Autonomie der Wissenschaft sei gefährdet, weil sie in einer aufkommenden „Audit Society“ (Power 1997) mithilfe von Evaluationen einer „second order observation via incongruente perspectives“ (Stichweh 1997: 99f.) unterzogen werde. An die Stelle professioneller Selbststeuerung trete deshalb zunehmend die Fremdbeobachtung und -kontrolle wissenschaftlicher Leistung nach neuen Kriterien (Matthies/Simon 2008, Kurtz 2005: 171f., Oevermann 2005: 33). Ob wissenschaftliche Erzeugnisse einen Beitrag für die Wissenschaft und die Gesellschaft liefern, ergebe sich nicht mehr bloß aus dezentralen und eigentätigen Systemprozessen, in denen Forschungsergebnisse aufgegriffen werden oder nicht (Gläser/Lange 2007), sondern darüber werde in zentralisierten Verfahren entschieden. In ihnen seien nur wenige Wissenschaftler mit Entscheidungsautorität ausgestattet, so dass „epistemic elites“ und hierarchische Entscheidungsstrukturen gestärkt, aber das Kollegialitätsprinzip geschwächt werde (Martin/Whitley 2010). Schließlich werden Evaluationen mit einer Formalisierung (Neidhardt 2000), Quantifizierung (Kieser 2010), Standardisierung und Routinisierung der wissenschaftlichen Leistungsbewertung in Zusammenhang gebracht, woraus sich eine „Abkehr von der fallspezifischen Begutachtung“ (Oevermann 2005: 47) ergeben könnte.

In institutionellen Evaluationen drücke sich also insgesamt ein Vertrauensverlust in die Selbststeuerungsfähigkeiten der Wissenschaft (Weingart 2005), eine Ausdehnung und Vermischung wissenschaftlicher mit außerwissenschaftlichen Leistungskriterien (Frederiksen et al. 2003), ein Niedergang kollegialer und eine Stärkung hierarchischer Beziehungsstrukturen sowie ein Übergang von fallspezifischen zu standardisierten Beurteilungsweisen aus. In allen Dimensionen, die es bislang erlaubten, Wissenschaft als Profession (autonome Selbstkontrolle, Kollegialität, Fallorientierung) zu kennzeichnen, drohe deshalb eine „Entprofessionalisierung“ (Schimank 2005). Die Ausleuchtung der These, dass Evaluierung zu einer Deprofessionalisierung der akademischen Profession führe, beschränkt sich hier auf einen Teilbereich und bedarf schon deshalb weitere Studien. Im Zentrum der Betrachtung in diesem Artikel stehen die Gutachter, die unter den besonderen Bedingungen von Evaluationen Entscheidungen treffen müssen und deshalb von herausragender Bedeutung sind.

Die Gutachterrolle ist selbst im Hinblick auf das wissenschaftseigene Peer Review, also unabhängig von den Sonderbedingungen institutioneller Evaluationen, umstritten. Zahlreiche Studien bemängeln etwa die geringe Reliabilität und Validität von Gutachterentscheidungen, weisen einen strukturellen Konservatismus nach, kritisieren die Abhängigkeit des Urteils von „Old Boy Networks“, „epistemic particularisms“ (Travis/Collins 1991), von „epistemological styles“ (Lamont 2009: 54) und subjektiven

Vorlieben und Interessen der Gutachter/innen.<sup>2</sup> In institutionellen Evaluationen, wo die Gutachter im Auftrag von wissenschaftspolitischen Instanzen verbindliche Urteile fällen, die für die finanzielle und inhaltliche Zukunft von ganzen Instituten einschneidend sein können, ist die Rolle der Gutachter zusätzlichen Spannungen ausgesetzt. Gutachter sind damit in einem Grenzbereich zwischen Wissenschaftspolitik und Wissenschaft positioniert, wo grundsätzlich fragwürdig wird, ob Gutachter ihre Urteilsweise noch an professionsinternen Normen orientieren (können). So könnte die *autonome Selbststeuerung* der akademischen Profession entlang geteilter wissenschaftsinterner Standards in eine Fremdsteuerung umkippen, wenn in Evaluationsverfahren wissenschaftspolitische und organisationsbezogene Kriterien die Oberhand gewinnen. Auch das *Kollegialitätsprinzip* könnte ausgehebelt werden, da einzelne Gutachter zur Intervention in die Arbeitsmöglichkeiten von Kollegen ermächtigt werden und unklar ist, ob die Loyalität der Gutachter den Wissenschaftsgemeinschaften oder wissenschaftspolitischen Instanzen gehört (Schimank 2004). Schließlich könnte sich die Art des Begutachtens ändern. Während die Begutachtung von Publikationen oder Forschungsanträgen sich am Einzelfall orientierte, müssen in institutionellen Evaluationen die heterogenen Gesamtleistungen von wissenschaftlichen Instituten beurteilt werden. Die Informationsmenge, der Zeitdruck unter dem Gutachter stehen und der Anspruch von Evaluationen, Vergleichbarkeit und Transparenz herzustellen, könnten zu einer standardisierten Urteilsweise führen.

Bei diesen Transformationsthesen handelt es sich allerdings um Prognosen, die noch wenig empirisch gesättigt sind und insofern eher offene Fragen darstellen. Die Forschung über Wissenschaftsevaluationen zeichnet sich nämlich dadurch aus, dass sie (a) mehr auf Vermutungen über die Folgen von Evaluationen basiert als auf Nachweisen, (b) die Forscher zugleich Betroffene und in Evaluationen Involvierte sind, so dass verständlicher Weise bevorzugt auf mögliche Gefahren aufmerksam gemacht wird, und (c) die teils weitreichenden Schlussfolgerungen oft nur auf Basis der allgemeinen Gestalt von Evaluationen (fremdinitiiert, verfahrensgesteuert, heterogene Kriterien, potentiell mit finanziellen Folgen verbunden) gezogen werden ohne jedoch die Prozesse innerhalb von Evaluationen empirisch zu analysieren. Man stößt auch auf gegenläufige Thesen, die Evaluationen als bloßes „akademisches Ritual“ (Michaels 2010) der Rechtfertigung verstehen, als „Formalisierung dessen, was [in der Wissenschaft] sowieso geschieht“ (Neidhardt 2000: 29), oder die sogar zu dem Schluss kommen, dass „fast jeder Aspekt wissenschaftlicher Kommunikation (...) durch Evaluationen bestimmt (ist)“ (Hirschauer 2002: 1). Evaluation heißt schließlich nicht mehr als Bewertung und stellt in sofern einen alltäglichen Vorgang in der Wissenschaft dar, der nun in neuen Praxisformen auf neuem Aggregationsniveau vollzogen wird.

Wie Wissenschaftler innerhalb von Evaluationen auf diese neuen Arrangements handelnd reagieren und mit Interpretationen versehen, und ob sie dabei wissenschaftsferne Maßstäbe übernehmen oder nicht, ist deshalb ein wichtiges und bislang wenig erforschtes Thema. Insbesondere fehlt es an Studien, die mikroanalytisch die ‚Innenwelt der Evaluation‘ und die von Gutachtern in Evaluationsverfahren praktisch in An-

---

<sup>2</sup> Für einen Überblick siehe Stefan Hirschauer (2002, 2004).

spruch genommenen Regeln des Beurteilens ergründen.<sup>3</sup> Erst auf dieser Basis können empirisch gesättigte Aussagen getroffen werden, inwiefern und inwieweit sich die Gutachterrolle in Evaluationen reproduziert oder transformiert.

Um den Realitätsgehalt dieser sehr weitreichenden Thesen über die Folgen von Evaluationen für die soziale Rolle der Gutachter in der Praxis des Evaluierens einschätzen zu können, werden im Folgenden umfangreiche Materialien und einige Ergebnisse aus einem Forschungsprojekt zu verschiedenen Evaluationsverfahren dargelegt.<sup>4</sup> In Großbritannien wurde das *Research Assessment Exercise* (RAE), eine zeitgleiche Evaluation aller Universitätsinstitute und Forschungsbereiche, in den Niederlanden die Evaluation von einzelnen Akademieinstituten entlang des sogenannten *Standard Evaluation Protocol* (SEP) und in Deutschland das auf heterogene außeruniversitäre Forschungsinstitute (ehemals „Blaue Liste Institute“) angewendete *Evaluationsverfahren der Leibnizgemeinschaft* (EL) untersucht. Damit umfasst die Datenbasis zum einen unterschiedliche Arten von wissenschaftlichen Einrichtungen und Wissenschaftsbereichen und zum anderen verschiedene Evaluationstypen. Letztere kann man mit Richard Whitley (2007: 9) als „two ideal types of strong and weak evaluation systems“ charakterisieren: das RAE als ein starkes System, weil es den direkten Vergleich aller nationalen Universitätsinstitute innerhalb eines Fachgebiets vorsieht und auf diese Weise ein numerisches Ranking produziert, das unmittelbare Konsequenzen für die zukünftige Finanzierung der Institute hat; das SEP und das EL als schwache, weil wissenschaftliche Einrichtungen nicht in einem vergleichenden Wettbewerb, sondern einzeln evaluiert werden, und am Ende fallbezogene „Empfehlungen“ stehen, die nur bei besonders gravierenden Fehlleistungen direkte Folgen für die Förderung haben.

Aus den genannten Verfahren wurde der Evaluationsprozess von 11 wissenschaftlichen Einrichtungen (Geistes-, Sozial-, Lebens-, Natur- und Technikwissenschaft) multiperspektivisch erfasst: Die Materialbasis besteht aus den offiziellen Verfahrensdokumenten und insgesamt 102 leitfadenstrukturierten Interviews mit Planern der Verfahren in den jeweiligen Evaluationsagenturen (11), leitenden Wissenschaftlern aus den evaluierten Institutionen (44) sowie den Evaluatoren selbst (47). Wir haben möglichst detaillierte Beschreibungen des Evaluationsprozesses erhoben, in dem nach Teilnahmemotiven von Gutachtern, Vorbereitungsarten, auffälligen Ereignissen und Diskussionen oder nach Beispielen für Votenvergaben gefragt wurde. Aus diesen Erzählungen wurden dann grundlegende Erwartungen rekonstruiert, die sich an die Gutachterrolle richten.<sup>5</sup> Diese sollen im Folgenden aus drei verschiedenen Perspektiven dargelegt

<sup>3</sup> Stefan Hirschauer (2005, 2010) hat die Begutachtungspraktiken im Kontext des Peer Review für wissenschaftliche Zeitschriften und Michèle Lamont (2009) im Kontext der Stipendienvergabe untersucht. An diese ‚Social Studies of Peer Review and Evaluation‘ schließt der vorliegende Beitrag an.

<sup>4</sup> Es handelt sich um das Forschungsprojekt ‚Formation of Judgement. International Case Studies on Evaluation of Scientific Institutions‘. Die Daten wurden zwischen 2007 und 2008 in der Forschungsgruppe Wissenschaftspolitik am Wissenschaftszentrum Berlin für Sozialforschung (WZB) zusammen mit Christoph Biester, Silke Gülker, Hildegard Matthias und Dagmar Simon erhoben.

<sup>5</sup> Einschränkung ist anzumerken, dass es nicht möglich war, die Gutachterrolle ‚at work‘ zu analysieren, da uns eine Teilnahme am Evaluationsprozess und der direkte Zugang zu den Gesprächen im Gutachtergremium verwehrt blieb. Das ist ein grundlegendes methodisches Problem bei der Erforschung der Innenwelt von Evaluationen, denn dabei handelt es sich um besonders herausgehobene Ereignisse, für die explizite Vertraulichkeitsregeln vorgesehen sind und die deshalb hinter verschlossenen Türen ablaufen. Wirkliche Praxisprotokolle sind nicht zu bekommen; allenfalls sind teilnehmende Beobachtungen möglich (z.B. Tra-

werden: der offiziell erwünschten Positionierung von Gutachtern in den formalen Verfahrensbeschreibungen (2), aus der Perspektive der Begutachteten darüber, wie die Gutachter vermutlich handeln werden (3) und aus der Perspektive der Gutachter hinsichtlich ihrer eigenen Praxis (4 und 5). Abschließend werde ich diskutieren, in welchem Verhältnis die empirischen Befunde zur Deprofessionalisierungsthese stehen und welche Anschlussfragen sich hieraus ergeben (6).

## 2. Die Positionierung von Gutachtern im Evaluationsverfahren

Will man den Folgen nachgehen, die Evaluationen für die Gutachterrolle haben können, dann ist das natürlich nur in solchen Evaluationsverfahren möglich, die bei der Entscheidungsfindung überhaupt auf das Urteil von Gutachtern zurückgreifen.<sup>6</sup> Das ist in den meisten und auch in den hier untersuchten Evaluationsverfahren noch immer der Fall. Inwiefern diese – zweifelsohne unter externem Druck der Wissenschaftspolitik – implementierten Verfahren überhaupt wissenschaftsferne Aspekte in die Begutachtung einspeisen und das Gutachterhandeln programmieren, kann man den jeweiligen Verfahrensordnungen entnehmen. Zunächst lässt sich feststellen, dass wissenschaftliche Gutachter und wissenschaftliche Vereinigungen eine zentrale Position im Evaluationsprozess und bei der konkreten Ausgestaltung von Evaluationsverfahren einnehmen. Den Verfahren sind umfangreiche Konsultationen mit der scientific community vorausgegangen, sie werden von wissenschaftsnahen und auch mit Wissenschaftlern besetzten Organisationen erarbeitet und durchgeführt.<sup>7</sup> Die Nominierung von Gutachtern erfolgt über Fachgemeinschaften oder die evaluierten Einrichtungen können selbst Vorschläge unterbreiten; Vertreter wissenschaftspolitischer Instanzen sind bei der Urteilsfindung der Gutachter allenfalls als Gäste beteiligt. Insofern wäre eine Charakterisierung von Wissenschaftsevaluationen als rein wissenschaftsexterne Kontrollinstanzen ebenso eine Überzeichnung, wie eine einfache Zuordnung zur Wissenschaft. Denn alle Evaluationsverfahren agieren aus einer spezifischen Organisationslogik heraus und versuchen eine Standardisierung der Begutachtung vorzunehmen. Die schriftlich fixierten Verfahrensvorschriften richten sich an alle Beteiligten und sie liegen diesen als objektiviert und verbindliche „Guidance“<sup>8</sup> (RAE), „Checklist“ (SEP) oder „klar definierte Evaluationskriterien“ (EL) vor, die das gutachterliche Handeln anleiten sollen. Hierzu werden der Gegenstand, der formale und zeitliche Ablauf, die Entscheidungsdimensionen und -kriterien sowie die von den Begutachteten beizubringenden Informationen definiert.

---

vis/Collins 1991, Langfeldt 2001, Lamont 2009). Insofern konnten lediglich die Deutungen einer Praxis, aber nicht die Praxis selbst analytisch erfasst werden. Deutungen und Zuschreibungen sind aber ein wesentlicher Bestandteil von sozialen Rollen, denn soziologisch betrachtet werden Rollen als überindividuell gültige Verhaltenserwartungen verstanden (Luhmann 2000: 82).

<sup>6</sup> Für den Sonderfall „Evaluation without Evaluators“, die ausschließlich auf formalen Indikatoren beruhen, siehe das australische Beispiel von Jochen Gläser und Grit Laudel (2007).

<sup>7</sup> In den hier untersuchten Fällen werden die Verfahren in Großbritannien von dem englischen (HEFCE), walisischen (HEFCW), schottischen (SFC) und nordirischen (DEL) Higher Education Funding Council, in den Niederlanden gemeinsam von der nationalen Akademie der Wissenschaft (KNAW), der Vereinigung der niederländischen Universitäten (VSNU) sowie der nationalen Forschungsförderorganisation (NWO) und in Deutschland von der Leibnizgemeinschaft organisiert.

<sup>8</sup> Bei doppelten Anführungszeichen handelt es sich stets um Zitate, die in Kapitel 2 aus den jeweiligen Verfahrensbeschreibungen, in Kapitel 3 von den Evaluierten und in den Kapiteln 4 und 5 von den Evaluierenden stammen.

Nimmt ein Gutachter an Evaluationen teil, so wird dieser sowohl als Angehöriger der scientific community adressiert, als auch als temporäres Mitglied eines organisierten Verfahrens, von dem qua Mitgliedschaftsrolle erwartet werden kann, den jeweiligen formalen Verfahrensvorgaben und Bewertungskriterien zu folgen. Diese sollen nun betrachtet werden, um genauere Auskunft zu erhalten, in welchem Verhältnis beide Rollen zueinander stehen.

Die erbrachte Leistung und zukünftig vermutete Leistungsfähigkeit von Instituten wird in den Verfahren anhand der wissenschaftlichen Erzeugnisse (Publikationen, Patente, Transfer- und Beratungsleistungen), der Reputation der Institution (Preise, Gremien, Einladung zu wichtigen Tagungen, bekannte Gastwissenschaftler) und der Fähigkeit der Institutsleitung, angemessene Führungsstrukturen, schlüssige Strategien, Profile und Infrastrukturen (Nachwuchsförderung, (Dritt-)Mittelakquise, Kooperationen) auszubilden und bereitzustellen, beurteilt. Dabei soll vor allem deren Qualität, aber (im SEP und EL) auch ihre „Produktivität“, „Relevanz“, „Effizienz“ oder „Machbarkeit“ bewertet werden. Die Bewertungsdimensionen sind also sehr breit gefasst. Ebenso beschreiben die Bewertungskriterien wissenschaftliche Leistung recht unspezifisch, da in einem einheitlichen Verfahren über wissenschaftliche Organisationen und Wissenschaftsbereiche mit ganz verschiedenen Ausrichtungen geurteilt werden soll. Das Verfahren verlangt den Gutachtern also in der Tat ab, in ihrem Urteil Betrachtungsweisen zu übernehmen und Aspekte zu berücksichtigen (z.B. effiziente Führungsstrukturen), die nicht unbedingt der Wissenschaft und ihrer Expertise entspringen. Auf der anderen Seite werden wissenschaftsspezifische Leistungsgesichtspunkte und -nachweise aber nicht einfach aufgelöst, sondern auf den neuen Gegenstand Organisation übertragen. Zum Beispiel werden Leistungen, die Personen in ihren Wissenschaftsgemeinschaften zuerkannt wurden (Publikationen, Preise, Einladungen etc.), nun der Organisation zugerechnet.<sup>9</sup> Insofern können Gutachter expansiven Gebrauch von ihren im Rahmen der scientific community erworbenen Beurteilungsmaßstäben und -kompetenzen machen. Hinzu kommt, dass die Verfahren weder die Kriterien, die Gewichtung zwischen Kriterien noch die Verwendungsweise von Kriterien exakt definieren, so dass die Gutachter über einen erheblichen Interpretationsspielraum verfügen. So erfährt man im SEP und im EL nichts über das wichtige Kriterium Qualität und auch im RAE wird die Interpretationsbedürftigkeit nicht getilgt wenn es heißt: „Quality that is world-leading in terms of originality, significance and rigour“.

In den Verfahrensordnungen stößt man immer wieder auf den Anspruch, das Gutachterhandeln über allgemeine und interpretationsbedürftige Kriterien programmieren zu wollen, ohne ihn wirklich einlösen zu können. Das RAE versucht das Problem, verallgemeinerbare und zugleich fachadäquate Kriterien zur Anwendung zu bringen,

---

<sup>9</sup> In der Wissenschafts- und Hochschulforschung gilt dieser Sachverhalt unter dem Stichwort „Organisationswerdung der Universität“ (Meier 2009, Krücken/Meier 2006) als ein historischer Einschnitt, dessen Ausdruck und Motor institutionelle Evaluationen sind. Organisationen und ihren Leitungen wird zugeschrieben, selbst für die dortigen Forschungserfolge verantwortlich zu sein, obwohl sie nicht von ihnen betrieben werden (Meier/Schimank 2010: 218). Das hat nicht nur zur Folge, dass Organisationen beginnen auf Forschungsprozesse Einfluss zu nehmen, sondern eben auch, dass die Güte der Organisation von den Errungenschaften abhängt, die ihre Mitglieder in den Wissenschaftsgemeinschaften, also außerhalb des organisatorischen Einflussbereichs, erzielen. Die *wechselseitigen* Abhängigkeiten scheinen somit größer zu werden.

über eine organisatorische Lösung zu bearbeiten. So bleibt es dort Sache der Gutachter in den fachübergreifenden main panels, „broadly similar approaches“ auszubilden, die dann in „sub-panels“ disziplinspezifisch ausgelegt werden müssen. Konsequenterweise beschreibt sich das Verfahren selbst als „expert review exercise“ und untermauert damit die zentrale Position der Gutachter.

Im SEP äußert sich die Spannung zwischen Programmierungswunsch und Abhängigkeit von den Interpretationsleistungen der Gutachter in widersprüchlichen Definitionen und Verbindlichkeiten der Verfahrensvorgaben. So ist die Evaluation einmal „based on expert assessments“, dann ist das Verfahren aber wieder selbst „the basis for research evaluation“.

Besonders anschaulich kann man beim EL nachverfolgen, auf welche Weise und aus welchen Gründen Autonomie und verfahrensmäßige Programmierung der Gutachter in Spannung geraten. Dort heißt es in den Verfahrensrichtlinien:

*„Die Bewertung der wissenschaftlichen Leistungen der Einrichtungen erfolgt durch Fachgutachter (peer review). Um zu gewährleisten, dass für alle Einrichtungen – unter Berücksichtigung ihres fachlichen sowie aufgabenspezifischen Charakters und unabhängig von der personellen Zusammensetzung der Bewertungsgruppe – im Wesentlichen vergleichbare Maßstäbe angesetzt werden, orientiert sich die fachliche Bewertung der wissenschaftlichen Qualität und Leistungsfähigkeit an einem standardisierten Kriterienkatalog (vgl. Abschnitt 1). Auswahl, Gewichtung und ggf. Ergänzung der Kriterien je nach fachlichem Umfeld und Auftrag der Einrichtung erfolgt durch die Fachgutachter.“*

In dieser Beschreibung werden zwei Modelle der Urteilsfindung kontrastiert und miteinander verschränkt. Das klassische Kollegialmodell, in dem auf Basis fachlicher Expertise Kollegen autonom über Kollegen befinden, bildet zwar die Grundlage („Bewertung ... erfolgt durch Fachgutachter“), gilt aber offenbar als wenig verlässlich, weil es alleine eben nicht „gewährleistet“, alle Einrichtungen mit ‚vergleichbaren Maßstäben‘ zu beurteilen. Dazu muss das Kollegialmodell durch einen von Personen und Typ der wissenschaftlichen Organisation unabhängigen, „standardisierten Kriterienkatalog“ ergänzt werden. Dem Kollegialmodell und der Urteilsweise von Fachgutachtern wird implizit also unterstellt, zu sehr nach fachlichen, persönlichen oder fallbezogenen Sonderperspektiven zu entscheiden („unabhängig von der personellen Zusammensetzung der Bewertungsgruppe“) und demnach über keine geteilten Standards der Urteilsfindung zu verfügen. Gleichwohl verzichtet das Verfahren nicht auf die den Gutachtern zugeschriebene Fähigkeit, Kriterien fallbezogen interpretieren und verwenden zu können. Ohne dem wäre nämlich der Anspruch des Verfahrens, fair zu sein, nicht mehr gewährleistet, weil die Gleichbehandlung aller wissenschaftlichen Einrichtungen mithilfe gleicher Kriterien in eine Gleichmachererei umkippen würde. Das Problem, sowohl vergleichbare als auch fallangemessene Kriterien zu verwenden, bearbeitet dieses Verfahren, indem es selbst Kriterien vorgibt, die dann von den Gutachtern fallspezifisch ausgelegt und verwendet werden sollen. Das Verfahren beschreibt die Gutachterrolle einerseits als Interpretatoren von extern vorgegebenen Regeln. Andererseits haben die Gutachter völlige Entscheidungsautonomie, wenn ihnen die Möglichkeit zur „Ergänzung der Kriterien“ eingeräumt wird.

Allen vorgestellten Verfahrensordnungen kann man somit entnehmen, dass die Autonomie der Gutachter und die professionelle Selbststeuerung im Verfahren nicht einfach ausgehebelt werden. Vielmehr haben Gutachter weiterhin eine zentrale Position

im Verfahren, weil sie stellvertretend ein Grundproblem bearbeiten sollen, das jedem Evaluationsverfahren zu Grunde liegt: unterschiedlichste Wissenschaftsorganisationen und -typen in einem einheitlichen Verfahren zu beurteilen. Ihnen kommt die Rolle zu, allgemeine Regeln und abstrakt formulierte Kriterien auf konkrete Fälle zu beziehen. Sie sind folglich Interpretatoren des Verfahrens, müssen zwischen verschiedenen Kriterien wählen und gewichten, Spezifikationen vornehmen oder wenn es der Fall erfordert, sogar neue Regeln entwickeln. Betrachten wir im Folgenden, welche Rolle oder Rollenkombination den Gutachtern von den Angehörigen der evaluierten wissenschaftlichen Einrichtungen zugeschrieben wird.

### 3. Die Gutachterrolle aus Sicht der evaluierten Institute

Wie jede praktizierte Rolle ist auch die des Gutachters nicht nur von der offiziellen Aufgabenbeschreibung oder ihren Selbstbeschreibungen abhängig, sondern auch von den an sie gerichteten Erwartungen der evaluierten Institute. Um herauszuarbeiten, welche Ausprägung die Gutachterrolle aus deren Sicht annimmt, ist es hilfreich, sich die allgemeine Handlungssituation zu vergegenwärtigen, vor der die Institute stehen und die ihr Vorgehen prägt. Kein Institut lässt sich freiwillig evaluieren, sondern wird hierzu turnusmäßig von Wissenschaftsorganisationen aufgefordert und kann sich dieser externen Erwartung auch nicht entziehen. Damit stellen Evaluationen den inhaltlichen und finanziellen Fortbestand von Instituten ohne akuten Anlass in Frage und fordern mit der Kopplung an Förderentscheidungen eine Rechenschaftslegung ein. Für die Institute sind Evaluationen unter diesen Vorzeichen künstlich erzeugte Krisensituationen, in denen ihre Fähigkeit, relevante Problemlagen zu erkennen und angemessen zu bearbeiten, in Zweifel gezogen wird. Entsprechend wirken die zur Evaluation ausgewählten Institute daraufhin, diese Situation möglichst unbeschadet zu überwinden und diesen Verdacht abzuwenden. Empirisch drückt sich das in der selbstverständlichen Erwartung „natürlich auf ne sehr gute Evaluierung“ und in der Hoffnung aus, von den Gutachtern in ihrer Handlungsweise bestätigt zu werden. Um dieses Ziel zu erreichen, bedienen sich die Institute typischer Vorgehensweisen. Erstens stößt man auf interne Reflektions-, Selbstverständigungs- und Selbstvergewisserungsprozesse, bei denen es darum geht, „the state of affairs in the institute“ herauszufiltern und herauszuarbeiten „what we would like the institute to be“. Das gesamte Institut und das Ineinandergreifen verschiedener Forschungsarbeiten zu reflektieren, gilt bei den Institutsangehörigen als ein positiver Effekt von Evaluationen. Zugleich ist dies eine zentrale Vorbereitungsmaßnahme für weitere typische Handlungsweisen, weil dabei etwaige Lücken und Ansatzpunkte für Kritiken sichtbar werden, auf die Institute strategisch reagieren können. Die umfangreichen und in den Instituten frühzeitig eingeleiteten Vorbereitungsmaßnahmen auf eine Evaluation<sup>10</sup> zielen nämlich darauf ab, bereits im Vorfeld alles zu tun, um den Ruf, die Finanzierung und die inhaltliche Ausrichtung der jeweiligen wissenschaftlichen Einrichtung nicht zu gefährden. Folglich muss man die Vorbereitungshandlungen der Institute auch als Überzeugungs- und Kritikvermeidungspraktiken interpretieren, die je nach Typus des Evaluationsverfahrens variieren. Idealerweise unterziehen die

<sup>10</sup> Die Vorbereitung auf eine Evaluation dauert meist mehrere Jahre. Dafür werden zunehmend eigens Stellen geschaffen und Monitoringsysteme errichtet, so dass eine Veralltäglichsung von Evaluationen gemäß dem Motto ‚nach der Evaluation ist vor der Evaluation‘ zu verzeichnen ist.



Institute alles, was die Gutachter beobachten und kritisieren könnten, einer intensiven institutsinternen Vorabüberprüfung, in die verschiedene Vorannahmen über das Gutachterhandeln einfließen.

Im RAE etwa, das maßgeblich auf die erneute Begutachtung bereits veröffentlichter Publikationen ausgerichtet ist, richtet sich die Aufmerksamkeit vor allem auf die Produktion, Selektion und Steigerung geeigneter Veröffentlichungen und Outputs. Dazu wird das RAE von einzelnen Instituten im Vorfeld intern simuliert, oftmals unter Hinzuziehung externer Gutachter. Auf der Grundlage dieser Selbstevaluationen werden dann mit zusätzlichen Stellen, durch die Straffung des Lehrangebots sowie die interne Differenzierung in stärker forschungs-, lehr- und verwaltungsbezogenes Personal Freiräume und Unterstützungsmöglichkeiten für die evaluierungsrelevanten als „research active“ in Erscheinung getretenen Personen geschaffen. In Absprachen über geplante Publikationen und der Gegenkontrolle des Produktionsprozesses werden zudem Verbindlichkeiten erzeugt, um „geeignete“ Publikationen rechtzeitig vor dem eigentlichen Verfahren fertig zu stellen. Diese Vorbereitungspraxis unterstellt, dass die Gutachterrolle im RAE mit der des Fachgutachters im klassischen Peer Review von Zeitschriften vergleichbar ist, und – wenn die Simulation des Verfahrens eine rationale Vorbereitungsmaßnahme sein soll –, dass die Gutachter keineswegs beliebig urteilen.

Neben der Erzeugung und Selektion von wissenschaftlichen Erzeugnissen gibt es aber auch strategische Entscheidungen, die im Feld explizit als „playing the game“ firmieren, also als Manipulation und Ausnutzung der Eigenlogik des Verfahrens gelten. So lassen Institute zur Verbesserung der durchschnittlichen Leistung<sup>11</sup> nur wenige Mitarbeiter an der Evaluation teilnehmen, wählen mutmaßlich zwischen härter oder weicher urteilenden Gutachterpanels aus oder versuchen, eigene Mitarbeiter in den Panels unterzubringen. Derartige Versuche der Einflussnahme auf die Begutachtungspraxis sowie die Ausnutzung der Verfahrenslogik bis zur Täuschung bringen klar zum Ausdruck, dass für die Institute der Erfolg im RAE-Wettbewerb im Zentrum steht. Noch vielfältiger sind die strategischen Vorbereitungen und Annahmen über das Gutachterhandeln in den evaluierten Instituten des niederländischen und deutschen Verfahrens, wo der zu begutachtende Fall keine anonymen Einzelpublikationen, sondern das Institut in seiner Gesamtheit ist. Bereits die Ausrichtung auf die Gesamtgestalt des Instituts enthält eine Typisierung der Gutachterrolle:

*„a scientist always wants to go into great detail, because this is his work [...] For the evaluation this is not relevant, the evaluation wants to have the broad picture of an institute, [...] so that was also saying implicit criteria's, not to go into this freebie nitty-gritty minute detail.“*

Die Gutachter werden also weniger als Experten mit detaillierten Sachkenntnissen, denn als Generalisten und „Repräsentanten des gesamten wissenschaftlichen Feldes“ (Barlösius 2008: 258) angesprochen, obwohl die Faszination für die Details des Forschungsgegenstands im Hinblick auf generelle Fragestellungen der jeweiligen Wissenschaft (vgl. auch Oevermann 2005, Franzmann 2008) eigentlich zum Kernbestand wissenschaftlichen Handelns gehören. Zur Erfassung der Gesamtgestalt des Instituts erhalten die Gutachter drei eigens für die Evaluation erzeugte Informationsquellen:

<sup>11</sup> Das numerische „quality profile“, in dem die Ergebnisse des RAE kommuniziert werden, bildet nur ab, wieviel Prozent der Arbeiten zwischen „world leading“ (4) und „unclassified“ (0) bewertet wurden, aber nicht die Gesamtzahl der Einreichungen.

Einen „Selbstevaluationsbericht“ mit Beschreibungen der besonderen Ausrichtung, wesentlichen Ergebnissen und Strukturdaten der Einrichtung (Drittmittel, Publikationen, Qualifikationen, Rufe, etc.), ein Veranstaltungsprogramm mit Präsentationen von Forschungseinheiten und die direkte Interaktion mit Mitarbeitern während der Institutsbegehung. Alle drei Informationsquellen sind Gegenstand von institutsinternen Selbstverständigungsprozessen und strategischen Überlegungen, die mit Typisierungen der Gutachterrolle einhergehen. Die Selbstverständigungsprozesse zwischen den Institutsmitarbeitern etwa zielen darauf ab, sich selbst und den Gutachtern das spezielle Profil der Arbeitsweisen im Institut verständlich zu machen. Das Institut in seiner Besonderheit herauszustellen, unterstellt aber, dass die Gutachter eine Einzelfallbetrachtung vornehmen:

*„But it yeah, it really depends on the people, if you have raw people, I don't know not raw people, but people that are, sometimes if you have a judge or an evaluator with very strong ideas of his own, that can be very dangerous because he or she can then tell you what's good for him or her but not good for the institute so they really, I think it's very difficult to be a good committee member. But it's yeah, it's a people's business, it's even more important than all these, output classification or which is very useful, very necessary but basically if you have a good committee, that's crucial.“*

Da nicht unbedingt mit einem Rollenverständnis der Gutachter der Art gerechnet werden kann, weder entlang standardisierter „output classification“, noch auf Grundlage eigener Präferenzen, sondern zum Wohle des je speziellen Instituts zu urteilen, treffen die Institute strategische Vorbereitungsmaßnahmen. Sie versuchen auf die Gutachterwahl Einfluss zu nehmen, indem sie Kollegen vorschlagen, die das Institut kennen, zum Institutsprofil passende Fachkompetenzen haben und dem Institut mutmaßlich wohl gesonnen sind. Sie versuchen aber auch die Informationsquellen Bericht, Präsentation und Interaktion so zu gestalten, dass sie zum vermuteten Gutachterhandeln passen. Der Selbstevaluationsbericht bietet vielfältige Möglichkeiten, Leistungen hervorzukehren und Schwierigkeiten zu verdecken, also zu kontrollieren, was die Gutachter zu sehen bekommen und was nicht. Deshalb geraten diese Dokumente bei Gutachtern schnell in den Verdacht, bloß „Selbstdarstellungen mit dem üblichen Selbstlob“ zu sein. Um diesen Eindruck zu verhindern und zu gewährleisten, dass „das nicht nur auf dem Papier jetzt steht, sondern gelebt wird“, zirkulieren die Berichte zwischen den Institutsmitgliedern. Letztere unterfüttern den Bericht mit Informationen aus den Forschungsbereichen, zeigen die Bedeutung ihrer Forschungen für die Programmatik des Gesamtinstituts auf, machen auf mögliche Kritikpunkte aufmerksam und tragen insgesamt dazu bei, dass Darstellung und Wirklichkeit des Instituts nicht zu sehr auseinander fallen.<sup>12</sup> Durch dieses Vorgehensweise werden Gutachter als Beobachter typisiert, die den Realitätsgehalt einer im Kontext von Evaluationen erwartbaren Leistungsschau kritisch überprüfen und die die Fähigkeit besitzen, die Beschreibungen, Strukturdaten und Informationen aus der direkten Interaktion in eine „perspective of real life“ zu übersetzen.

<sup>12</sup> Die Beteiligung an der Berichtsabfassung erfolgt aus der Perspektive der Institutsmitarbeiter nicht ohne Eigeninteresse. Positive Erwähnung im Bericht und bei der Evaluation zu finden, gilt ihnen als Möglichkeit, ihre institutsinterne Position zu sichern und zu stärken.

Aber auch jenseits wissenschaftlicher oder vom Verfahren vorgegebener Bezüge werden Gutachter in ihrer alltagsweltlichen Verankerung und den dort vermutlich vorherrschenden, sehr allgemeinen Urteilkriterien, adressiert. Denn die Institute versuchen nicht nur über die kommunizierten Inhalte, sondern auch über die ästhetische Gestaltung von Dokumenten und Präsentationen sowie mit einem aufwendigen ‚Eventmanagement‘ Gutachter für sich einzunehmen. Die Institute imaginieren mit diesen Vorbereitungen die Situation der Gutachter als Gäste mit wenig Zeit, die eine zuvorkommende Sonderbehandlung ebenso erwarten dürfen wie einen schnell und einfach zu lesenden, übersichtlich gestalteten und auf das Wesentliche beschränkten Bericht. Solchen Vorbereitungsmaßnahmen kann man die Annahme entnehmen, dass Gutachter als ‚ganze Personen‘ handeln und jeder Widerstand der sich ihnen bietet, Anlass (nicht jedoch die nachfolgende Begründung!) für Kritik liefern könnte. Damit basiert gutachterliches Handeln aus Sicht der in diesen Verfahren evaluierten Institute neben Fach- und Feldwissen eben auch auf allgemeinen Kriterien der Stimmigkeit, Konsistenz oder Eleganz des Beobachtungsgegenstandes und ist insofern in der ästhetischen Erfahrung (Oevermann 2005) fundiert. Die strategisch handelnden Institute rechnen mit möglichen Eigenarten von Gutachtern und richten ihre Vorbereitungsmaßnahmen daran aus. So dienen die aufwendigen Proben und Planungen des Veranstaltungsprogramms nicht nur der Information, sondern auch der Kontrolle von Gutachtern. Denn im Gegensatz zu offenen Interaktionen lassen sich die Präsentationen des Veranstaltungsprogramms vorbereiten, üben und verfeinern. Zudem schränkt ein straffes Programm die Möglichkeit unkalkulierbarer Interaktionen ein. In Generalproben wird sogar versucht, diese interaktiven Aspekte unter Kontrolle zu bringen. Präsentationen werden auf mögliche Reaktionsweisen von Gutachtern untersucht und selbst das Verhalten von Mitarbeitern in der direkten Interaktion mit Gutachtern wird diskutiert. Hierbei werden basale Erwartungen an die bevorzugte Kommunikationsform der Gutachter aufgerufen:

*„(D)ann wurde sagen wir mal wirklich dramaturgisch operiert, also wo ein Poster steht, wer sich mit wem unterhält, ob man sehr viel selber sagt oder ob man die Gutachter auch was sagen lässt. Das war manchmal ein sehr wichtiger Hinweis zu sagen, also Gutachter bewerten dann besonders gut, wenn sie selber viel reden, weil sie das Gefühl haben, das ist interessant, ja, wenn sie sich das alles anhören müssen und zugeknallt werden mit irgendwelchem Kram, das mögen die gar nicht, sondern man hat ein kurzes Eingangs-Statement, hat ein paar Kernpunkte und vielleicht ne Frage an die und gibt es an die zurück und wenn die dann reden, dann fühlen die sich gut“.*

Die Strategie, mithilfe der Programmplanung offene Interaktionen zu begrenzen, stößt aus dieser Sicht an ihre Grenzen, weil Gutachter damit eine passive Zuschauerrolle erhalten würden, die in Spannung zur antizipierten Gutachterrolle gerät. Sie zeichnet sich durch die freie, aktive Suche nach für sie wichtigen Informationen und die Präferenz für eine offene Kommunikationsform aus: „es sollte ja dazu kommen, dass ein Gespräch geführt wird“.

Insgesamt lässt sich den Vorbereitungsmaßnahmen der evaluierten Institute entnehmen, dass sie mit allen Mitteln versuchen, gegenüber den Evaluatoren ihre Autonomie zu behaupten und von ihnen Bestätigung zu erlangen. Dafür bilden sie Strategien aus, die auf unterschiedlichen Konzepten der Gutachterrolle beruhen. Gutachter werden zugleich als ganze Personen mit alltagsweltlichen Beobachtungsweisen (z.B. als Gast), als Mitglied der gleichen Fachgemeinschaft mit prinzipiell vergleichbaren

Urteilsweisen, als kritische Interpreten heterogener Informationen oder potentiell wohl gesonnene Kollegen sowie als an eigenen Sonderperspektiven oder am evaluierten Einzelfall orientierte Beobachter angesprochen. Die heterogenen Erwartungen an die Gutachterrolle und die vielfältigen Strategien, Gutachter für sich einzunehmen, zeigen, dass weder die Verfahrensordnungen noch der Umstand, dass Kollegen die Evaluation vornehmen, Erwartungssicherheit bei den Instituten erzeugt. In dieser ‚Notlage‘ greifen Institute auf die Einschätzungen von Mitarbeitern und Kollegen zurück, regen Verständigungsprozesse an, um implizite Kritik-, Sicht- und Urteilsweisen sichtbar zu machen, und unterstellen dabei, dass die Evaluatoren ähnlich Handeln werden.

#### 4. Selbstbeschreibungen der Gutachterrolle

Zerfällt auch aus Sicht der Gutachter ihre Rolle in unterschiedliche Konzepte oder stößt man doch auf gemeinsame Grundorientierungen? Schauen wir uns dazu zunächst an, weshalb die Gutachter an diesen von außen initiierten und mit erheblichem Arbeitsaufwand versehenen Evaluationen teilgenommen haben. Bei diesen ist auffällig, dass erst auf explizite Nachfrage eigene Motivationen genannt werden und keiner der interviewten Gutachter überhaupt in Erwägung zieht, die Anfrage als Gutachter mitzuwirken, abzulehnen. Die Teilnahme an einer Evaluation ist für die Gutachter eine Selbstverständlichkeit, die auf einer Pflicht gegenüber der scientific community beruht, der man sich nicht entziehen darf: „I think part of my responsibility is not only to do my job in isolation, but to do my job in the context of the community that I relate to, and it’s in the interest of that community“. Evaluationen als ‚community work‘ zu verstehen, kann verschiedene Ausformungen haben. Es gibt Gutachter, die damit eine prinzipielle Förderung des eigenen Wissenschaftsgebiets in Verbindung bringen, aber auch solche, die intervenieren wollen: „to get rid of all the dead wood“. Die durch Erwartungsunsicherheit und Misstrauen gekennzeichnete Evaluationssituation, die Gefahr unangemessener Urteilsweisen<sup>13</sup> und das Wissen, dass mit dem Urteil der Gutachter einschneidende Konsequenzen für die Evaluierten verbunden sind, fördert aber vor allem eine „Hochachtung vor der Aufgabe“ und folglich einen „honest intent to do a good job“ bei den Gutachtern zu Tage. Um diesem Anspruch gerecht werden zu können, stößt man auf eine Grundhaltung, die zwei Pole zugleich berücksichtigt: „to make sure we did *the job properly* but that we were also *very fair to the community*“. Geteilte Standards (‚proper job‘) und ihre dem Fall angemessene Verwendung (‚fair job‘) gehören für die Gutachter zusammen.<sup>14</sup>

Die Pflicht zur ‚community work‘ bedeutet nicht, dass Gutachter keine, sondern spezielle Eigeninteressen mit ihrer Teilnahme als Gutachter verknüpfen: Gutachter fühlen sich geehrt, diese Rolle zu übernehmen, sehen in der Anfrage einen Hinweis, „nicht völlig ohne Spuren durch’s Wissenschaftssystem“ gelaufen zu sein, interessieren

<sup>13</sup> Überwiegend beziehen die Gutachter diese Gefahr nicht auf Ereignisse im eigenen Panel, sondern auf andere.

<sup>14</sup> Auch Mallard, Lamont und Guetzkow (2009) sind bei ihren Untersuchungen über Verfahren der Stipendienvergabe auf die Regel gestoßen, dass in Gutachtergruppen „Fairness as Appropriateness“ aufgefasst wird. Beim Urteil nur theoretischen und methodischen Vorlieben zu folgen, ohne die Ziele des Antrags zu berücksichtigen, gilt als unangemessen. Ob und auf welche Weise diese doppelte Grundhaltung in den Gutachterteams vor Vereinseitigungen gesichert und wirksam wird, wird im nächsten Abschnitt behandelt, wenn wir uns den Begutachtungspraktiken zuwenden.

sich für das Innenleben der jeweiligen wissenschaftlichen Einrichtung, können einen Überblick über die Entwicklung und den aktuellen Stand von Forschungsgebieten bekommen, sind auf die Sicht- und Begründungsweise der Gutachterkollegen gespannt, können über den Ablauf solcher Evaluationen etwas für die zukünftige Beurteilung der eigenen Institution lernen und wollen schließlich die professionelle Selbstkontrolle gewahrt wissen: „I think it’s, since somebody has to do it, it’s better that [...] you don’t refuse to participate“. Die Gutachter machen aus der Not, extern eingeforderte Leistungsüberprüfungen durchführen zu sollen, eine Tugend, indem sie ihre Tätigkeit als Evaluatoren mit wissenschaftsimmanenten Beweggründen durchsetzen. Man stößt auf Interpretationsweisen, die Evaluationen in ein Licht rücken, als wären sie ein Element der akademischen Profession und könne an dort vorherrschende Grundorientierungen und Praxisformen angeschlossen werden. Die unhinterfragte Selbstverständlichkeit der Teilnahme, die verbreitete Auffassung unter den Gutachtern, dass gelegentliche Leistungskontrollen nötig und hilfreich sind und der ausbleibende Zweifel, selbst urteilen zu können<sup>15</sup>, obwohl ein neuer Gegenstand mit zusätzlichen Gütekriterien zur Debatte steht, liefern erste Hinweise hierfür. Inwiefern handelt es sich hierbei aber nur um Versuche der Selbstlegitimierung und –illusionierung, die praktisch folgenlos sind? Kann ein solcher Anschluss gelingen?

Betrachten wir hierzu, wie die Gutachter sich in einer Situation verorten, in der ihr Urteil gleich mehreren Funktionen dient und in der ungewiss ist, ob ihre Loyalität wissenschaftspolitischen Instanzen und Kriterien gilt oder doch den eigenen Kollegen.<sup>16</sup> Die Gutachter antworten auf die Frage, wozu die von ihnen erbrachten Bewertungen in erster Linie dienen, immer wieder, dass sie eine *für* die Institute und Wissenschaftsbereiche wichtige und möglichst belastbare Informationsgrundlage erzeugen. Selbst im dafür wenig geeigneten RAE, das keine inhaltlichen Rückmeldungen, sondern letztlich ein fachbezogenes, numerisches Rating erarbeitet, welches der Politik als Verteilungsschlüssel dient, stellen Gutachter ihre Tätigkeit als eine Form der *kollegialen Unterstützung* dar: „The first reason [for the RAE] which is probably the only valid one, really is to help universities benchmark their research against their competitors“. Dabei gibt es ein Selbstverständnis, dass die Möglichkeit sich untereinander vergleichen zu können, an sich zur Verbesserung der Forschungsleistungen anhält: „because, people who don’t do so well think ‘well let’s show these buggers what’s going on’ and let’s do a better [job...] ‘let’s really improve our act and learn from this’.“ Die Gutachter stehen allerdings vor dem Problem, dass sie nicht kontrollieren können, von wem und in welcher Weise die von ihnen erarbeiteten und veröffentlichten Informationen interpretiert und verwendet werden. Ihnen ist es vom Verfahren sogar verboten, über Interna der Begutachtung zu reden und den evaluierten Instituten ausführlich Rückmeldung zu geben. Ihre Kritik, dass in den Medien oder in den Universitäten ihre disziplinspezifischen Urteile in Rankings ganzer Institutionen münden und als Vergleichsfolie zwischen verschiedenen Wissenschaftsgebieten genutzt werden, die Klage, dass sich die Wissenschaftspolitik die Mittelverteilung nicht vollständig an den Ergebnissen der Begutach-

<sup>15</sup> Eva Barlösius (2008) nennt dies eine „Urteilsgewissheit“, die Gutachter selbst haben und die ihnen als Vertreter des wissenschaftlichen Feldes auch zugeschrieben wird.

<sup>16</sup> Uwe Schimank (2004) hat den Loyalitätskonflikt, der sich aus einer Konstellation ergibt, „in der ein Berater einem Beratenem empfiehlt, wie mit Dritten umzugehen ist, die Fachkollegen des Beraters sind“, drastisch als einen Übergang der informellen Norm des „Takts“ zum „Verrat“ bezeichnet.

tung orientiert (Roberts 2003) oder über Regelbrüche, bei denen Gutachter trotz Verbot dennoch in Kontakt zu den evaluierten Instituten treten, „to explain what that means, what it doesn't mean“, sind Ausdruck und Folge des Grundproblems, dass die Evaluationsergebnisse beginnen, ein Eigenleben zu führen (vgl. Espeland/Sauder 2007).

In den niederländischen und deutschen Verfahren, die neben Förderempfehlungen konkrete Hinweise zur inhaltlichen und strukturellen Weiterentwicklung einzelner Institute geben sollen, tritt sogar ein Selbstverständnis der Gutachter als *kollegial beratende Instanz* deutlich hervor: „it was not as much to find faults but, are they going in the right direction? And are there points of improvement? Can things be done in a better way?“. Diese Rolle als Berater einzunehmen, steht allerdings ebenfalls vor verfahrensbedingten Schwierigkeiten. Denn die beratenden Institute handeln in erster Linie strategisch, geben Problemlagen nicht unbedingt offen preis und erhalten am Ende verbindliche Urteile und Handlungsempfehlungen, deren Umsetzung bei der nächsten Evaluation überprüft wird: „So it's a combination, because advising is too, too optional, as if you can do with my advise what you want, you don't have to“.

Das Selbstverständnis der Gutachter, Instituten und Wissenschaftsbereichen unterstützende und Verbesserungen anstoßende Informationen und Vorschläge zu liefern, stößt auf strukturell ungünstige Situationsbedingungen. Die Gutachter rekurren zwar weiterhin auf Handlungsformen (valide Informationsbasis, kollegiale Beratung), die zum Kollegialitätsprinzip und zu professionsinternen Begutachtungsstandards (regel- und fallorientiert) passen, deren selbstverständliche Geltung aber im Kontext von Evaluationen und bei der Verwendung der erzielten Ergebnisse nicht mehr gesichert ist.

## 5. Handlungs- und Urteilsweisen von Gutachterteams

Wendet man sich den Entscheidungsprozessen in Gutachterteams zu und fragt, welche Regeln dort praktisch in Anspruch genommen und welche zurückgewiesen werden, dann stößt man durchaus auf geteilte Grundverständnisse, die allerdings erst in der Interaktion von Gutachtern hervortreten und gegen Devianz gesichert werden müssen. Da bislang wenig über Entscheidungsstrukturen und -prozesse von Gutachterteams bekannt ist<sup>17</sup>, stehen diese im Folgenden im Zentrum der explorativen Analyse. Die Begutachtungspraktiken werden entlang der beiden untersuchten Evaluationstypen vorgestellt, weil die Evaluationsverfahren den Rahmen vorgeben, anhand welcher Materialien Gutachter ihr Urteil treffen und auf welche Weise sie dieses kommunizieren sollen.

### 5.1 Begutachtung im Research Assessment Exercise (RAE)

Im britischen RAE treten die Gutachtergruppen mehrfach und lange vor der eigentlichen Evaluation zusammen. Sie müssen zumindest die groben Linien ihrer „Panel criteria and working methods“ festlegen und veröffentlichen, bevor die Evaluierten ihre Unterlagen einreichen. Sie entscheiden über die Gewichtung zwischen den drei Bewertungsdimensionen „Output“, „Esteem“ und „Environment“, über die Zuteilung von Informationen zu den Kategorien und über den Anteil der detailliert begutachteten Outputs. Die Bewertungsdimensionen werden also an das jeweilige Wissenschaftsgebiet

<sup>17</sup> Vgl. aber die Arbeiten von Travis/Collins 1991, Langfeldt 2001, Hirschauer 2005, Lamont 2009, Olbrecht/Bornmann 2010.

angepasst und mit den dort üblichen Errungenschaften inhaltlich spezifiziert, damit Bewertungskriterien möglichst im Sinne und zum Vorteil der in einem Wissenschaftsfeld geltenden Qualitätsvorstellungen ausgelegt werden. Bei diesen ersten Entscheidungsprozessen geht es darum, im Gutachtergremium zu klären, woran in einem Forschungsgebiet die Werthaltigkeit von Arbeiten festgemacht werden kann und welche Informationsquelle evidente Urteile zulässt. Vom Verfahren gesetzt und von den Gutachtern unterstützt, hat die Beurteilung wissenschaftlicher Erzeugnisse den Vorrang, weil man sein Urteil auf Evidenzen im jeweiligen Text stützen kann, wohingegen andere Informationsquellen (z.B. Strategiepapiere) wenig verlässlich, in ihrer Bedeutung für die eigentlichen Forschungsergebnisse kaum einschätzbar und deshalb im Gutachterkreis auch strittig sind. Die Kategorien „Esteem“ und „Environment“ dienen als Korrektiv für Disziplinen, die stark anwendungsbezogen sind und für die Publikationen nicht unbedingt im Zentrum stehen.<sup>18</sup> Die Qualität wissenschaftlicher Elementarakte ins Zentrum der Begutachtung zu stellen, hat gleich mehrere Folgen. Zum einen werden damit akademische Wertorientierungen gegenüber Wissenschaftsbereichen geschützt, die diesen nicht unbedingt folgen wollen.<sup>19</sup> Zweitens wird der zu begutachtende Fall ‚wissenschaftliche Einrichtung‘ in Einzelakte aufgelöst und auf diese Weise handhabbar gemacht. Die Gutachter können so an eine aus anderen Peer Review-Prozessen anerkannte und eingeübte Praxis anschließen. Die zentrale Bedeutung der sachhaltigen Beurteilung von Einzelerzeugnissen führt schließlich drittens dazu, dass die Gutachter statt einer möglichen Auswahl alle eingereichten Outputs (bis über Tausend!) inhaltlich beurteilen.

Beobachtet man, auf welche Weise das Panel ihr immenses Pensum bearbeitet, dann tritt die zentrale Bedeutung professionsinterner Standards deutlich zu Tage. Eine Reduktion der Arbeitslast, den Verzicht auf eine (zumindest) von zwei Gutachtern vorgenommene Bewertung oder eine schematisierte Beurteilung entlang von Impactfaktoren und Zitationshäufigkeiten wäre in dieser Lage zwar erwartbar, findet aber nicht statt.<sup>20</sup> Der prinzipielle Einwand gegen eine solche vielleicht effektivere Vorgehensweise ist die geringere Verlässlichkeit des Urteils gegenüber der Lektüre.<sup>21</sup> Statt solche Abkürzungsstrategien zu verwenden, investieren die Gutachter mehr Zeit und kritisieren folglich den überhand nehmenden „workload“ des Verfahrens: „the major

<sup>18</sup> Zum Beispiel ist für den Bereich „Engineering“ die Einwerbung von Industriegeldern ein Beleg für den Erfolg, so dass Drittmittel als Reputation verstanden und folglich der Kategorie „Esteem“ zugeordnet werden, wobei diese Bewertungsdimension mit 30% einen auffällig starken Einfluss auf das Gesamtergebnis hat. Zugleich wird die Bedeutung der Outputs mit den minimal zu veranschlagenden 50% hier am geringsten bewertet.

<sup>19</sup> So kommen Ben Martin und Richard Whitley zu dem Schluss, das RAE stärke „basic rather than applied“, „monodisciplinary rather than inter- or multidisciplinary“ und „academic rather than professional research“ (2010: 72).

<sup>20</sup> Das RAE wird jedoch zukünftig durch das „Research Excellence Framework“ (REF) abgelöst. Dort spielt der erzielte Impact von Outputs eine größere Rolle und die Gutachter können ihr Urteil an metrischen Daten spiegeln. Da es ihnen freigestellt ist, ob und wie solche bibliometrischen Informationen genutzt werden, darf man gespannt sein, ob professionelle Standards zukünftig an Stellenwert einbüßen oder nicht. In unseren Gutachterinterviews galten bibliometrische Verfahren jedoch als unzuverlässig.

<sup>21</sup> Viele Gutachter erläutern dies gerne am Beispiel von Review-Artikeln. Diese werden häufig zitiert und gerne auch in angesehenen Zeitschriften veröffentlicht, weil sie eine wichtige Informationsfunktion in der Wissenschaft haben. Ohne zu lesen, könne man allerdings nicht wissen, ob es sich nur um eine Zusammenfassung von Bekanntem oder doch um einen eigenständigen Beitrag handelt.

challenge for quality was getting the job done“. Um dennoch eine angemessene inhaltliche Bewertung einer großen Menge an Einzeloutputs gewährleisten zu können, werden im Panel spezielle Rationalisierungsstrategien entwickelt. Eine erste und besonders wichtige ist, dass im heterogen besetzten Gutachterteam ein Konsens über die Qualität eines jeden Outputs erzeugt werden soll. Einzelbewertungen der Gutachter einfach zu addieren, genügt nicht, sondern es bedarf ein kollektiv geteiltes Gesamturteil. Dazu werden in so genannten „calibration sessions“ anhand exemplarisch ausgewählter Arbeiten geteilte Standards des Begutachtens entwickelt, überprüft und durch die Offenlegung der einzelnen Gutachtervoten im Team über den gesamten Evaluationsprozess hinweg gesichert: „very quickly you create an ethos for the panel“. <sup>22</sup> Daran ist zum einen bemerkenswert, dass entgegen der Verfahrensordnung <sup>23</sup> nicht nach absoluten Maßstäben gesucht wird, sondern aus der Menge vorhandener Arbeiten typische Referenzfälle für hohe bzw. niedrige wissenschaftliche Qualität ausgewählt und verglichen werden. Würde man absolute Maßstäbe anlegen, so argumentieren Gutachter, dann könnte die Bestnote nur äußerst selten an Arbeiten vergeben werden, da sie das Potential für einen Paradigmenwechsel haben müssten. Insofern handelt es sich nicht um ein absolutes, sondern um ein relatives Urteil über realistisch erwartbare Forschungsleistungen. Zum anderen waren die Gutachter überrascht, dass sowohl in den „calibration sessions“ und der nachfolgenden Begutachtung aller Einzeloutputs, eine hohe Übereinstimmung in den Einzelbewertungen vorhanden war.

Worauf lässt sich dieses Phänomen gleicher Urteile von heterogenen Gutachtern zurückführen? Wir stoßen hierbei auf eine weitere Rationalisierungspraktik. Die Einzeloutputs werden nämlich nicht in aller Ausführlichkeit, sondern nur auf wesentliche Aspekte hin betrachtet. Hierdurch lässt sich zum einen eine erhebliche Beschleunigung der Begutachtung erzielen und zum anderen besteht bei der grundsätzlichen Einschätzung des wissenschaftlichen Wertes von Beiträgen Einigkeit, weil die Gutachter mit grundlegenden Standards operieren. Ist die grundsätzliche Frage geklärt, ob es sich überhaupt um einen eigenständigen Forschungsbeitrag handelt, richtet sich die Aufmerksamkeit der Gutachter auf den vom vorliegenden Text selbst gesetzten Anspruch. Neuartigkeit und Generalisierbarkeit, Klarheit und Konsistenz der Argumentation und die Überprüfung, ob die bereitgestellten Evidenzen auch die gezogenen Schlussfolgerungen erlauben, sind in der Wissenschaft schon immer herangezogene Deutungsdimensionen. Im Fokus der Gutachter steht also die Textgestalt. Vermag der Text keine gelungene Kommunikation entlang dieser Dimensionen herstellen, dann schließen Gutachter auf Schwächen der unternommenen Forschung:

*„So if you as a clever scientist, engineer, sociologist, are writing it badly, I can't judge the quality of the research as well so it's bound to get a lower score, isn't it? If you, I mean surely that is*

<sup>22</sup> Natürlich gibt es in Gutachterteams Gutachter, die nach Gesichtspunkten urteilen, die nicht zu einem solchen „Ethos“ passen. Ein solches Verhalten gilt jedoch als nicht akzeptable Devianz und setzt Maßnahmen frei, die auf die Einhaltung der Norm zielen. So berichtet ein Vorsitzender von einem Kollegen, dessen Voten auffällig von anderen abwichen. Es stellte sich heraus, dass er Arbeiten aus einem speziellen, von ihm selbst vertretenen Bereich stets positiv, andere jedoch negativ beurteilte. Daraufhin wurde er aufgefordert, Voten nicht an seinen Präferenzen, sondern an der Qualität der jeweiligen Publikation zu verteilen.

<sup>23</sup> Dort heißt es explizit: „Each submission will be assessed against absolute standards and will not be ranked against other submissions.“



*the art of communication, so if you're asking the reader to think too much, there's a problem isn't there? I mean it's all about communication“.*

Eine dritte Rationalisierungsform lässt sich schließlich im Umgang mit abweichenden Gutachtervoten erkennen. Wenn überhaupt, dann werden nur Outputs mit erheblich abweichenden Urteilen diskutiert und gegebenenfalls ein entscheidungsgebender dritter Gutachter hinzugezogen. Bei weniger großen Abweichungen, ist die Bereitschaft, einen Konsens zu erzielen und die Voten anzugleichen, hingegen groß. Das ist ein klarer Effekt des Verfahrens, muss man doch in Rechnung stellen, dass die Bewertung eines einzelnen Outputs unter hunderten für die kumulative Gesamtnote kaum ins Gewicht fällt.

An den geschilderten Entscheidungsprozessen der Gutachterteams im RAE ist bemerkenswert, dass die Gutachter selbst unter widrigen und bis an ihre Belastungsgrenze gehenden Umständen, nicht einfach auf professionsinterne Standards verzichten. Man gewinnt sogar den Eindruck, dass diese in Gutachtergruppen gestärkt und bei der Anwendung radikalisiert werden, damit eine angemessene Form der Beschleunigung erzielt werden kann.

### *5.2 Begutachtung nach dem Standard Evaluation Protocol und in der Leibniz-Gemeinschaft*

Die Anforderungen an die Gutachterrolle unterscheiden sich beim zweiten Evaluations-typ in wesentlichen Hinsichten von denen des RAE. Erstens soll ein einzelnes Institut und keine ganze Forschungslandschaft evaluiert und sollen fallspezifische „Empfehlungen“ formuliert werden, womit die Gutachterrolle auch beratungsartige Züge annimmt. Zweitens ist das Gutachterteam interdisziplinär zusammengestellt und trifft seine Entscheidung nicht auf Grundlage einzelner Forschungsbeiträge, sondern auf Basis eines Globalberichts des Instituts, den Präsentationen der Forschungseinheiten und der direkten Interaktion mit Institutsangehörigen.

Der Beurteilungsprozess ist dreistufig angelegt. Die einzelnen Gutachter gewinnen anhand der eingereichten Unterlagen einen ersten Eindruck, der in der direkten Interaktion überprüft und ins Gutachtergremium eingespeist wird, wo er sich bewähren muss, um als kollektives Urteil gelten zu können. Betrachten wir im Folgenden genauer, wie die Gutachter und das Gutachterteam ihr Urteil bilden und welche Grundorientierungen dabei zum Tragen kommen.

Zunächst geht aus den Interviews hervor, dass die speziellen Verfahrensordnungen und Kriterienkataloge den Gutachtern kaum präsent sind, weniger während der Urteilsfindung, sondern erst bei der Abfassung des Bewertungsberichts Beachtung finden<sup>24</sup> und die sehr umfangreichen Unterlagen „in der Regel auch nicht gelesen werden oder sehr selektiv gelesen werden“. Dieser Selbstverständlichkeit kann man bereits entnehmen, dass die Gutachter wenn überhaupt, dann impliziten Regeln des Beurteilens folgen. Die Gutachter wissen, dass die Institute im Kontext von Evaluationen strategisch agieren und sie mit dem Bericht überzeugen wollen, so dass ihr Handeln vornehmlich darauf ausgerichtet ist, die sachliche Angemessenheit, Glaubwürdigkeit und Realisier-

<sup>24</sup> Die Verfahrensordnung der Leibnizgemeinschaft stützt diese Vorgehensweise, wenn es heißt: „Die Kriterien werden bei der Erarbeitung des Bewertungsberichtes zugrunde gelegt“.

barkeit dieser Selbstbeschreibungen zu überprüfen. Dazu rekurrieren Gutachter auf fachspezifisch ausgeprägte „ganz bestimmte Normen im Kopf“ und auf eine intuitive „Pi-mal-Daumen-Beurteilungskraft“. Sie spiegeln die Selbstdarstellungen des Instituts an Normalitätstsvorstellungen und zusätzlichen Informationen, die teilweise in Form von harten Indikatoren (Outputs, Infrastruktur, Finanzierung) zur Verfügung gestellt werden, aber den Gutachtern „auch ohne [...] Unterlagen“ gegenwärtig sind, weil die Gutachter auf ein „gewisses Vorwissen“ rekurrieren können. Dieses Vorwissen ist ein Erfahrungswissen, dass die Besonderheiten des wissenschaftlichen Themenbereichs und der institutionellen Einbettung des zu begutachtenden Instituts ebenso umschließt, wie erwartbare Publikationsleistungen, Qualifikationszeiten oder typische Karriereprobleme des wissenschaftlichen Nachwuchses. Es handelt sich dabei also weniger um ein spezialisiertes und formalisiertes Ableitungswissen („irgendwie weiß man das“) als um eine generelle Kenntnis des Handlungsfeldes mit seinen typischen Herausforderungen und Problemlagen, mit dem der jeweilige Fall dann (extern) kontrastiert und abgeschlossen werden kann – analog zu dem typologischen Wissen über Standardfälle, Symptomatiken oder Motivkonstellationen im juristischen und ärztlichen Handeln.

Zudem nehmen die Gutachter eine textimmanente Konsistenzprüfung vor, denn die „Papiere müssen in sich logisch sein oder ne gewisse, einen Zusammenhang aufweisen“. Wie wir bereits bei der Begutachtung von einzelnen Outputs im RAE gezeigt haben, geben Inkonsistenzen, Widersprüchlichkeiten oder fehlende Explikationen den Anlass für Nachfragen und Kritiken oder die Suche nach Verbesserungsvorschlägen. Oder sie nähren den Verdacht, dass die Autoren selbst nicht genau wissen, wofür sie was tun. Auf diese Weise wird bereits beim Studium der Akten nach möglichen Problemlagen und Indizien für und wider die Glaubwürdigkeit oder Angemessenheit der Selbstdarstellungen gesucht, die dann während der direkten Interaktion in der Institutsbegehung eine weitere Überprüfung erfahren. Der Maßstab ist: „die Papiere [müssen] mit dem was die Leute sagen in Übereinstimmung stehen, so dass es nicht auseinanderfällt“.

Bevor die eigentliche Begehung beginnt, bringen die Gutachter bei ihrer ersten Zusammenkunft ihre anhand der Unterlagen gewonnenen Eindrücke in das Gremium ein. Einzelsichtweisen können dabei von anderen unterstützt oder bestritten werden, sich dadurch verfestigen oder verflüssigen und weitere Informationen oder ein anderer Ablauf der Begehung können eingefordert werden. Über solche diskursiven Prozesse bildet sich beim ersten Zusammentreffen der Gutachter ein vorläufiger Gesamteindruck und eine kollektive Einschätzung darüber heraus, ob es sich im Fortgang um eine mehr oder minder problematische Evaluation handeln wird und ob der inhaltliche und finanzielle Fortbestand des Instituts in Frage steht oder die Suche nach Verbesserungsmöglichkeiten die Evaluation anleitet. Da in den meisten Fällen der Extremfall einer Institutsschließung nicht zur Debatte steht, verschiebt sich die Doppelrolle von Gutachtern stärker in eine Richtung: „Das hat sich viel stärker eben in Richtung Berater als in Kontrolleur sag' ich jetzt mal entwickelt im Laufe der [Evaluation]“.

Die direkte Interaktion mit den Mitarbeitern der evaluierten Einrichtung ist die zweite wichtige Informationsquelle für die Gutachter. Im Unterschied zu den eingereichten Unterlagen können sie aus der direkten Beobachtung der Handlungsweisen und des Antwortverhaltens von Institutsmitarbeitern weniger geschönte Anhaltspunkte für ihr Urteil erhalten.

*„Das war mir schon ein bisschen überorganisiert, das muss man noch ein bisschen aufbrechen [lacht], aber das sind dann also ja die Gespräche mit den Leuten vor Ort, die einem also auch helfen, einzuordnen, wie dort das Alltagsgeschäft läuft, ob es wirklich alles so perfekt ist oder nur dafür [die Evaluation, der Verf.] organisiert ist.“*

Zunächst liefert die Interaktionsweise unter den Mitarbeitern „gewisse nicht verbale Indikatoren“, ob die Formen der Zusammenarbeit im Institut angemessen sind. Die Gutachter deuten die Handlungs- und Kommunikationsformen im Institut entlang von impliziten Standards aus. Wenn zum Beispiel ersichtlich ist, dass nur der „Chef [...] redet und die andern sitzen als Statisten dabei“, dann ist dies ein Hinweis auf zu hierarchische Organisationsformen, und wenn umgekehrt „fünf reden und [...] überhaupt keine Ordnung da ist“, auf mangelnde Koordination.

Anhand der genauen Beobachtung der Reaktionsweisen auf Fragen versuchen die Gutachter herauszufinden, ob Darstellung und Wirklichkeit des Instituts stimmig sind:

*„Also das ist eben die Art, antworten die Leute auf Fragen, die man ihnen stellt, auch auf kritische Fragen und wie gehen sie mit diesen Fragen um? Weichen sie denen aus, beantworten sie die gar nicht, beantworten sie die klar? Wenn sie die klar beantworten und auch ein Problem eingestehen, ist das im Prinzip schon mal ein Indiz, dass es in die richtige Richtung geht. Und wenn sie ein realistisches Selbstbild auch haben, wie sie sich selber einschätzen, ist das auch ein positives Indiz. Also das sind Indizien, die was mit Glaubwürdigkeit zu tun haben.“*

Diese Vorgehensweise ist in mehreren Hinsichten bemerkenswert. Zunächst richtet sie sich an einem allgemeinen Strukturproblem von Evaluationen aus. Um nämlich ein sachlich angemessenes Urteil treffen zu können, sind die Gutachter auf eine glaubwürdige Offenlegung von internen Informationen angewiesen. Davon können sie jedoch in einem prüfungsartigen Kontext von stets unfreiwilligen und potentiell einschneidenden Evaluationen gerade nicht ausgehen. Die Handlungssituation Evaluation ist vielmehr durch ein strukturelles Misstrauen gekennzeichnet, selbst wenn dort Kollegen aufeinander treffen. Während das Handeln der Begutachteten darauf abzielt, Problemlosigkeit zu kommunizieren, ist das gutachterliche Handeln ganz im Gegenteil auf die Suche nach Indizien für Problemlagen ausgerichtet. Die von den Gutachtern in Anspruch genommene Situationsdeutung einer nach Verbesserungsmöglichkeiten suchenden, kollegialen Beratung oder Supervision erfolgt aber in einem ungünstigen, unter Kontroll- und Sanktionsvorbehalt stehenden Handlungskontext, und kann sich deshalb nicht stabil einrichten. Das zeigt sich bis hinein in die Ergebnisformulierung, bei der zwar nur „Empfehlungen“ ausgesprochen werden sollen, schriftlich aber verbindliche Handlungsaufforderungen fixiert werden, deren Umsetzung in der nächsten Evaluation überprüft wird.

Die Ergebnisse des Evaluationsprozesses sind wenig vorhersehbar, weil man trotz allen Vorbereitungsmaßnahmen nicht wissen kann, auf welche Widersprüche die Gutachter in den Unterlagen und während der Interaktion stoßen oder welche Einwände und Vorschläge das Gutachterteam passieren und zu einem Kollektivurteil werden und welche nicht. Es kann zum Beispiel passieren, dass ein Nebenaspekt ins Zentrum rückt, sofern dieser im Gutachterteam auf Unterstützung trifft und sich so diskursiv festsetzt, dem man kaum Relevanz beigemessen hat.<sup>25</sup> Bei der speziellen Interaktionsform in

<sup>25</sup> So attestierten in einem Fall die Gutachter einem Institut hervorragende Leistungen in allen Dimensionen. Nur die relativ langen Promotionszeiten fielen ihnen auf, so dass dieser Aspekt ins Zentrum der Begutach-

Gutachterteams greifen diskursive, normbildende, disziplinierende und die Urteilsfindung beschleunigende Momente ineinander, die ausufernde Diskussionen verhindern und erklären, warum es den analysierten Interviews zufolge nur sehr selten zum offenen Streit unter Gutachtern kommt:

*„Also es ist oft so, ich meine bei solchen Gutachtersitzungen ist selten, dass einer sagt, das sehe ich ganz anders, Gutachten laufen meist so ab, dass einer was sagt und wenn's keiner aufgreift, dann war es ein Flop [...] und ein guter Vorsitzender sagt dann nicht um Gottes Willen, was haben die da gesagt, sondern geht dann zum Nächsten über und wenn es keiner mehr aufgreift, ist es tot“.*

Kommunikative Anschlussfähigkeit zu erlangen ist ein wesentlicher und kaum berechenbarer Selektionsmechanismus in Gutachtergruppen, bei dem Gutachter beobachtet werden, wie sie begutachten, ob sie geteilten Standards oder Sonderperspektiven folgen. So wird auch die Verfolgung von eigenen Interessen sichtbar und sie werden nach dem gleichen diskursiven Prinzip des Nichtaufgreifens ausgeschlossen, sofern Einwände oder Vorschläge sich weder aus dem diskursiven Zusammenhang erschließen, noch vor den Gutachterkollegen begründet werden:

*„[Gutachter X] hatte das vorgebracht, die Kenner guckten sich an und irgendwo merkten alle andern, klar, also war so was Spezielles, das war jetzt nicht irgendwo aus der Diskussion, dass irgendetwas an- und ableitbar [gewesen wäre], so dass jeder wusste, das kann nur deren Interesse sein.“*

Fasst man die Begutachtungspraktiken und Rollenverständnisse dieses Evaluationstyps zusammen, dann fällt auf, dass das Gutachterhandeln weniger expliziten Verfahrenskriterien denn impliziten „Normen im Kopf“ folgt und trotz ungünstiger Umstände die Rolle des kollegialen Beraters oder Supervisors zentral ist. Das bereits im Verfahren angelegte Strukturproblem führt dazu, dass diese Rolle bisweilen in die des investigativen Detektivs übergeht, der erst hinter die Fassade geschönter Selbstberichterstattungen gelangen muss, bevor er Probleme entdecken und einer Bearbeitung zugänglich machen kann.<sup>26</sup> Gutachter sollen verbindliche Empfehlungen gegenüber Kollegen aussprechen, die darum nicht gebeten haben. Probleme werden deshalb auch nicht feilgeboten, sondern sie müssen erst gefunden und aufgedeckt werden.

## 6. Der Evaluator im Spannungsfeld widerstreitender Situations- und Rollendeutungen

Das Ziel dieses Beitrages war es, einen Einblick in die Innenwelt einer neuen Form der Leistungsbewertung von Wissenschaft zu geben, die in der Wissenschafts- und Professionssoziologie gemeinhin als Deprofessionalisierungserscheinung der akademischen Profession verhandelt wird (Stichweh 1997: 99f., Kurtz 2005: 171f. Oevermann 2005:

---

tion rückte und Neuinterpretationen anleitete. Die hervorragenden Publikationen waren jetzt ein Indiz für ein überzogenes Anspruchsniveau, die einzigartigen Apparaturen galten als Hinweis für Überspezialisierungen und Überbelastung von Doktoranden mit der Aufrechterhaltung der technischen Infrastruktur.

<sup>26</sup> Tilman Allert (1998: 26) verortet den Supervisor vor allem im Gebiet professioneller Tätigkeiten generell „im habituellen Umfeld zwischen Detektiv und Jongleur“. Das an sich attraktive Angebot des Supervisors, Alltagsroutinen sichtbar zu machen, einer Reflexion auszusetzen und gegebenenfalls zu überdenken, stößt auf eine prinzipielle Abwehr, weil Professionelle ihrem Selbstverständnis nach „Selbstsupervisoren“ (ebd.: 16) sind.

33, Schimank 2005). Dabei konzentrierte sich die Analyse auf die in den Verfahrensordnungen institutionalisierte, auf die von den Evaluierten erwartete und auf die von den Evaluatoren ausgeübte Gutachterrolle. Was kann man abschließend aus den hier vorgestellten Praktiken, Deutungen und Rollenverständnissen hinsichtlich der allgemeinen Diagnose, dass Evaluierung zur Deprofessionalisierung führe, lernen?

Insgesamt führt die Analyse zu facettenreicheren und dieser Diagnose teilweise sogar widersprechenden Eindrücken, die zum Überdenken der über verschiedene theoretische Lager hinweg ja doch recht einhellig formulierten Deprofessionalisierungsthese anregen. So konnte gezeigt werden, dass sich klassische Merkmale einer Profession wie professionelle Selbstkontrolle, die Orientierung am Kollegialitätsprinzip und die Fallorientierung des professionellen Urteils im Rahmen von Evaluationen aus Sicht der an Evaluationen Beteiligten nicht auflösen. Wenn aber zentrale Bezüge der akademischen Profession weiterhin Geltung beanspruchen, dann stellen sich neue Fragen an die Deprofessionalisierungsthese und an das Professionskonzept, die hier nur aufgeworfen aber nicht beantwortet werden können.

1. Die extern initiierten Evaluationsverfahren weisen den wissenschaftsintern rekrutierten Gutachterteams eine zentrale Position zu und räumen ihnen im Entscheidungsprozess erhebliche Freiheiten ein. Die Gutachter sollen abstrakt formulierte und stark interpretationsbedürftige Kriterien an die Eigenheiten von speziellen Wissenschaftsgebieten oder wissenschaftlichen Einrichtungen anpassen, verschiedene Kriterien fallbezogen gewichten und sie können sogar eigene Kriterien entwickeln. Wissenschaftsgemeinschaften haben also einen zentralen Stellenwert jenseits ihrer angestammten Handlungsbereiche, so auch im Grenzbereich zwischen Wissenschaft und Politik. Werden aber die scientific communities einfach für fremde Zwecke instrumentalisiert – sind sie somit nur Anwender und Interpretatoren extern vorgegebener Maßstäbe – oder gelingt es ihnen, ihren Zuständigkeitsbereich und ihre Praxisformen auf neue Tätigkeitsgebiete auszudehnen?

2. Auch das Kollegialitätsprinzip ist in Evaluationsprozessen nicht einfach hinfällig, wenn man den Beschreibungen der Evaluierten und Evaluierenden folgt. Die Gutachter sehen ihre Teilnahme als wichtige und verantwortungsvolle Pflicht zur ‚community work‘, der man sich nicht entziehen darf und verbinden mit ihr professionseigene Motivationen, angefangen beim Nachweis, Reputation im Feld erlangt zu haben, über ein Sachinteresse an Forschungsständen und Urteilsweisen von Kollegen bis zur Wahrung der wissenschaftlichen Selbstkontrolle. Bei den evaluierten Einrichtungen kann man eine Aktivierung von Kollegen zur Vorbereitung und internen Simulation von Evaluationsprozessen beobachten, bei denen auf mutmaßlich geteilte Einschätzungen zurückgegriffen wird und Verständigungsprozesse über solche ausgelöst werden. Ebenso finden sich in den Interaktionen der Gutachter Praktiken, fachübergreifend geteilte Standards des angemessenen Urteilens emporzuheben, zu normieren und gegen Devianz zu sichern. Ist das alles nur Ausdruck professioneller Ideologien, rhetorischer Außendarstellung oder professioneller Selbstüberschätzung und Selbstillusionierung? Oder kommt hier nicht doch ein professioneller Habitus zum Ausdruck, der selbst unter ungünstigen Rahmenbedingungen nicht zu tilgen ist?

3. Desweiteren finden sich Hinweise, dass der zentrale Grundsatz professionellen Handelns, allgemeine Regeln fallspezifisch zu verwenden, auch in den Evaluationsverfahren und -prozessen handlungsleitend ist. Die Verfahren überantworten diesen An-

spruch den Gutachtergruppen. Diese wehren schematisierte Beurteilungsweisen als unzuverlässig ab, setzen sich bis zur Belastungsgrenze und ohne Notwendigkeit dem Zwang aus, Unmengen an Publikationen zu lesen und entlang sehr allgemeiner, in der Wissenschaft akzeptierter Deutungsdimensionen (Neuheitswerte und Generalisierungsniveau, argumentative Klarheit und Konsistenz, Passung zwischen erhobenem Anspruch und beigebrachten Evidenzen) zu beurteilen. Sie suchen sogar den Anschluss an eine professionalisierte Praxis der kollegialen Unterstützung und Beratung, bei der die Suche nach Problemen nicht der vernichtenden Kritik und Sanktion dient, sondern der Entwicklung von Verbesserungsmöglichkeiten. Warum aber wählen Gutachter Vorgehensweisen, die im Kontext von Evaluationen nicht explizit verlangt und nur schwerlich durchzuhalten sind? Eine mögliche Antwort ist, dass die Gutachter Evaluationen als eine für die Wissenschaft anschlussfähige Praxisform gestalten wollen.

4. Eine stabil eingerichtete Handlungssituation oder sogar ein klar konturiertes „Arbeitsbündnis“ (Oevermann 1996) zwischen Evaluierten und Evaluatoren ist bislang allerdings nicht erkennbar, da auch strukturelle Widersprüche und ungeklärte Probleme sichtbar wurden. Auf der einen Seite erzeugt nämlich die ungefragte, nicht selbst initiierte und unter Sanktionsvorbehaltung stehende Überprüfung von Handlungsroutrinen strategische Handlungsweisen der evaluierten Institute, die an verschiedenste Rollenkonzeptionen des Evaluators anschließen und dazu dienen, den freien Blick und die Kritik der Gutachter zu führen damit bestenfalls der Eindruck von Problemlosigkeit oder zumindest von Problemlösungsfähigkeit in den Vordergrund rückt. Auf der anderen Seite stehen hingegen Gutachter, zu deren Selbstverständnis es gehört, ein hilfreiches kollegiales Feedback zu geben. Dafür müssen sie aber erst auf Problemsuche gehen und bei der Ergebniskommunikation im Blick behalten, dass sie letztlich nicht kontrollieren können, von wem, auf welche Weise, für welche Zwecke ihr einmal veröffentlichtes Urteil interpretiert und verwendet wird. Die Etablierung eines offenen Gesprächs zwischen Kollegen steht also vor strukturellen Schwierigkeiten. Ob solche Strukturprobleme von Evaluationen, die zugleich kontrollieren, sanktionieren, legitimieren und Verbesserungen anstoßen sollen, überhaupt bearbeitet werden können, ist zumindest fraglich.

Die Exploration von Rollenverständnissen, Deutungs- und Handlungsweisen in der Innenwelt von Evaluation liefert Hinweise, die zum Öffnen des Diskurs über Evaluationen und die damit in Verbindung gebrachten Deprofessionalisierungstendenzen der akademischen Profession Anlass geben. Den Blick verstärkt auf die konkreten Reaktionsweisen, Anpassungs- oder Formierungsprozesse der akademischen Profession in ihrer zweifelsohne veränderten und sich schon immer wandelnden institutionellen Umwelt zu richten, könnte eine für die Professions- und Wissenschaftssoziologie attraktive Forschungsperspektive sein.

### *Literatur*

- Allert, T. (1998): Braucht jede Profession eine Supervision? Soziologische Anmerkungen zum Supervisionsbedarf und zum Autonomieanspruch der Professionen, in: Berker, P.; Buer, F. (Hrsg.): Praxisnahe Supervisionsforschung. Felder – Designs – Ergebnisse, Münster, 16-45
- Barlösius, E. (2008): Urteilstgewissheit und wissenschaftliches Kapital, in: Matthies, H.; Simon, D. (Hrsg.): Wissenschaft unter Beobachtung – Effekte und Defekte von Evaluationen, Wiesbaden, 248-264

- Espeland, W. N.; Sauder, M. (2007): Rankings and Reactivity: How Public Measures Recreate Social Worlds, in: *American Journal of Sociology* 113, 1, 1-40
- Franzmann, A. (2008): Biographische Ursprungsconstellationen des Wissenschaftlerberufs, in: *Sozialer Sinn*, 9, 329-355
- Frederiksen, L. F.; Hannson, F.; Wenneberg, S. B. (2003): The Agora and the Role of Research Evaluation, in: *Evaluation*, 9, 149-172
- Gläser, J.; Laudel, G. (2007): Evaluation Without Evaluators: The Impact of Funding Formulae on Australian University Research, in: Whitley, R.; Gläser, J. (eds.): *The Changing Governance of the Sciences: The Advent of Research Evaluation Systems*, Dordrecht, 127-151
- Hirschauer, S. (2002): Die Innenwelt des Peer Review. Qualitätszuschreibung und informelle Wissenschaftskommunikation in Fachzeitschriften  
<http://www.sciencepolicystudies.de/dok/expertise-hirschauer.pdf> (Download am 19.04.2011)
- Hirschauer, S. (2004): Peer Review auf dem Prüfstand. Zum Soziologiedefizit der Wissenschaftsevaluation, in: *Zeitschrift für Soziologie*, 33, 62-83
- Hirschauer, S. (2005): Publierte Fachurteile. Lektüre und Bewertungspraxis im Peer Review, in: *Soziale Systeme*, 11, 52-82
- Hirschauer, S. (2010): Editorial Judgements. A Praxeology of 'Voting' in Peer Review, in: *Social Studies of Science*, 40, 71-103
- Hornbostel, S. (2010): (Forschungs-)Evaluation, in: Simon, D.; Knie, A.; Hornbostel, S. (Hrsg.): *Handbuch Wissenschaftspolitik*, Wiesbaden, 293-309
- Kieser, A. (2010): Unternehmen Wissenschaft?, in: *Leviathan*, 38, 347-367
- Krücken, G.; Meier, F. (2006): Turning the University into an Organizational Actor, in: Drori, G.; Meyer, J.; Hwang, H. (eds.): *Globalization and Organization: World Society and Organizational Change*, Oxford, 241-257
- Kurtz, T. (2005): *Die Berufsform der Gesellschaft*, Weilerswist
- Lamont, M. (2009): *How Professors Think. Inside the Curious World of Academic Judgment*, Cambridge
- Langfeldt, L. (2009): The Decision-Making Constraints and Processes of Grant Peer Review, and Their Effects on the Review Outcome, in: *Social Studies of Science*, 31, 820-841
- Leibniz-Gemeinschaft (2007): Kriterien für die Evaluierung von Einrichtungen der Leibniz-Gemeinschaft (SEN 0057/07, 18.07.2007)
- Loer, T.; Liebermann, S. (2006): Krise der Kritik. Die Misere der Universität, eine Krise der Kollegialität, in: *Forschung & Lehre* 6, 322-326
- Luhmann, N. (2000): *Organisation und Entscheidung*, Opladen
- Mallard, G.; Lamont, M.; Guetzkow, J. (2009). Fairness as Appropriateness: Negotiating Epistemological Differences in Peer Review, in: *Science, Technology and Human Values* (published online 2/10/09. <http://sth.sagepub.com/content/36/3/413.full.pdf+html>) (Download: 19.04.2011)
- Martin, B.; Whitley, R. (2010): The UK Research Assessment Exercise: A Case of Regulatory Capture?, in: Whitley, R.; Gläser, J.; Engwall, L. (eds.): *Reconfiguring Knowledge Production. Changing Authority Relationships in the Science and their Consequences for Intellectual Innovation*, Oxford, 51-80
- Matthies, H.; Simon, D. (2008) (Hrsg.): *Wissenschaft unter Beobachtung – Effekte und Defekte von Evaluationen*, Wiesbaden, 15-24
- Meier, F. (2009): *Die Universität als Akteur. Zum institutionellen Wandel der Hochschulorganisation*. Wiesbaden
- Meier, F.; Schimank, U. (2010): Mission Now Possible: Profile-Building and Leadership in German Universities, in: Whitley, R.; Gläser, J.; Engwall, L. (eds.): *Reconfiguring Knowledge Production. Changing Authority Relationships in the Science and their Consequences for Intellectual Innovation*, Oxford, 211-236

- Michaels, A. (2010): Die große Begehung der Mittelbaustelle, in: Frankfurter Allgemeine Zeitung, 11.08.2010, N5
- Neidhardt, F. (2000): Über die Kunst der Selbstverteidigung: und dass sie manchmal nichts nützt, in: Gegenworte 5, 26-29
- Neidhardt, F. (2010): Selbststeuerung der Wissenschaft: Peer Review, in: Simon, D.; Knie, A.; Hornbostel, S.: Handbuch Wissenschaftspolitik, Wiesbaden, 280-292
- Oevermann, U. (1996): Skizze einer revidierten Theorie professionalisierten Handelns, in: Combe, A.; Helsper, W.: Pädagogische Professionalität. Untersuchungen zum Typus pädagogischen Handelns, Frankfurt am Main, 70-182
- Oevermann, U. (2005): Wissenschaft als Beruf - Die Professionalisierung wissenschaftlichen Handelns und die gegenwärtige Universitätsentwicklung, in: Die Hochschule. Journal für Wissenschaft und Bildung 22, 15-51
- Olbrecht, M.; Bornmann, L. (2010): Panel Peer Review of Grant Applications: What Do We Know from Research in Social Psychology on Judgment and Decision Making in Groups?, in: Research Evaluation 19, 293-304
- Power, M. (1997): The Audit Society. Rituals of Verification, Oxford
- Power, M. (2008): Research Evaluation in the Audit Society, in: Matthies, H.; Simon, D. (Hrsg.): Wissenschaft unter Beobachtung – Effekte und Defekte von Evaluationen, Wiesbaden, 15-24
- Research Assessment Exercise (2008): <http://www.rae.ac.uk/> (Download 31.05.2011)
- Roberts, Sir G. (2003): Review of research assessment. Report by Sir Gareth Roberts to the UK funding bodies. <http://www.ra-review.ac.uk/reports/roberts.asp> (Download 30.05.2011)
- Schimank, U. (2004): Leistungsbeurteilung von Kollegen als Politikberatung – Am Beispiel von Evaluationen im Hochschulsystem, in: Schützeichel, R.; Brüsemeyer, T. (Hrsg.): Die beratene Gesellschaft. Zur gesellschaftlichen Bedeutung von Beratung, Wiesbaden, 39-56
- Schimank, U. (2005): Die akademische Profession und die Universitäten: “New Public Management“ und eine drohende Entprofessionalisierung, in: Klatetzki, T.; Tacke, V. (Hrsg.): Organisation und Profession, Wiesbaden, 143-164
- Stichweh, R. (1997): Professions in Modern Society, in: International Review of Sociology, 7, 95-102
- Standard Evaluation Protocol 2009-2015: Protocol for Research Assessment in the Netherlands. <http://www.knaw.nl/Pages/DEF/28/499.bGFuZz1FTkc.html> (Download 31.05.2011)
- Torka, M. (2011): Institutioneller gleich handlungspraktischer Wandel? Das Beispiel von Begutachtungspraktiken bei der Evaluation wissenschaftlicher Einrichtungen, in: iFQ-Tagungsband „Evaluation: New Balance of Power“ (im Erscheinen)
- Travis, G.D.L.; Collins, H.M. (1991): New Light on Old Boys: Cognitive and Institutional Particularism in the Peer Review System, in: Science, Technology, & Human Values, Vol. 16, 322-341
- Weingart, P. (2005): Das Ritual der Evaluierung und die Verführung der Zahlen, in: ders.: Die Wissenschaft der Öffentlichkeit. Essays zum Verhältnis von Wissenschaft, Medien, Öffentlichkeit, Weilerswist, 102-122
- Whitley, Richard (2007): Changing Governance of the Public Sciences, in: Whitley, R.; Gläser, J. (eds.): The Changing Governance of the Sciences, 3-27