

Beobachtungen wissenschaftlicher Selbstkontrolle

Qualität, Schwächen und die Zukunft des *Peer Review*-Verfahrens

Bei der Gründung der Zeitschrift für Internationale Beziehungen haben sich deren Herausgeber für ein doppelt anonymes Peer Review-Verfahren entschieden und sich fortan in ihren Publikationsentscheidungen von den Voten der GutachterInnen leiten lassen. Welche Validität kann das bei der ZIB praktizierte Verfahren für sich in Anspruch nehmen? Die Analyse von 452 Begutachtungen im Peer Review-Verfahren der ZIB ergibt, dass sich bezogen auf das Geschlecht und das Qualifikationsniveau der AutorInnen und der GutachterInnen keine signifikanten Verzerrungen in der Bewertung von Manuskripten zeigen. Bei der Qualität der Gutachten zeigen sich allerdings erhebliche Unterschiede, die eine starke Korrelation mit dem Qualifikationsniveau der GutachterInnen aufweisen. Vor diesem Hintergrund werden Qualitäts-Standards für Gutachten und einige Vorschläge für ein verbessertes Peer Review-Verfahren entwickelt, das die Leistungen der GutachterInnen höher honoriert. Neben ihrer Hilfestellung für Herausgeber-Entscheidungen sind Gutachten aus Peer Review-Verfahren vor allem wichtige Teile innerdisziplinärer Kommunikation. Sie gilt es, durch eine Qualitätssteigerung der Gutachten zu verbessern.

Kann es sein, dass die Mitglieder ein und derselben wissenschaftlichen Teildisziplin bezüglich der Frage, ob ein bestimmtes Manuskript in einer angesehenen Fachzeitschrift veröffentlicht werden soll, zu fundamental gegensätzlichen Einschätzungen gelangen? Bei den 175 Manuskripten, welche die Redaktion der *Zeitschrift für Internationale Beziehungen* (ZIB) in den ersten acht Jahren in die externe Begutachtung verschickt hat, kam dies in 15 Fällen vor, also immerhin bei fast neun Prozent der Begutachtungsverfahren: Ein Gutachten empfahl, das jeweilige Manuskript wie vorgelegt zu veröffentlichen, ein anderes, es ohne weitere Möglichkeiten der Überarbeitung sofort abzulehnen. Bei den restlichen 91 Prozent der extern begutachteten Manuskripte herrschte gerade mal bei der Hälfte der Fälle prinzipielle Einigkeit, dass das Manuskript – entweder ohne oder mit Überarbeitungen – veröffentlicht werden soll bzw. in der vorgelegten Fassung für eine Publikation noch nicht oder gar nicht in Frage kommt. Das heißt, dass sich die GutachterInnen der ZIB bei mehr als der Hälfte der extern begutachteten Manuskripte uneinig darüber waren, ob es sich jeweils um ein publikationsfähiges Manuskript handelt oder nicht.¹ Anlass

1 Der *intraclass-correlation coefficient* bezogen auf übereinstimmende Gutachtenden-Voten entsprechend der vier verwendeten Beurteilungskategorien beträgt +0,14; er wurde hier berechnet auf der Grundlage aller paarweisen Vergleiche von Gutachten bezogen auf dasselbe Manuskript ($n=379$; zu 102 Manuskripten liegen jeweils drei Gutachten vor). Zur Messung der Reliabilität von Gutachter-Urteilen und zur Berechnung solcher Koeffizienten und ihren Werten in anderen Disziplinen vgl. Marsh/Ball (1989); Hargens/Herting (1990); Daniel (1993: 3f) und Hirschauer (2004: 67f).

genug, das *Peer Review*-Verfahren selbst einer Analyse zu unterziehen und aus den bei der ZIB gemachten Erfahrungen einige Schlussfolgerungen abzuleiten.²

In den seit vielen Jahren geführten Diskussionen über Vorzüge, Schwächen und Nachteile von Evaluationen wissenschaftlicher Forschung ist auch das bei wissenschaftlichen Zeitschriften praktizierte *Peer Review*-Verfahren vielfach in die Kritik geraten (vgl. etwa Kornhuber 1988; Daniel 1993; Weingart 2001: 290f; Fröhlich 2003; Hirschauer 2004). Die dabei herausgearbeiteten Schwächen nehme ich als Ausgangspunkt einer systematischen Überprüfung des bei der ZIB praktizierten *Review*-Verfahrens. Nach der Erläuterung von Funktionen, Problemen und Erfahrungen mit *Peer Review*-Verfahren (Abschnitt 1) werde ich anschließend die *Review*-Praxis bei der ZIB beschreiben und die Datengrundlage für die Analyse dieser Form wissenschaftlicher Selbstkontrolle darstellen (Abschnitt 2). Dem folgt eine erste Auswertung der Datensätze zu verschiedenen Aspekten eines *gender Bias* im *Peer Review*-Verfahren der ZIB (Abschnitt 3). Die Qualität von Gutachten steht im Zentrum des zweiten Teils dieses Beitrags (Abschnitt 4): Zunächst werden mögliche Erklärungen für deren große Varianz diskutiert, dann zentrale Qualitäts-Standards für Gutachten entwickelt und in Beziehung zu Merkmalen der Gutachtenden gesetzt. Abschließend werden noch einige Vorschläge skizziert (Abschnitt 5), wie sich das *Peer Review*-Verfahren über die Herbeiführung von Publikationsentscheidungen bei Zeitschriftenartikeln hinaus im Sinne der Verbesserung wissenschaftlicher Kommunikationsprozesse weiterentwickeln ließe.

1. Die Bedeutung von Peer Review-Verfahren für die Wissenschaft

Die meisten AutorInnen, die ein Manuskript bei einer wissenschaftlichen Zeitschrift einreichen, sind davon überzeugt, dass es dem Fortgang wissenschaftlicher Kommunikation diente, wenn ihr Beitrag veröffentlicht würde.³ Es sprechen viele Gründe dafür, diesem Ansinnen nicht einfach zu folgen, sondern nur ausgewählte Manuskripte zu publizieren. Doch die Auswahl gestaltet sich schwieriger als allgemein vermutet wird. Zwar unterscheidet sich die Wissenschaft von anderen gesellschaftlichen Teilsystemen durch ihre Leitdifferenz zwischen »wahr« und »unwahr« (Luhmann 1994: 273), was der wissenschaftlichen Selbstkontrolle außerordentlich dienlich sein müsste, stehen doch selten solch klare Entscheidungskriterien bereit wie in diesem Fall. Doch so sehr dieser Bezugspunkt bei der Begutachtung von Zeitschriften-Manuskripten auch mitschwingen mag, folgen die Publikationsentscheidungen bei Zeitschriften doch eher Nützlichkeitserwägungen angesichts von Ressourcen-

-
- 2 Für hilfreiche Anmerkungen zu einer früheren Fassung dieses Beitrags danke ich Nicole Deitelhoff, Gabriele Matern, den mir anonym gebliebenen GutachterInnen der ZIB sowie Monika Pavetic für ihre Unterstützung bei der Datenedition, der Berechnung statistischer Kennwerte und der Darstellung der Ergebnisse.
 - 3 Vgl. aber Hirschauer: »Publizität« durch wissenschaftliche Aufsätze ist eine Autorenfiktion, die für die Mobilisierung der narzistischen Brennstoffe wissenschaftlicher Arbeit nützlich ist, aber nicht zur Modellierung wissenschaftlicher Kommunikation taugt« (Hirschauer 2004: 17).

knappeit: Die Aufmerksamkeit der LeserInnen, materielle Aufwendungen für gedruckte Publikationen sowie wissenschaftliches Ansehen und Renommee stehen nicht unbegrenzt, sondern nur in eingeschränktem Umfang zur Verfügung. Wer daran teilhaben will, muss sich den Standards anpassen, die sich für hoch angesehene Publikationsorte herausgebildet haben.⁴ Für Zeitschriften heißt dies: doppelt anonymes *Peer Review*-Verfahren!

»Peer review« steht für die Begutachtung und Bewertung von Publikationen und Forschungsanträgen, das heißt wissenschaftlicher Wissensbehauptungen durch die dazu allein kompetenten Kollegen (»peers«) (Weingart 2001: 284f; vgl. auch POST 2002). Es ist keine andere Instanz vorstellbar, die angesichts der immer weiter wachsenden Spezialisierung des Wissens dazu in der Lage sein könnte. Durch diese interne Prüfung sichert sich die Wissenschaft aus der Perspektive der Gesellschaft ihren Status als Produzentin verlässlichen Wissens. Und auch WissenschaftlerInnen wollen bzw. müssen sich in besonderer Weise auf solcherart gesichertes Wissen verlassen, sind sie doch niemals in der Lage, jenes Wissen zunächst selbst zu überprüfen, auf das sich ihre eigenen Untersuchungen stützen. Zugleich wollen sie die eigenen Einsichten dieser Form der Prüfung unterwerfen, um ihnen und sich selbst wissenschaftliche Anerkennung zu verschaffen.

»Das ›Peer review‹-System hat die Funktion, Vertrauen zu erzeugen, und zwar nach ›innen‹ das Vertrauen in die Verlässlichkeit und Wechselseitigkeit der wissenschaftlichen Kommunikation zur Sicherung ihrer Offenheit; und nach ›außen‹, gegenüber der Öffentlichkeit, Vertrauen in die Verlässlichkeit des produzierten Wissens, um die Ressourcen für die Forschung zu legitimieren« (Weingart 2001: 287).⁵

Das *Peer Review*-Verfahren bei wissenschaftlichen Zeitschriften zur Beurteilung von Manuskripten »wird gemeinhin als Kernstück wissenschaftlicher Kommunikation betrachtet, das ihren ›organisierten Skeptizismus‹ institutionalisiert und gute von schlechter Forschung unterscheidet« (Hirschauer 2004: 62). Bei Publikationsentscheidungen soll das *Peer Review*-Verfahren zudem dazu dienen, einer Zeitschrift ein bestimmtes Profil und Ansehen zu verschaffen, sowohl was die inhaltliche Ausrichtung als auch die Qualität der veröffentlichten Beiträge betrifft. Auf diesem Wege will eine Zeitschrift die ihr entsprechende Aufmerksamkeit einer *scientific community* gewinnen. Organisiert sie ihre Hefte allein über die Festlegung von Themen, sind die Entscheidungen über Aufnahme oder Zurückweisung eines Manuskripts leichter transparent zu machen, als wenn sie, wie die *Zeitschrift für Internationale Beziehungen*, vornehmlich »hohe Qualitätsstandards« (ZIB-Editorial, Hellmann/Müller 2003a: 5) zur Grundlage ihrer Publikationsentscheidungen machen möchte. Dabei stellt sich die Frage, wer am besten die Qualität eines Manuskripts bewerten kann im Hinblick auf seine Veröffentlichung in einer Zeitschrift, die den Anspruch verfolgt, »auch international als die führende wissenschaftliche Zeitschrift der Internationalen Beziehungen in deutscher Sprache anerkannt« (Hellmann/Müller 2003a: 3) zu werden.

4 Vgl. Plümper (2003). Zur Diskussion über unterschiedliche Maßstäbe für die Qualität und Reputation wissenschaftlicher Zeitschriften vgl. Norris/Crewe (1993).

5 Zu den Problemen und Gefahren dieser doppelten Zielsetzung von Forschungsevaluation vgl. Röbbecke/Simon (2001: 33-43).

Die HerausgeberInnen einer Zeitschrift gelten für diese Qualitätskontrolle als nur bedingt geeignet, stehen sie doch in der Gefahr, zu sehr ihrer individuellen Perspektive zu folgen und dabei zugleich den sozialen Dimensionen ihrer Entscheidungen zu großes Gewicht beizumessen: Wer sollte bereit sein, die möglichen Konsequenzen auf sich zu nehmen, wenn es etwa darum geht, das von einem DFG-Gutachter eingereichte Manuskript abzulehnen, weil es aus Sicht der HerausgeberInnen kein guter Text ist? Wer will für sich in Anspruch nehmen, das gesamte Themen-, Methoden- und Theoriefeld einer Disziplin oder auch nur einer Teildisziplin so gut zu überblicken, dass sie oder er in der Lage wäre, Qualität und Relevanz aller möglichen Manuskripte – zumindest in der Selbstwahrnehmung – sicher einzuschätzen? Diese Schwierigkeiten scheinen durch das doppelt anonyme Peer Review-Verfahren zunächst ausgeräumt: Indem die Gutachtenden nicht wissen, von wem ein Manuskript verfasst wurde, entsteht nicht jener bekannte Bias zugunsten etablierter, mächtiger, männlicher Autoren,⁶ sondern eine sich allein am Text und dessen Lesarten orientierende Beurteilung – wenn die Anonymisierung eines Manuskripts erfolgreich war. Indem zugleich die Gutachtenden für die AutorInnen der begutachteten Manuskripte anonym bleiben – wenn sie nicht aus irgendeinem Grund zur Selbstbeziehtigung neigen –, wird gewährleistet, dass Gutachtende bei negativen Beurteilungen keine individuellen Konsequenzen von möglicherweise mächtigen AutorInnen zu befürchten haben. Die doppelte Anonymität eines *Peer Review*-Verfahrens – AutorInnen bleiben für die Gutachtenden und diese auch für die AutorInnen anonym – kann folglich jene Verzerrungen der Begutachtung verhindern, deren Ursache in den sozialen Beziehungen innerhalb einer wissenschaftlichen Disziplin liegen.

Zugleich lässt sich durch Einbeziehung externer Expertise die Wissensgrundlage für die zu treffende Entscheidung verbreitern. Das *Peer Review*-Verfahren nutzt die Kompetenzen derjenigen, die im selben wissenschaftlichen Feld arbeiten wie die AutorInnen der zu begutachtenden Manuskripte. Sie werden für die Erstellung von Gutachten herangezogen, weil sie nicht nur die Qualität der Bearbeitung einer Fragestellung am genauesten beurteilen, sondern auch die Relevanz der Forschungsergebnisse und den Nutzen ihrer Publikation am besten abschätzen können. Doch mit dieser inhaltlichen Nähe zwischen AutorInnen und Gutachtenden gerät auch schon die Unabhängigkeit des Urteils in erhebliche Gefahr. Wer im selben Themenfeld arbeitet, der kennt sich – zumindest aus der Literatur. Von den Gutachtenden »muss man fürchten, dass sie dem Objekt ihrer Bewertung entweder als allzu gute Kollegen oder aber als das gerade Gegenteil davon herhaft verbunden sind« (Neidhardt 2000: 28). Das schafft zudem nicht selten ein Anonymisierungsproblem bei den Manuskripten, denn was zitiert wird und was nicht, kann sehr verräterisch sein. Häufig provoziert die formale Anonymität von zu begutachtenden Manuskripten einen detektivischen Impuls, zumindest starke Vermutungen über die AutorInnenschaft des Manuskripts anzustellen. Für die Gutachtenden und ihr Urteil aber ist es

6 Daniel (1993: 73); Felt et al. (1995: 107f); Wennerås/Wold (1997); Krais (2000a); Europäische Kommission (2001); Hirschauer (2004: 65f).

nicht bedeutungslos, ob sie in ihrer Beziehung zu den möglichen AutorInnen von Kollegialität oder Konkurrenz ausgehen sollen (Hirschauer 2004: 71).

»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« (Neidhardt 2000: 28f).⁷

Zudem lassen sich durch die Anonymisierung von Manuskripten und Gutachten all jene Beurteilungsfaktoren nicht aus dem *Review*-Verfahren eliminieren, die aus den individuellen Motiven und Perspektiven der Gutachtenden in ihren Bezügen zu einer wissenschaftlichen *community* liegen: Sollen sie einem Manuskript, das dem selbst vertretenen Ansatz kritisch gegenübersteht, zur Publikation verhelfen, um die wissenschaftliche Debatte darüber zu fördern, oder besser seine Ablehnung empfehlen, um den für die eigenen Arbeiten förderlichen Mainstream nicht zu schwächen? Welche Überlegungen werden angestellt, wenn es gilt, ein Manuskript zu begutachten, das den eigenen Arbeiten eng verwandt ist? Soll man es mit kritischen Überarbeitungshinweisen überhäufen, um die eigenen Kompetenzen zu verdeutlichen, und damit die Publikation möglicherweise verhindern oder zumindest verzögern, damit Raum bleibt für die Veröffentlichung der eigenen Arbeiten? Oder soll man seine Publikation befürworten, um den entsprechenden Strang der Forschung zu stärken, wovon man auch selbst profitieren könnte? Wie sind Manuskripte zu begutachten, die neue, innovative Ansätze in die Diskussion einbringen wollen? Sie werfen zwangsläufig mehr ungelöste und damit auch unbeantwortete Fragen auf als die *x-te* Fallstudie mit einem breit etablierten Ansatz und liefern damit viel einfachere Möglichkeiten zur Begründung einer Ablehnung als jene methodisch korrekte, aber langweilige, weil innovationsfreie Fallstudie. Befürwortet man die Veröffentlichung von Beiträgen mit neuen Ansätzen, weil man wissenschaftliche Innovationen fördern will, ist damit das Risiko verbunden, sich das Image des unkritischen Gutachtenden zuzulegen.⁸ Plädiert man für »Ablehnen«, lässt sich dies nicht nur leichter begrün-

7 Bei einer deutschsprachigen Zeitschrift verschärft sich dieses Problem noch zusätzlich, da aus Gründen der Sprachkompetenz nur in sehr begrenztem Umfang auf ausländische Gutachtende zurückgegriffen werden kann.

8 Zudem ist zu befürchten, dass auch die HerausgeberInnen eher zurückhaltend mit der Veröffentlichung innovativer Forschungsergebnisse umgehen: »Die Herausgeber renommierter Zeitschriften können es sich nicht erlauben, Aufsätze zu veröffentlichen, die das Risiko in sich tragen, nicht häufig zitiert zu werden, denn das senkt den *impact factor* der Zeitschrift. [...] Das Streben, viele Aufsätze in Zeitschriften mit einem hohen *impact factor* zu veröffentlichen, fördert einen wissenschaftlichen Konservatismus. Wissenschaftler betreiben keine innovative Forschung mehr, sondern konzentrieren sich auf Weiterentwicklungen von Bewährtem. [...] So fördert die Evaluation auf der Basis von Zitationen die Standardisierung der Wissenschaft und beschleunigt, vor allem in den Sozial- und Wirtschaftswissenschaften, den Prozess der Abkopplung des Wissenschaftssystems von den Feldern der praktischen Umsetzung. Nicht mehr die Aufhellung ungeklärter Phänomene und erst recht nicht die Lösung praktischer Probleme ist das zentrale

den, sondern damit auch dem Risiko vorbeugen, dass die Aufmerksamkeit der *community* durch innovative Ansätze zu sehr abgelenkt wird von den Debatten, an denen man selbst beteiligt ist.

Diese und ähnliche soziale Dimensionen des doppelt anonymen *Peer Review*-Verfahrens spielen für die Urteile der Gutachtenden und damit für die Entscheidung über die (Nicht-)Veröffentlichung von Manuskripten eine nicht unbedeutende, aber kaum systematisch erfassbare Rolle. Selbstkritische Gutachtende werden sich im Laufe ihres Beurteilungsprozesses – vom ersten Vorurteil bei Kenntnis des Titels über die Reflexion ihrer Leseindrücke und die Überprüfung ihres Urteils bis hin zum Ausfüllen des *Review*-Bogens – solche Fragen, wie sie oben angedeutet wurden, selbst stellen und möglicherweise erste Bewertungen revidieren. Doch eine Unabhängigkeit von solchen Faktoren wird es bei der Begutachtung wissenschaftlicher Veröffentlichungen nicht geben: Wissenschaft ist und bleibt eine soziale Veranstaltung (Mittelstraß 1989),⁹ auch und gerade bei ihrer Evaluation und Selbstkontrolle (Hirschauer 2004: Kap. 3).

Für Differenzen in der Begutachtung fallen insbesondere all jenen, die schon begutachtet haben oder deren Manuskripte schon begutachtet wurden, noch weitere Erklärungen ein: Der Wissens- und damit Beurteilungs-Hintergrund ist verschieden, sei es bei theoretischen Prämissen, methodischer Herangehensweise, metatheoretischer Position oder einfach bei der Kenntnis für wichtig erachteter wissenschaftlicher Debatten. Genau diese Verschiedenheit der Expertise wollen die HerausgeberInnen jedoch durch den gezielten Rückgriff auf unterschiedliche Gutachtende nutzen. So wurde bei der ZIB zumeist versucht, mit den zwei oder drei Gutachtenden einerseits jemand Kompetentes für die behandelte Empirie zu finden und andererseits besondere Expertise bezüglich der gewählten theoretischen Herangehensweise in das *Peer Review*-Verfahren einzubeziehen. Wenn beispielsweise in einem Manuskript internationale Vereinbarungen zum Umweltschutz mit einem Regime-Ansatz analysiert werden, bietet es sich an, zum einen jemand um ein Gutachten zu bitten, der internationale Umweltpolitik als bevorzugtes Themenfeld angegeben

Anliegen, im Vordergrund steht vielmehr die Herstellung von Aufsätzen, deren Chancen, in Zeitschriften mit hohem *impact factor* veröffentlicht zu werden, sehr gut sind. Forschungseinrichtungen degenerieren zu Fabriken zur Produktion hoch standardisierter Aufsätze. [...] Wie alle Evaluationskriterien schafft auch der *citation index* die Wirklichkeit, die zu messen er vorgibt. Er ändert das Verhalten der Evaluierten« (Kieser 2003: 31, Hervorh. dort). Zur unerwünschten Anpassung der Evaluierten an die vermeintlichen Maßstäbe der Beurteilung vgl. auch Röbecke/Simon (2001) und Frey (2003).

9 Vgl. auch Krais (2000c: 34–42). »Kurz zusammengefasst geht die ›neue‹ Wissenschaftsforschung davon aus, dass wissenschaftliche Erkenntnis nicht einfach durch die Realität oder ›Natur‹ des Gegenstandes bestimmt wird, die dann durch immanentes Fortschreiten der Wissenschaft immer besser erfasst werden kann. Vielmehr ist wissenschaftliche Erkenntnis ein sozialer Prozess, in den nicht nur theoretische Vorannahmen, sondern auch der jeweilige soziale Kontext, Gewohnheiten und Verfasstheiten der *scientific community* und historisch gebundene Denkstile, wie Fleck (Fleck 1980) sie nennt, eingehen. Wissenschaftliche Wahrheit ist nicht zu begreifen als ›Entdeckung‹ von etwas, das unabhängig von der Wissenschaft, als Gegenstand für sich – in der ›Natur‹ oder in der ›Realität‹ – existiert, sondern viel eher als Konstruktionsprozess im Rahmen wissenschaftlicher Arbeit« (Krais 2000c: 32, Hervorh. dort).

hat,¹⁰ und ein zweites Gutachten von einer Regime-Expertin einzuholen. So können sich verschiedene Gutachten gegenseitig ergänzen und zusammengenommen eine möglichst umfassende Beurteilung des Manuskripts ermöglichen. Dass dabei die Bewertungen der Gutachten nicht immer einheitlich ausfallen, wird kaum jemand verwundern. Vielmehr handelt es sich um einen erwünschten Effekt der Auswahl der Gutachtenden, denn würden die verschiedenen Gutachten zu einem Manuskript in ihren Bewertungen und inhaltlichen Anregungen in aller Regel weitgehend übereinstimmen, könnte man leicht darauf verzichten, mehrere Gutachten einzuholen.

Zwar lässt sich unter Hinweis auf die Motive, die der GutachterInnen-Auswahl zugrunde liegen, argumentieren, dass Differenzen in der Begutachtung nicht als Schwäche des *Peer Review*-Verfahrens, sondern als sein erwünschter Effekt zu interpretieren sind (vgl. Daniel 1993: 6; Hirschauer 2004: 76). Doch andere Bewertungsfaktoren treten hinzu, bei denen die Gutachtenden erheblich differieren können und damit uneindeutige *Review*-Ergebnisse produzieren. So kann etwa das Interesse an der bearbeiteten Fragestellung sehr unterschiedlich sein, oder es können die grundsätzlichen Maßstäbe für publikationswürdige Manuskripte differieren. Für die einen hängt ein theoretisches Argument ohne empirische Anwendung oder zumindest Veranschaulichung in der Luft, für die anderen sind empirische Fallstudien ohne theoretische oder metatheoretische Reflexion kein die Forschung weiterbringender Beitrag, um nur einen Aspekt der Maßstäbe-Debatte zu erwähnen. Zudem können Stil und Argumentationsweise eines Manuskripts der einen Gutachterin sympathisch sein, bei einem anderen aber leicht aggressive Reflexe auslösen, was alles nicht ohne Einfluss auf das Ergebnis der Begutachtung sein wird. Am Ende des Begutachtungsprozesses schließlich wird sich kaum jemand gegen das eigene, intuitiv-emotional gefallte Urteil entscheiden, das sich während des Lesens im Hintergrund der inhaltlichen Aufmerksamkeiten gebildet hat. Für dieses Urteil lassen sich anschließend allemal rationale Gründe auffinden und anführen, denen dann nur noch das entsprechende Gewicht beigelegt werden muss.¹¹

Ist angesichts all dieser – gemessen am Ideal einer Qualitätsmessung von Manuskripten – sachfremden Einflüsse auf die Begutachtungsergebnisse das praktizierte *Peer Review*-Verfahren ungeeignet für die Publikationsentscheidungen bei wissenschaftlichen Zeitschriften? Dient das *Peer Review*-Verfahren wie viele andere Evaluierungen in der Wissenschaft mehr der Legitimation (vgl. Mittelstraß 2000: 25) als der Qualitätskontrolle? Ein systematisches Begutachtungsverfahren, in dem verbindlich über die (Nicht-)Veröffentlichung eines wissenschaftlichen Manuskripts entschieden wird, macht nur Sinn, wenn es zumindest teilweise einheitliche Maßstäbe

-
- 10 Die Mitglieder des ZIB-*Review-Panels* wurden, in aller Regel bevor ihnen ein Manuskript zur Begutachtung zugeschickt wurde, nach ihrer prinzipiellen Bereitschaft zur Mitwirkung im *Peer Review*-Verfahren der ZIB gefragt und dabei um die Angabe bevorzugter Themenfelder gebeten.
- 11 Vgl. Hirschauer: »Die Bewerkstelligung des ›Urteils‹ besteht eben auch in der rhetorischen Herstellung von Konsistenz durch die Legitimation einer einmal getroffenen (Vor)entscheidung. Dabei sind die an der Textoberfläche erscheinenden Gründe post hoc Rationalisierungen, die mit den im Entscheidungsprozess wirksamen Gründen und Motiven u. U. nicht viel zu tun haben« (Hirschauer 2004: 71).

gibt, anhand derer die Veröffentlichungswürdigkeit eines Manuskripts beurteilt wird. Ein *Peer Review*-Verfahren bei einer Zeitschrift geht davon aus, dass es eine ausreichend große und zugleich grob abgrenzbare disziplinäre Fach-community gibt, die in dieser Frage einigermaßen übereinstimmt.¹² Inwieweit dies auf das *Review-Panel* der *Zeitschrift für Internationale Beziehungen* zutrifft, soll in den Abschnitten 3 und 4 anhand einiger quantitativer Analysen überprüft werden. Zunächst aber ist die dafür verwendete Datengrundlage und das *Peer Review*-Verfahren, wie es bei der ZIB in den ersten acht Jahren ihres Bestehens praktiziert wurde, vorzustellen.

2. Das *Peer Review*-Verfahren der *Zeitschrift für Internationale Beziehungen*

Für die ersten acht Jahrgänge der ZIB wurden von Sommer 1993 bis Sommer 2001 insgesamt 175 Manuskripte eingereicht.¹³ Ihre Zahl stieg in den ersten Jahren und stabilisierte sich dann bei etwa 20 Einreichungen in zwölf Monaten. Davon wurden 145 (83%) extern begutachtet und 80 (45%) publiziert. Die Ablehnungsquote der ZIB pendelte über die Jahre zwischen 50% und 60%, bei Aufsatz-Manuskripten lag sie fast immer über 60% und hat sich auch in den letzten Jahren nicht spürbar verändert.

Etwa ein Sechstel (17%) der eingereichten Manuskripte wurde nach einer ersten Evaluation durch den ZIB-Redakteur und den geschäftsführenden Herausgeber ohne externe Begutachtung abgelehnt, weil sie aus unterschiedlichsten Gründen dem Profil der Zeitschrift nicht entsprachen und mit Sicherheit im *Review*-Verfahren abgelehnt worden wären.¹⁴ Diese Vorauswahl eingereichter Manuskripte dient vor allem der Entlastung des *Review-Panels*, das nicht mit der Begutachtung chancenloser Manuskripte behelligt werden sollte. Trotz dieser Vorselektion votierten die Gutachter noch bei 25 Manuskripten (14% der extern begutachteten Manuskripte) mehrheitlich für Ablehnung.¹⁵

-
- 12 »Es spricht viel dafür, dass die Idee des ›richtigen‹ Gutachtens eine Illusion darstellt. Wenn dem so ist, bliebe freilich zu klären, wie man sich als Gutachter auch unter verrückten Bedingungen noch professionell verhalten kann. Vermutlich reduziert es sich auf zwei Punkte: auf einen Schuss methodisch eingesetzter Naivität und auf die Bereitschaft, etwas für die eigene Irritierbarkeit zu tun. Gutachter müssen nämlich, um ihren Job überhaupt tun zu können, auf der Idee der rationalen Entscheidbarkeit von Gutachtenfragen beharren. Trotz aller Belege für deren faktische Unmöglichkeit bleibt dies eine ebenso unentbehrliche wie hilfreiche Fiktion. Sie verdient es, mit Bedacht – sozusagen im ›Modus des Als-ob‹ – kultiviert und dann gleichermaßen gegen Relativisten wie gegen jene verteidigt zu werden, die sie bloß wörtlich nehmen« (Wolff 2000: 39).
- 13 In diese Zählung wurden auch Forums-Manuskripte und Literaturberichte einbezogen, wenn sie – wie seit 1997 geschehen – im *Peer Review*-Verfahren begutachtet wurden. Zur aktuellen, nur marginal veränderten Form des *Peer Review*-Verfahrens der ZIB vgl. Hellmann/Müller (2002: 4f).
- 14 Ohne *Review*-Verfahren wurden Manuskripte dann abgelehnt, wenn geschäftsführender Herausgeber und ZIB-Redakteur in dieser Frage übereinstimmender Auffassung waren.
- 15 Dies entspricht dem in fast allen Studien zur Zuverlässigkeit von Urteilen in *Peer Review*-Verfahren beobachtbaren höheren Ablehnungskonsens in Relation zum Akzeptanzkonsens (vgl. Hirschauer 2004: 67f; Giles et al. 1989: 60). Da den Urteilen der Gutachter differierende Kriterien und Maßstäbe zugrunde liegen, gibt es Übereinstimmung vor allem bei den Ausreißern und nach unten. »Urteile über wissenschaftliche Manuskripte verlangen die Gewichtung unzähliger Kriterien. Sie sind so komplex wie

Für die externe Begutachtung wurden in der Anfangsphase, in der die anonyme Begutachtung von Zeitschriften-Manuskripten in der deutschen Politikwissenschaft noch unüblich war¹⁶ und das ZIB-Review-Panel aufgebaut werden musste, zwei, später fast immer drei im entsprechenden Theorie- oder Themenfeld ausgewiesene PolitologInnen, vornehmlich aus der Teildisziplin Internationale Beziehungen, um Gutachten gebeten. Thematisierten die Manuskripte auch Problemstellungen benachbarter Disziplinen, wurden z. B. auch SoziologInnen, HistorikerInnen, PhilosophInnen, PsychologInnen oder JuristInnen angefragt. Allen Gutachtenden wurde vorgegeben, sich in ihrer Beurteilung für eine von vier Kategorien¹⁷ zu entscheiden, ihr Urteil in einem ausführlichen Gutachten zu begründen und dabei auch konkrete Überarbeitungshinweise zu geben, wenn für eine Überarbeitung plädiert wird. Das externe Begutachtungsverfahren konnte zu fünf verschiedenen Ergebnissen führen, wie mit dem Manuskript weiter verfahren wird (siehe Abbildung 1).

- Direkte Publikationszusage an die AutorInnen, ggf. mit der Bitte um Anpassung des Beitrags an die Formatvorlage der ZIB (zwei Fälle);
- Publikationszusage mit der Bitte um Berücksichtigung inhaltlicher Anmerkungen aus den Gutachten zur Erstellung der Endfassung (46 Fälle);
- Konditionale Publikationszusage an die AutorInnen, bei der die Veröffentlichung von der Erfüllung bestimmter Überarbeitungsaufgaben abhängig gemacht wird (acht Fälle);
- Aufforderung zur intensiven Überarbeitung und Wiedereinreichung des Manuskripts bei der ZIB unter Berücksichtigung der Anmerkungen und Hinweise aus den Gutachten (32 Fälle);
- Ablehnung des Manuskripts unter Verweis auf die entsprechend übereinstimmenden Voten der externen Gutachtenden (19 Fälle).

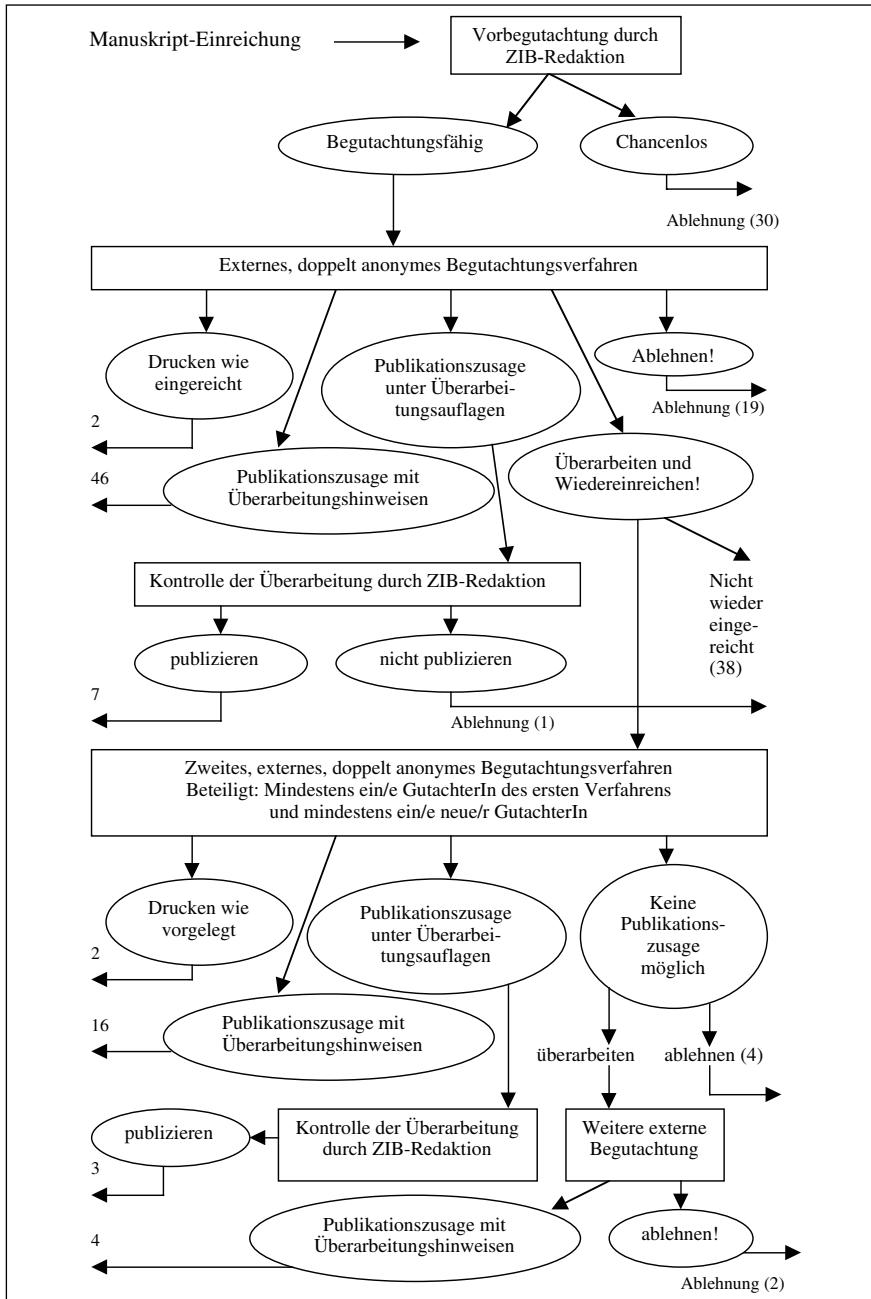
In allen Fällen wurden den AutorInnen alle Gutachten – in einer anonymisierten Form – zur Verfügung gestellt und in einem Begleitschreiben auch kommentiert. Letzteres war insbesondere dann erforderlich, wenn gutachterliche Überarbeitungshinweise in verschiedene Richtungen zielten, sich Gutachtende im Ton vergriffen hatten oder aus Redaktionssicht eine motivierende Geste zur Umsetzung der Überarbeitungshinweise angebracht schien.

Urteile über Attraktivität, über den Geschmack von Wein oder das Aroma von Parfums. Zwar kann man gravierende Mängel recht gut spezifizieren (daher die höhere Übereinstimmung bei Ablehnungen), aber man kann wie bei Bewerbern im Einstellungsgespräch so auch bei Manuskripten ganz unterschiedliche Qualifikationsprofile beschreiben, je nachdem welchen Einsatzzweck man im Auge hat« (Hirschauer 2004: 68f.).

16 So etwa bei der *Politischen Vierteljahrsschrift* (PVS): »Eine erhebliche Veränderung der Arbeitsweise der [PVS-]Redaktion wird sich ab 1994 durch die Umstellung des Begutachterverfahrens [sic!] auf ein ›external referee-‹System ergeben, das vom Vorstand und Beirat der Deutschen Vereinigung für Politikwissenschaft angeregt wurde« (Héritier 1994: 1). »Die Auswahl der Beiträge [erfolgt] auf der Grundlage eines doppelt-anonymen Begutachtungsverfahrens« (Seibel 1998: 1).

17 »Wie vorliegend ohne Änderungen zur Veröffentlichung annehmen«, »mit Veröffentlichungszusage zur Überarbeitung zurückgeben«, »ohne Veröffentlichungszusage zur Überarbeitung zurückgeben«, »ablehnen«.

Abbildung 1: Begutachtungsprozesse im Rahmen des Peer Review-Verfahrens



Dieses erste externe Begutachtungsverfahren (siehe Abbildung 1) brachte für weitere 67 Manuskripte – nach 30 vor dem externen Review-Verfahren abgelehnten – ein eindeutiges Ergebnis:¹⁸ Zwei wurden ohne inhaltliche Überarbeitung und 46 nach der Berücksichtigung gutachterlicher Anmerkungen publiziert (27% von 175 eingereichten Manuskripten). 19 Manuskripte wurden abgelehnt.¹⁹ In acht Fällen wurde eine konditionale Publikationszusage erteilt, die nur bei einem Manuskript nicht zum erhofften Erfolg der Veröffentlichung führte. Bei den restlichen 70 Manuskripten wurde den AutorInnen die Überarbeitung und Wiedereinreichung empfohlen, was in weniger als der Hälfte der Fälle (32, entspricht 46%) gemacht wurde. Wenn diese Überarbeitung entsprechend den gutachterlichen Hinweisen vorgenommen wurde, führte sie in 78% der Fälle zum Publikationserfolg, allerdings teilweise erst nach weiteren intensiven Überarbeitungen aufgrund einer konditionalen Publikationszusage (drei Fälle) oder einer nochmaligen externen Begutachtung mit erneuten Anmerkungen aus dem *Review*-Verfahren, die vor der Veröffentlichung zu berücksichtigen waren (vier Fälle). Das zweite externe *Review*-Verfahren zu einem Manuskript führte in vier Fällen zur Ablehnung, bei sieben Manuskripten wurde eine erneute Überarbeitung und Wiedereinreichung empfohlen,²⁰ was am Ende zu vier Publikationszusagen und zwei Ablehnungen führte; eines dieser Manuskripte wurde nicht wieder eingereicht.

Entsprechend des hier beschriebenen *Peer Review*-Verfahrens bei der ZIB wurden innerhalb von gut acht Jahren bis Ende 2001 insgesamt 452 externe Gutachten erstellt, mit denen 175 eingereichte oder schon überarbeitete Manuskripte beurteilt wurden. 145 Review-Verfahren bezogen sich auf die erste Einreichung und 26 Verfahren auf schon begutachtete und nach einer Überarbeitung wiedereingereichte Manuskripte. In vier Fällen war ein zum zweiten Mal überarbeitetes Manuskript auf seine Publikationsfähigkeit hin zu beurteilen.²¹ Diese 175 *Review*-Verfahren, in

-
- 18 Lagen zwei Gutachten mit gegensätzlichen Voten vor, wurde in aller Regel ein weiteres Gutachten eingeholt und in der herausgeberischen Entscheidung dann der überwiegenden Tendenz gefolgt. Bei den in der Einleitung erwähnten 15 Manuskripten, zu denen fundamental gegensätzliche Urteile der Gutachtenden vorlagen, handelte es sich mehrheitlich um Aufsatz-Manuskripte, 40% waren Literaturberichte oder Forumsbeiträge. In 13 der 15 Fälle waren nur PolitologInnen an der Begutachtung beteiligt, in zehn der 15 Fälle stammten alle Gutachtenden aus dem Bereich der Internationalen Beziehungen.
 - 19 Bei sechs weiteren Manuskripten plädierten die Gutachtenden mehrheitlich für Ablehnung, Herausgeber und Redaktion folgten jedoch der Minderheiten-Meinung und empfahlen den AutorInnen die intensive Überarbeitung und Wiedereinreichung bei der ZIB.
 - 20 Eine solche Entscheidung stellt sehr hohe Ansprüche und grenzt häufig an eine Zumutung, sowohl für AutorInnen wie für die mehrfach einbezogenen Gutachtenden. Die Verbesserung eines Manuskripts durch *Peer Review*-Verfahren muss irgendwann abgebrochen werden, auch um die Gutachtenden nicht zu ungenannten KoautorInnen werden zu lassen (vgl. Hirschauer 2004: 67). Diesem Problem wird bei der ZIB inzwischen dadurch begegnet, dass den Gutachtenden bei der Beurteilung eines wiedereingereichten Manuskripts die Kategorie »überarbeiten und wiedereinreichen!« nicht mehr zur Verfügung steht.
 - 21 Die Analyse bezieht sich auf alle bis Ende 2001 erstellten Gutachten. Zu einigen Manuskripten war damit allerdings das Verfahren noch nicht abgeschlossen. Für die Darstellung in Abbildung 1 wurde das jeweilige Endergebnis der Verfahren berücksichtigt, weshalb die Summe der dort verzeichneten *Peer Review*-Verfahren leicht höher liegt.

denen derselbe Text von verschiedenen – mindestens zwei – Gutachtenden im Hinblick auf seine Veröffentlichung beurteilt wurde, eröffnen eine erste Möglichkeit, das doppelt anonyme *Peer Review*-Verfahren der ZIB zu evaluieren. Um die Anonymität des Verfahrens zu wahren, werden im Hinblick auf mögliche Schwächen und Schieflagen des Verfahrens, wie sie oben beschrieben wurden, quantitative Analysen auf der Grundlage kumulierter bzw. aggregierter Daten vorgenommen.²² Sie kommen zu ähnlichen Ergebnissen wie viele vorhergehende Studien zu *Peer Review*-Verfahren (vgl. dazu Hirschauer 2004), gehen aber bei der Analyse der Gutachten noch einen Schritt darüber hinaus und bringen damit Zusammenhänge zwischen der Qualität von Gutachten und Merkmalen der Gutachtenden zum Vorschein (siehe unten Abschnitt 5), die Anlass geben, über Verbesserungen des *Peer Review*-Verfahrens nachzudenken.

3. Gender-Bias?

Die Diskriminierung von Wissenschaftlerinnen durch männlich geprägte Funktionsweisen von Wissenschaft ist in der Wissenschaftsforschung vielfach herausgearbeitet worden (vgl. etwa Krais 2000a; Europäische Kommission 2001). Dabei ist von »agonalem Verhalten« (Krais 2000b: 44) die Rede, bei dem es darauf ankommt, »untereinander Rangordnungen herzustellen; Leistungen nicht ausschließlich um ihrer selbst willen zu erbringen, sondern um einen Positionsgewinn zu erzielen; Gegner zu haben, die sie herausfordern und von denen sie sich ihrerseits herausfordert fühlen; sich gegen ›Mitspieler‹ durchsetzen zu wollen. Diese Aspekte gehören zum Kern der Funktionsweise von Wissenschaft« (Krais 2000b: 44).

Diese Konkurrenz zwischen Individuen stellt einen wesentlichen Beweggrund des Verhaltens in der Wissenschaft dar und ist zugleich nur als Ritual zu verstehen. Der wissenschaftliche »Wettkampf« wird nach bestimmten Regeln ausgetragen (Krais 2000b: 45), die sicherstellen, dass auch anschließende Kooperation immer möglich bleibt. In der Konkurrenz werden die eigenen Leistungen mit denen anderer verglichen und damit zugleich die »Gegner« bzw. Mitspieler anerkannt. »In dieser agonalen Form des Sich-Aufeinander-Beziehens erst bildet sich das gemeinsam betriebene ›Spiel‹ aus, entsteht und aktualisiert sich Wissenschaft als ›gesellschaftliche Veranstaltung‹« (Krais 2000b: 46). Und diese agonale Funktionsweise von Wissenschaft stellt eine mindestens zweifache Barriere für Wissenschaftlerinnen dar. Frauen werden nicht im selben Maße wie Männer in den wissenschaftlichen Wettbewerb einbe-

22 Um die zugesicherte Anonymität insbesondere der Gutachtenden weiterhin zu gewährleisten, kann der zwar anonymisierte, aber an verschiedenen Stellen auch problemlos entanonymisierbare Datensatz nicht öffentlich zugänglich gemacht werden. Er steht den Herausgebern der ZIB zur Verfügung und kann von diesen bei einem begründeten Interesse an der Reevaluierung der hier vorgenommenen Auswertungen zur Verfügung gestellt werden.

zogen,²³ womit ihnen Teile von Anerkennung und gleichberechtigte Integration in eine *scientific community* versagt bleiben (Krais 2000b: 46f). Zudem verweigern sie sich vielfach den Ritualen des Konkurrenzverhaltens in der Wissenschaft, das sie als substanzloses Spiel wahrnehmen und kritisieren. Wissenschaft erscheint ihnen als zu ernste Sache, als dass die Wettbewerbe nur ein Spiel sein sollten.

»Die Verweigerung agonalen Verhaltens in der ›arena of contest‹ von Seiten vieler WissenschaftlerInnen allerdings ist vor dem Hintergrund ihrer Nicht-Anerkennung im agonalen Spiel ihrer männlichen Kollegen nicht einfach das Verhalten von SpielverderberInnen: Der für Frauen bitter ernste Kampf um die Anerkennung als Wissenschaftlerin hindert sie daran, sich auf die agonalen, ritualisierten Kämpfe des wissenschaftlichen Feldes einzulassen, in denen die Anerkennung praktisch hergestellt wird« (Krais 2000b: 48).

Hat dies auch Auswirkungen auf ihre Publikationstätigkeit, auf die Bewertung der von Frauen verfassten Manuskripte und auf ihre Urteile, wenn sie im *Peer Review*-Verfahren Manuskripte zu begutachten haben? Der Bericht der ETAN-Expertinnenarbeitsgruppe »Frauen und Wissenschaft« fordert dazu auf, dass auch das *Peer Review*-System »auf mögliche geschlechtsspezifische Verzerrungen in der Konzeption und Durchführung überprüft werden muss« (Europäische Kommission 2001: 33). Viele Studien haben gezeigt, dass die Begutachtung wissenschaftlicher Leistungen nicht unabhängig vom Geschlecht vorgenommen wird. Vielmehr zeigt sich eine Überbewertung der Leistungen von Männern, während die Leistungen von Frauen unterbewertet werden (Wennerås/Wold 2000: 108; Europäische Kommission 2001: 33-45). Dieser Bias zeigt sich aber nur dann, wenn den GutachterInnen das Geschlecht der zu begutachtenden Person bzw. das AutorInnen-Geschlecht bekannt ist (vgl. Giles et al. 1989: 62; Wennerås/Wold 2000: 117).²⁴ Genau dies wird durch das doppelt anonyme *Peer Review*-Verfahren ausgeschlossen. Ist damit aber auch sichergestellt, dass Manuskripte von Frauen nicht aufgrund sprachlicher oder anderer Besonderheiten auffallen und daraufhin – bewusst oder unbewusst – schlechter bewertet werden als Manuskripte von Männern? Dies könnte möglicherweise einhergehen mit einer unterschiedlichen Bewertung der Manuskripte von Frauen, abhängig vom Geschlecht der Gutachtenden. Hypothetisch ließe sich annehmen, dass GutachterInnen die Manuskripte von Frauen besser bewerten als Manuskripte von Männern.

-
- 23 Vgl. Hagemann-White/Schultz: »Der Werdegang des männlichen Hochschullehrers kann als *berufliche Sozialisation* bestimmt und als Fortsetzung der typisch männlichen Sozialisation in Kindheit und Jugend gesehen werden: der Werdegang der Frau an der Hochschule wird hingegen als ein Prozess der *Akkulturation* zu bestimmen sein und steht vielfach im Widerspruch zu der geschlechtstypischen Sozialisation in Kindheit und Jugend« (Hagemann-White/Schultz 1986: 101, Hervorh. dort).
- 24 Der Anteil von Frauen unter den ZIB-AutorInnen bestätigt diesen Befund. Er ist für die Jahre 1994 bis 2001 etwa doppelt so hoch (14%) als beispielsweise bei der *Politischen Vierteljahresschrift* in den Jahrgängen vor Einführung des anonymen *Peer Review*-Verfahrens (1990-1995: 6,5%). Berechnet man den AutorInnen-Anteil für die darauf folgenden Jahrgänge (1996-2003) der PVS, beträgt er 15,6%. Der geschäftsführende Redakteur der PVS berichtet für den Jahrgang 1997 von einer »weit überdurchschnittlichen Annahmequote bei den Manuskripten weiblicher Autoren. Lediglich 16% aller von Männern eingesandten Manuskripte wurden letzten Endes abgedruckt, jedoch 57% aller von weiblichen Autoren eingesandten Manuskripte. Da die Auswahl der Beiträge auf der Grundlage eines doppelt-anonymen Begutachtungsverfahrens erfolgt, ist ein geschlechtspezifischer Bias bei der Manuskriptauswahl so gut wie ausgeschlossen« (Seibel 1998: 1).

Von den insgesamt 33 allein von Frauen verfassten Manuskripten, die ins ZIB-Review-Verfahren verschickt wurden, waren bei zwölf Manuskripten sowohl Gutachterinnen als auch Gutachter beteiligt. Sie stimmten in vielen Fällen bei ihren Voten überein, und die Gutachterinnen neigten insgesamt sogar zu leicht schlechteren Bewertungen als die Gutachter.²⁵ Betrachtet man die arithmetischen Mittelwerte der Voten von weiblichen und männlichen Gutachtenden bezüglich der von Frauen verfassten Manuskripte,²⁶ sind sie nahezu identisch (3,08; 3,00; Standardabweichung 0,79 bzw. 0,83). Auch weitere statistische Überprüfungen eines *gender*-Bias im ZIB-Review-Verfahren bestätigen den Befund der weitgehenden Geschlechter-Neutralität des doppelt anonymen *Peer Review*-Verfahrens. So kann auch die Hypothese, dass Manuskripte von Autorinnen insgesamt schlechter oder uneinheitlicher begutachtet werden, ebenfalls statistisch nicht bestätigt werden. Der Mittelwert für die Bewertung der Manuskripte von Autoren beträgt 2,7, jener für Manuskripte von Autorinnen 2,9 bei einer Standardabweichung von 0,72 (Autoren: 0,65).²⁷

Auch auf Seiten der Gutachtenden zeigen sich nur minimale Auffälligkeiten bei der Überprüfung eines *gender*-Bias. So kann etwa die These, Gutachterinnen beurteilten Manuskripte grundsätzlich besser als Gutachter, nicht bestätigt werden. Tendenziell ist eher das Gegenteil der Fall, aber die Differenz zwischen Gutachterinnen und Gutachtern hinsichtlich der Bewertung von Manuskripten ist sehr gering.²⁸ Und der Vergleich von Gutachterinnen und Gutachtern zum jeweils selben Manuskript zeigt nur eine leichte Tendenz zum mildereren Urteil durch Gutachterinnen.²⁹ Vergleicht man die Publikationschancen von Frauen mit denen von Männern, zeigen sich zumindest bis zum Abschluss des ersten *Review*-Verfahrens keinerlei Differenzen. 27% aller eingereichten Manuskripte und 27% der von Frauen eingereichten Manuskripte wurden nach dem ersten *Review*-Verfahren publiziert und nahezu genau so viele vor bzw. in der Begutachtung abgelehnt (gesamt: 28%, Frauen: 27%). Eine geschlechtsspezifische Differenz ergibt sich allerdings bei der Wiedereinreichung zur nochmaligen Begutachtung: Während die Begutachtung in 75% der von Frauen wiedereingereichten Manuskripte erfolgreich endete (78% bei allen wiedereingereichten Manuskripten), nutzten nur 27% der Frauen, aber 51% der

25 Bei sieben Manuskripten gab es übereinstimmende Voten der beteiligten Gutachterinnen und Gutachter; wenn sie differierten, gab bei neun Vergleichen der Gutachter das bessere Votum ab und nur bei vier Vergleichen die Gutachterin. Da in fast allen Fällen drei Gutachten pro Manuskript vorlagen, gibt es mehr als zwölf Vergleichsmöglichkeiten zwischen weiblichen und männlichen Gutachtenden.

26 Da es sich bei den vier Bewertungskategorien für Manuskripte im *Peer Review*-Verfahren der ZIB nicht um metrisch skalierte Variablen handelt, kann der arithmetische Mittelwert aus den Voten streng genommen nur eine Hilfsfunktion übernehmen. Er liefert jedoch interessante Hinweise bezüglich der vermuteten Zusammenhänge.

27 Für den Zusammenhang von gutachterlicher Bewertungsdifferenz und Geschlecht der AutorInnen eines Manuskripts beträgt der Rangkorrelationskoeffizient r_s 0,004.

28 Der Kontingenzkoeffizient beträgt 0,069. Zudem ist dieses Ergebnis nicht signifikant ($p = 0,543$).

29 Beim Vergleich von 52 Review-Verfahren stimmten Gutachterinnen und Gutachter in 22 Fällen überein, in 13 Fällen bewerteten Männer das Manuskript besser als Frauen, und umgekehrt war dies 17 Mal der Fall.

männlichen Autoren diese Möglichkeit! Liegen die Ursachen dafür vor allem bei den Wissenschaftlerinnen oder im männlich geprägten Wissenschaftssystem, verkörpert etwa in einer rein männlichen Herausgeber-Riege bei der *Zeitschrift für Internationale Beziehungen* inklusive eines männlichen Redakteurs in den ersten acht Jahren, der für die Kommunikation mit den AutorInnen verantwortlich war?

Während das Geschlecht von AutorInnen an diesem Punkt des *Peer Review*-Verfahrens Einfluss auf das Publikationsverhalten hat, scheinen die Beurteilungsergebnisse des Verfahrens selbst geschlechtsblind (Krais 2000c: 33) zu sein, sowohl was das Geschlecht der AutorInnen als auch das der Gutachtenden betrifft. Erstes kann vornehmlich als Erfolg der Anonymisierung von Manuskripten betrachtet werden (vgl. Wennerås/Wold 1997), während Letzteres dem ebenfalls nicht signifikanten Einfluss anderer Gutachtenden-Merkmale auf ihre Bewertungen entspricht. Auch das Qualifikationsniveau der Gutachtenden (ProfessorIn, habilitiert, promoviert, nicht-promoviert) wirkt sich nicht systematisch auf ihre Urteile aus. Weder ProfessorInnen noch Habilierte bewerten dieselben Manuskripte bzw. bewerten grundsätzlich schlechter oder besser als geringer Qualifizierte. Die Mittelwerte der verschiedenen Qualifikationsgruppen unter den Gutachtenden weichen nur minimal voneinander ab,³⁰ und beim Vergleich der Urteile bezogen auf dasselbe Manuskript zeigen sich auch keine aussagekräftigen Zusammenhänge.³¹

4. Die Qualität von Gutachten im Peer Review-Verfahren

Objektivierbare Eigenschaften von Gutachtenden wie ihr formales Qualifikationsniveau oder ihr Geschlecht können ganz offensichtlich die Beurteilungsdifferenzen hinsichtlich der Publikationsfähigkeit von Manuskripten ebenso wenig erklären wie AutorInnen-Eigenschaften, etwa deren Geschlecht (vgl. Abschnitt 3). Die Unterschiede bei der Begutachtung sind wohl vor allem auf individuelle Motive, Maßstäbe und Strategien der Gutachtenden zurückzuführen (vgl. Abschnitt 1). Das macht es weder für die HerausgeberInnen einer Zeitschrift, noch für die potenziellen AutorInnen leicht, das *Peer Review*-Verfahren als Entscheidungsverfahren über die (Nicht-)Publikation von Aufsätzen in renommierten Zeitschriften – und damit auch über wissenschaftliche Karrierechancen – uneingeschränkt anzuerkennen. Hinzu kommt die große Uneinheitlichkeit in der Qualität von Gutachten und die darin auch zum Ausdruck kommende varierende Anerkennung dieses Verfahrens. Dafür lassen sich verschiedene Gründe anführen (Abschnitt 4.1), die partiell auch in der Verantwortung von HerausgeberInnen und Zeitschriften-Redaktionen liegen könnten. Einer davon scheint mir die mangelnde Kommunikation über Qualitäts-Standards für Gutachten zu sein, die im dann folgenden Abschnitt (4.2) angeregt werden soll.

30 Die Mittelwerte (und Standardabweichungen) betragen für Promovierte 2,82 (0,859), für Habilitierte 2,74 (0,994), für ProfessorInnen 2,68 (0,959) und insgesamt 2,72 (0,931). Die dabei erkennbar werdende leichte Tendenz zum milderen Urteil der höher Qualifizierten entspricht den Ergebnissen anderer Studien (vgl. Hirschauer 2004: 66).

31 Spearman's r_s liegt jeweils im Intervall $\pm 0,2$.

4.1. Ungerechtigkeiten des Peer Review-Verfahrens

HerausgeberInnen und Zeitschriften-Redaktionen sehen sich mit widersprüchlichen Voten aus den *Review*-Verfahren konfrontiert und müssen eigene Maßstäbe entwickeln, welchen Gutachten sie in ihrer Entscheidung eher folgen wollen. Je bedeutsamer aber solche HerausgeberInnen-Entscheidungen bei uneinheitlichen GutachterInnen-Voten sind, desto fragwürdiger wird der enorme Aufwand an Zeit und intellektueller Arbeit, der von einer *scientific community* in die *Peer Review*-Verfahren ihrer Zeitschriften investiert werden muss. Ein 30- bis 40-seitiges Manuskript aufmerksam zu lesen und ein etwa zweiseitiges Gutachten mit Gründen für die Bewertung und hilfreichen Überarbeitungshinweisen zu erstellen erfordert selten weniger als einen halben Arbeitstag, bei schwierigen Manuskripten und eingehend-ausführlichen Gutachten auch gerne einen ganzen oder noch mehr. Für die 452 Gutachten, die das ZIB-*Review-Panel* in den ersten acht Jahren (einer nur halbjährlich erscheinenden Zeitschrift!) erstellt hat, wurde etwa die Arbeitszeit eines Wissenschaftlers bzw. einer Wissenschaftlerin von deutlich über einem ganzen Jahr investiert. Ist das zu rechtfertigen, wenn am Ende, vor allem in uneindeutigen Fällen, doch die Entscheidungen der HerausgeberInnen, sowohl durch die Auswahl der Gutachtenden als auch bei der Bewertung von Beurteilungsdifferenzen im externen *Review*-Verfahren, den größten Einfluss auf das Ergebnis der Publikationsentscheidung haben?

Dass sich HerausgeberInnen einer Zeitschrift sowohl zur Steigerung des Renommee ihrer Zeitschrift als auch aus sozialen Gründen gerne eines *Peer Review*-Verfahrens zur Vorbereitung ihrer Publikationsentscheidungen bedienen, ist nahe liegend und verständlich. Wie engagiert sich dagegen die betroffene *scientific community* an dieser Aufgabe beteiligt, ist äußerst fraglich, denn zu viele Ungerechtigkeiten behindern ein qualitativ hochstehendes *Peer Review*-Verfahren:

- Gutachten zu erstellen kostet Zeit, ohne dass die Gutachtenden einen dem Zeitaufwand äquivalenten Nutzen daraus ziehen könnten.
- Qualitativ hochstehende Gutachten zu erstellen erfordert einen höheren Zeitaufwand und mehr Arbeit, wird aber wegen der Anonymität des Verfahrens von niemandem positiv honoriert oder gar belohnt.
- Wer qualitativ hochstehende Gutachten schreibt, wird von HerausgeberInnen lieber und damit häufiger als »schlechte« Gutachtende für das *Peer Review*-Verfahren herangezogen – und damit »bestraft«, weil sie oder er mehr Zeit und Arbeitskraft für diese Kollektivaufgabe bereitstellen muss.
- Wer qualitativ schlechte Gutachten schreibt, wird nicht sanktioniert, wahrscheinlich sogar »belohnt«, indem sie oder er in Zukunft nicht mehr um weitere Begutachtungen gebeten wird.
- Qualitativ schlechte Gutachten haben ähnlich großen Einfluss auf die Publikationsentscheidung wie qualitativ hochstehende Gutachten.³²

32 Erst im Falle eines nicht-eindeutigen Gesamtergebnisses des *Review*-Verfahrens werden HerausgeberInnen in besonderer Weise auf die Konsistenz zwischen den Voten und ihrer gutachterlichen Begründung achten und in diesem Zusammenhang einem qualitativ hochstehenden Gutachten – wenn es sie überzeugt – mehr Gewicht geben als schlecht gemachten Gutachten.

- Wer die Bitte um Beteiligung am *Peer Review*-Verfahren ablehnt, hat keine Nachteile und entzieht sich sanktionslos der Aufgabe, zu einem Kollektivgut beizutragen (vgl. Meier 1997: 563).

Dies führt insgesamt dazu, dass das *Peer Review*-Verfahren häufig eher als notwendiges Übel hingenommen denn als gemeinsame Aufgabe wahrgenommen wird. Auch wenn die GutachterInnen sich darüber im Klaren sind, dass sie weniger an der Qualitätsmessung von Forschung als vielmehr an der »Steuerung von Leseraufmerksamkeit« (Hirschauer 2004: 79) beteiligt sind, tragen sie doch gerade damit zur Verbesserung der Kommunikation innerhalb einer *scientific community* bei. Gutachten im *Peer Review*-Verfahren sind ein Teil dieser Kommunikation, deren Ziel auch darin besteht, zur Optimierung der Texte beizutragen, die von vielen gelesen werden (sollen).

»Peer review« sichert die Offenheit der Kommunikation bei gleichzeitiger Qualitätskontrolle. Das bedeutet: Sie schafft intern das Vertrauen der Wissenschaftler in die Verlässlichkeit ihrer Kommunikationen. Das wechselseitige Vertrauen der Wissenschaftler in die Wahrhaftigkeit der Behauptungen des anderen ist die Voraussetzung für die erfolgreiche Kommunikation. Sie ist damit der institutionelle Kern des wissenschaftlichen Kommunikationsprozesses« (Weingart 2001: 285).

Auch das Renommee einer begutachteten wissenschaftlichen Zeitschrift ist in entscheidender Weise vom Ansehen des *Review*-Verfahrens abhängig. Gewinnen die GutachterInnen den Eindruck, auch mit schlecht begründeten *Reviews* Einfluss auf die Veröffentlichungsentscheidungen der HerausgeberInnen nehmen zu können, oder dass diese sich regelmäßig über die – schlecht begründeten – Voten der GutachterInnen hinwegsetzen (müssen), wird das Niveau von Gutachten ständig weiter sinken und auch dem Renommee der Zeitschrift schaden. Gewinnen auch AutorInnen vermehrt den Eindruck, dass die Gutachten, mit denen über die Nicht-Publikation ihrer Arbeiten entschieden wird, oberflächlich, unangemessen und überheblich sind, indem sie sich auf das Manuskript, die Fragestellung und Argumentation nur teilweise oder gar nicht einlassen, wird auch das Interesse eines Großteils der *community*, Manuskripte bei dieser Zeitschrift einzureichen, spürbar nachlassen und die Gefahr entstehen, dass die Zeitschrift vornehmlich als Organ einer einzelnen Forschungsrichtung und der sie repräsentierenden Personen wahrgenommen wird. Auch wenn das doppelt anonyme *Peer Review*-Verfahren weitgehend im Verborgenen stattfindet, trägt seine Qualität doch entscheidend zum Image und damit auch zur wissenschaftlichen Anerkennung einer Zeitschrift und ihrer *scientific community* bei.

Hinzu kommt die Gefahr, dass die oben genannten Ungerechtigkeiten auch Einfluss auf die Voten der GutachterInnen nehmen können. So mag sich der eine oder die andere wegen viel zu vieler Verpflichtungen und Termindruck vor allem aus Zeitmangel für die Kategorie »wie vorliegend ohne Änderungen zur Veröffentlichung annehmen« entscheiden, weil für dieses Votum ein Manuskript nur oberflächlich gelesen und keine inhaltlich aufwändigen Begründungen und Überarbeitungsanregungen geliefert werden müssen.³³ Auch eine Publikationszusage mit Überarbeitungsbedarf lässt sich

33 Dies bestätigt auch die Untersuchung der Gutachten, die eine Veröffentlichung ohne Überarbeitung vorgeschlagen haben, von denen 44% als mangelhaft (zu den Bewer-

leichter und schneller begründen als eine grundlegende Überarbeitung im Hinblick auf eine Wiedereinreichung des Manuskripts. Während sich mit Leichtigkeit zu jedem Manuskript diverse mehr oder weniger periphere Überarbeitungsvorschläge anführen lassen, setzt die Begründung einer grundlegenden Überarbeitung ein stärkeres Einlassen auf den Text und seine Argumentation voraus.³⁴

Aus vielerlei Gründen erscheint es notwendig, nicht nur an die zu publizierenden Manuskripte, sondern auch an die zu dieser Frage erstellten Gutachten hohe Qualitätsmaßstäbe anzulegen. Die Mitglieder eines *Review-Panels* zu motivieren oder gar zu verpflichten, gewisse Standards bei ihren Gutachten einzuhalten, setzt eine veränderte Anreizstruktur voraus. Nicht die Schnell-Gutachtenden und die TrittbrettfahrerInnen dürfen bei *Peer Review*-Verfahren belohnt werden, sondern diejenigen müssen öffentliche Anerkennung finden, die qualitativ hochstehende Gutachten schreiben. Dies erfordert ein behutsames Auflösen der Anonymität nach Abschluss des Begutachtungsprozesses als auch neue Publikationsforen, in denen die inhaltlichen Kontroversen zwischen AutorInnen und Gutachtenden der gesamten *community* zugänglich gemacht werden können (siehe unten Abschnitt 5). Zunächst aber stellt sich die Frage: Was zeichnet ein qualitativ hochstehendes Gutachten zu einem Zeitschriften-Manuskript aus? Sowohl aus der Funktion von Gutachten im Kommunikationsprozess zwischen AutorInnen und der Redaktion einer Zeitschrift als auch aus der systematischen Evaluation von 452 Gutachten lassen sich die folgenden fünf Anforderungen ableiten, die für die Qualität von Gutachten im *Peer Review*-Verfahren von Zeitschriften eine wichtige Rolle spielen.

4.2. Qualitäts-Standards für Gutachten

Gutachten im *Peer Review*-Verfahren von wissenschaftlichen Zeitschriften wenden sich sowohl an die HerausgeberInnen der Zeitschrift als auch an die AutorInnen der begutachteten Manuskripte. Während Letzteren in aller Regel Hinweise gegeben werden, wie sich ein Manuskript verbessern ließe,³⁵ sollen die HerausgeberInnen eine begründete Entscheidungsgrundlage für den weiteren Umgang mit dem Manuskript erhalten. Dabei gelten der Neuigkeitswert der Forschungsresultate, Genaugigkeit und Richtigkeit der Untersuchung sowie die Relevanz der Ergebnisse als weithin akzeptierte – wenngleich meist nur implizit verwendete – Indikatoren für Publikationsentscheidungen (vgl. Daniel 1993: 1; Felt 1999: 15; Prowiss 2004). Doch solche und ähnliche Aspekte, die aus Sicht der Gutachtenden entscheidend zu ihren Gesamtvoten beitragen, verlieren sich oft in den gutachterlichen Texten, die

-
- tungsmaßstäben siehe unten Abschnitt 4.2) einzuschätzen waren. Überlastung und Zeitnot gelten als bedeutsame Faktoren für Schwächen oder das teilweise Versagen wissenschaftlicher Qualitätskontrolle durch Begutachtungssysteme (vgl. Finetti 2000).
- 34 Das bedeutet nicht, dass Zeitdruck und Arbeitsvereinfachung die zentralen Motive bei der Beurteilung von Manuskripten sind, aber doch Faktoren, die sich auf das Beurteilungsergebnis auswirken können und deren Einfluss sich durch höhere Qualitätsansprüche an die Gutachten möglicherweise begrenzen ließe.
- 35 Auch die Begründung einer Ablehnung basiert in den meisten Fällen auf hypothetischen und als unerfüllbar eingeschätzten Überarbeitungserfordernissen.

immer zugleich Ausweis eigener Expertise, Mitteilung an die AutorInnen im Hinblick auf Verbesserungsmöglichkeiten des Textes und Rechtfertigung für die erfolgte Bewertung sein wollen. Um allen Beteiligten den Umgang mit dieser Textsorte »Gutachten« zu erleichtern, bietet sich eine stärkere Formalisierung gutachterlicher Texte an (vgl. auch Daniel 1993: 63f). Diese sollte sich jedoch weniger an den verschiedenen Adressaten der Gutachten (Redaktion, HerausgeberInnen, AutorInnen) orientieren, als vielmehr an den Funktionen des Verfahrens: Die Gutachten aus dem *Peer Review*-Verfahren sollen den HerausgeberInnen fundierte Publikationsentscheidungen ermöglichen und den Verbesserungsbedarf eines Manuskripts so darlegen, dass die Redaktion einer Zeitschrift seine Realisierbarkeit gut einschätzen kann und die AutorInnen ihn leicht nachvollziehen und bei einer Überarbeitung gut umsetzen können.

Manche Gutachtenden kommen diesen verschiedenen Anforderungen in der Weise nach, dass sie parallel zum für die AutorInnen bestimmten Gutachten einen expliziten Kommentar für die HerausgeberInnen und die Redaktion schreiben, den die AutorInnen nicht zu Gesicht bekommen sollen. Diese getrennten Botschaften der Gutachtenden werfen aber mehr Probleme auf als sie beseitigen können, nähren sie doch den Verdacht einer Komplizenschaft zwischen Gutachtenden und HerausgeberInnen zuungunsten bestimmter AutorInnen. Damit wird die Transparenz des Verfahrens reduziert, was der breiten Anerkennung des *Peer Review*-Verfahrens eher abträglich ist. Zum anderen reduziert sich damit die Anforderung an den für die AutorInnen geschriebenen gutachterlichen Text, darin eine nachvollziehbare Begründung für das abgegebene Votum zu liefern. Diese argumentative Auseinandersetzung mit dem zu begutachtenden Text muss aber als eigentlicher Kern des *Peer Review*-Verfahrens betrachtet werden. An kaum einer anderen Stelle wissenschaftlicher Kommunikation findet eine eingehendere Auseinandersetzung mit den Forschungsergebnissen statt, die den jeweils eigenen Fragestellungen in aller Regel eng verwandt sind. Deshalb würde ein erhöhtes Qualitätsniveau von Gutachten nicht nur zur Steigerung von Akzeptanz und Validität des *Peer Review*-Verfahrens und damit auch des Renommes der entsprechenden Zeitschriften beitragen, sondern auch eine wesentliche Verbesserung innerdisziplinärer Kommunikation bedeuten. Dass davon auch die Produktivität einer Disziplin profitieren könnte, scheint keine unbegründete Erwartung zu sein.

Vor diesem Hintergrund werden nachfolgend fünf Anforderungen an Gutachten im *Peer Review*-Verfahren von Zeitschriften dargestellt und dieser Kriterien-Katalog anschließend (Abschnitt 4.3) zur Bewertung der vom ZIB-*Review-Panel* erstellten Gutachten herangezogen.

(1) Grundlage der Beurteilung eines Manuskripts ist dessen Argumentation im Kontext eines bestimmten wissenschaftlichen Diskurses. Ihre Relevanz, Stichhaltigkeit oder Überzeugungskraft kann unterschiedlich eingeschätzt werden, aber von großer Bedeutung bezüglich der Publikationsentscheidung ist, dass möglichst von allen Gutachtenden und den HerausgeberInnen dieselbe Argumentation als inhaltlicher Kern des Manuskripts und damit als Begutachtungsobjekt wahrgenommen wird. Nicht selten haben AutorInnen nach der Lektüre kritischer Gutachten zu ihren Manu-

skripten den Verdacht, ihr Text sei nicht richtig gelesen, nicht verstanden oder zumindest ihr Hauptargument falsch interpretiert worden. Ursache hierfür kann ein schlecht geschriebener, falsch strukturierter oder begrifflich unzugänglicher Text sein, aber auch die oberflächliche, interessenlose oder mehrfach unterbrochene Lektüre durch die Gutachtenden oder die HerausgeberInnen. Zudem können die inhaltlichen Aufmerksamkeiten anders gesteuert sein, etwa durch einen anderen Wissens-, Fragen- und/oder Theoriekontext oder die von den eigenen Forschungsinteressen beeinflussten Erwartungen an einen Text. Jeder Text wird von unterschiedlichen Personen verschieden gelesen,³⁶ weshalb jedes Gutachten zunächst deutlich machen sollte, wie der Text verstanden wurde, was als zentrales Argument, als roter Faden der inhaltlichen Stoßrichtung identifiziert wurde. Die *einleitende Rekonstruktion der Argumentation* ermöglicht den HerausgeberInnen zu erkennen, ob das Manuskript bei allen Gutachtenden zumindest ähnlich interpretiert wurde und ob diese Rekonstruktion mit der eigenen Wahrnehmung des Textes übereinstimmt. Sie bietet AutorInnen die Möglichkeit zu erfahren, wie ihr Manuskript verstanden wurde bzw. werden kann und auf der Grundlage welcher Lesart die Bewertung vorgenommen wurde.

(2) Nicht allein die Qualität eines Manuskripts, sondern auch die Relevanz der Fragestellung, der Stellenwert der behandelten Thematik im aktuellen wissenschaftlichen Diskurs spielt eine wesentliche Rolle im Entscheidungsprozess um die (Nicht-)Publikation eines Zeitschriften-Aufsatzes. Bei höchster Relevanz der vorgebrachten Argumente für die aktuelle wissenschaftliche Debatte dürften Gutachtende durchaus kompromissbereiter sein hinsichtlich einzelner Qualitätsmerkmale eines Manuskripts als bei einer nebенästhetischen Thematik oder gar abseitigen Fragestellung. Die Relevanz der behandelten Fragestellung und die Qualität ihrer Bearbeitung und sprachlichen Darstellung sind zwei voneinander weitgehend unabhängige Kennzeichen eines Manuskripts, die in einem je spezifischen Mischungsverhältnis empirisch auftreten und entsprechend auch getrennt voneinander in die Bewertung der Publikationsfähigkeit eingehen sollten. Aus Sicht der HerausgeberInnen kann es wichtiger sein, dass möglichst relevante Fragestellungen in den veröffentlichten Aufsätzen behandelt werden, als dass allein die – vermeintliche – Manuskriptqualität die Publikationsentscheidungen bestimmt. Deshalb sollten Gutachten explizit auf die Frage nach der *Relevanz der behandelten Fragestellung* eine Antwort geben, liegt doch genau darin einer der wesentlichen Gründe zur Einbeziehung externer Gutachtender in die Publikationsentscheidungen über Zeitschriftenaufsätze. Nur diese können für spezifische

36 Dies sollte eigentlich nicht überraschen, bedeutet doch Lektüre die eigene Rekonstruktion des Gemeinten, eine Deutung auf der Grundlage eigenen Wissens und bewusster wie unbewusster Erwartungen (vgl. dazu Weller 2002: 18-30). Ganz ähnlich konstituieren sich auch die Eigenschaften eines Manuskripts »in der Beziehung zwischen den Absichten des Autors und den Erwartungen des Lesers. Seine Eigenschaften werden nicht bloß in einer statisch konzipierten Sozialbeziehung von Autor und Gutachter (Unterschiede von Rang, Alter, Geschlecht etc.) richtig oder falsch eingeschätzt, sie entfalten sich vielmehr erst in einer dynamischen Kommunikationsbeziehung, in der ein Autor Ansprüche erhebt und enttäuscht, Sympathien und Antipathien weckt, und ein Leser Erwartungen und ganz unterschiedliche Nutzungsinteressen im Kontext seiner eigenen Arbeiten mitbringt« (Hirschauer 2004: 74, Hervorh. dort).

Themenfelder und Teildiskurse einer Disziplin diese Relevanz angemessen beurteilen. Und die Einschätzungen der Relevanz der Fragestellung durch die Gutachtenden werden auch für die AutorInnen – unabhängig vom Ergebnis des *Review*-Verfahrens – von Bedeutung und Interesse sein. So können sie möglicherweise erkennen, dass nicht die Qualität ihres Manuskripts, sondern die eingeschränkte Relevanz der Fragestellung den Ausschlag gegen die Veröffentlichung in dieser Zeitschrift gegeben hat, und daher nicht die Überarbeitung, sondern die Suche nach einem anderen Publikationsort die angemessene Strategie sein könnte.

(3) Nur in den seltensten Fällen werden die Gutachtenden keine Ideen und Vorschläge zur weiteren Verbesserung eines Manuskripts haben – dies bestätigt auch die Häufigkeitsverteilung bei den Review-Kategorien.³⁷ Eher schon werden es zeitliche Einschränkungen oder die mangelnde Motivation sein, entsprechenden intellektuellen Aufwand für eine ausführliche Stellungnahme aufzubringen, die die Gutachtenden darauf verzichten lassen, bei einem zur Veröffentlichung empfohlenen Manuskript noch Überarbeitungsanregungen zu formulieren. Doch nicht alle Kommentare und Anmerkungen zur Überarbeitung sind für den weiteren Umgang mit dem Manuskript gleich hilfreich. Wenn Überarbeitungshinweise oder gar -auflagen gegeben werden, sollten sie immer mit konkreten Textbezügen versehen sein. »Die Argumentation muss stringenter dargestellt werden«, »das Ergebnis bleibt unverständlich und sollte besser an die Fragestellung rückgebunden werden«, »wichtige Literatur wurde nicht berücksichtigt« und ähnliche Kritikpunkte mögen die Eindrücke eines Gutachtenden widerspiegeln, sind aber nur von geringem Nutzen, sowohl für die AutorInnen als auch für die HerausgeberInnen und eine Zeitschriften-Redaktion. Letzteren fehlen die konkreten Anhaltspunkte, anhand derer sich eine Verbesserung des Manuskripts bei der Überarbeitung identifizieren ließe, sowie eine tragfähige, substanziale Grundlage, auf der sich die Kommunikation mit den AutorInnen zur Umsetzung der Überarbeitungshinweise gestalten lässt. Und AutorInnen bleiben mit solchen Überarbeitungshinweisen zumeist rat- und hilflos zurück, sind sie doch in aller Regel insgesamt mit ihrem Manuskript zufriedener als die Gutachtenden, während sie sich aber in Einzelfragen gerne darauf hinweisen lassen, dass etwas missverständlich, unpräzise oder zu ausführlich beschrieben wird, an einer anderen Stelle die argumentative Verknüpfung geklärt, ein bestimmter Aufsatz noch berücksichtigt oder noch ein Beispiel eingefügt werden sollte. In vielen Fällen ist auch die Formulierung von Fragen, die den Gutachtenden wichtig sind, aus ihrer Sicht im vorliegenden Text aber bisher unbeantwortet bleiben, hilfreich für den Überarbeitungsprozess, denn dadurch können AutorInnen sehr gut die Lücken erkennen, die sie in der sprachlichen Darstellung ihrer Argumentation bisher gelassen haben. Möglicherweise können Gutachtende an verschiedenen Stellen sogar Überarbeitungsvorschläge entwickeln und bezogen auf zentrale Stellen der Argu-

37 Wenn die Veröffentlichung eines Manuskripts empfohlen wird, geschieht dies in 76% der Fälle anhand der Kategorie »Mit Veröffentlichungszusage zur Überarbeitung zurückgeben!«. Insgesamt wurde diese Kategorie bei 31,4% der Gutachten gewählt, 10% plädierten für »drucken wie vorgelegt!«, 23,2% für »ablehnen!« und 35,4% für »revise and resubmit!«.

mentation darstellen, *wie sie* das Manuskript verbessern würden. Dies aber setzt voraus, dass sich die Gutachtenden auf die behandelte Fragestellung und die Bearbeitungsstrategie der Autorin bzw. des Autors einlassen und Empathie für die ihnen unbekannten AutorInnen entwickeln.³⁸

(4) Die besten Gutachten sind jene, denen der Wunsch zugrunde liegt, dass zu der Fragestellung, die das zu begutachtende Manuskript aufgeworfen hat, ein möglichst guter, überzeugender Aufsatz erscheint. Auch dies setzt voraus, dass die Gutachtenden sich die jeweilige Problemstellung der AutorInnen zu eigen machen und sich zugleich – weitgehend – auf deren Bearbeitungsstrategie einlassen. Dies ist nicht nur bei Manuskripten, für die eine grundlegende Überarbeitung und anschließende Wiedereinreichung empfohlen wird, nicht einfach, sondern auch bei einer Publikationszusage mit Überarbeitungshinweisen. Zu viele individuelle Motive der Gutachtenden können diesem inhaltlichen Einlassen, dieser empathischen Lektüre und der Formulierung konstruktiver Verbesserungsvorschläge in einem motivierenden Stil entgegenstehen, sei es der Zeitmangel, die ausbleibende »Belohnung« für solche Arbeit oder mögliche Konkurrenzgefühle gegenüber den unbekannten AutorInnen. Nichtsdestotrotz sollte *das inhaltliche Einlassen auf die Fragestellung des Manuskripts* und nicht die Projektion eigener Forschungsprobleme in das Manuskript eine wichtige Orientierungsgröße und selbstkritische Prüffrage beim Fertigstellen von Gutachten sein.

(5) Ein letztes Kennzeichen qualitativ hochstehender Gutachten betrifft die *Form*, die insbesondere für den Umgang der AutorInnen mit den Gutachten von großer Bedeutung ist. Unabhängig davon, ob Überarbeitungshinweise mit einer Publikationszusage oder der Empfehlung zur Wiedereinreichung verknüpft sind, können sie motivierend wirken oder eine ablehnende Haltung bei den AutorInnen hervorrufen. Setzt ein Gutachten gleich im ersten Satz mit Kritik ein, wird manche Autorin oder mancher Autor den Hinweisen weniger gern und motiviert folgen wollen, als wenn zunächst die Inhalte des Manuskripts gewürdigt werden. Die einleitende Rekonstruktion der Argumentation, wie sie oben (1) vorgeschlagen wurde, erzwingt gewissermaßen diese Form, dass Kritik, die den Überarbeitungsvorschlägen zugrunde liegt, an zweiter Stelle vorgebracht wird. Und ob sie eher als destruktive Kritik am gesamten Manuskript oder als konkrete Verbesserungsvorschläge bestimmter Textabschnitte in einer zurückhaltenden, aber motivierenden Sprache formuliert werden, macht einen großen Unterschied. Auch in diesem Aspekt werden sich selbstkritische Gutachtende vor der Abgabe des Gutachtens noch fragen, welche versteckten Motive sie möglicherweise dazu verleitet haben könnten, eher konfrontative, provokative als empathisch-wohlwollende Formulierungen benutzt zu haben oder warum ein Gutachten allzu knapp ausgefallen ist. Ist es die Sorge, dass andere von den eige-

38 Neben allerhand Kritik an der Begutachtung ihrer Texte überwog in meiner Kommunikation mit den AutorInnen der ZIB doch die Anerkennung und der Dank für die zahllosen gutachterlichen Hinweise zur Verbesserung der publizierten Beiträge. »So intensiv hat sich noch nie jemand mit meiner Arbeit auseinandergesetzt« mag als beispielhafte Rückmeldung diese Wahrnehmung des *Peer Review*-Verfahrens auf AutorInnen-Seite veranschaulichen.

nen Ideen, Überlegungen und Vorschlägen profitieren könnten? Im Extremfall können Gutachtende auf einem solchen Wege zu »latenten KoautorInnen« werden (Hirschauer 2004: 67), was einen weiteren Schwachpunkt der Anonymität im *Peer Review*-Verfahren darstellt. Doch in den meisten Fällen wird auch ein *sehr ausführliches und motivierendes Gutachten* nicht zu einem grundlegend veränderten, aber möglicherweise doch zu einem substanziel verbesserten Aufsatz führen (vgl. aber Frey 2003). Dies ist, neben fundierteren Entscheidungen über die (Nicht-)Publikation von Zeitschriftenaufsätzen, ein wesentliches Ziel des *Peer Review*-Verfahrens, dem durch ausführliche und eine die Überarbeitung motivierende Form der Gutachten entsprochen wird.

4.3. Die Qualität der ZIB-Gutachten

Die Fähigkeit, Texte zu beurteilen, entwickelt sich mit der Menge des Gelesenen, vielleicht schon in der Schule, hoffentlich während des Studiums. Sie wird bedeutsamer, je breiter die bearbeiteten Fragestellungen und damit die Mengen an einschlägiger Literatur werden, denn das zu Lesende muss vom Unlesbaren unterschieden werden. Doch Texte zu *begutachten*, einen eigenen Text zu der Frage zu verfassen, warum ein Manuscript (nicht, noch nicht) publiziert werden soll und wie es noch verbessert werden könnte, wird an keiner Stelle politikwissenschaftlicher Ausbildungen gelehrt oder gar geübt.

»Da im Grunde niemand angeben kann, was einen guten Antrag oder was ein gutes Gutachten ausmacht und die Rückmeldungen auf entsprechende Versuche eher dürfzig (meist nur ›ja‹ oder ›nein‹) sind, bleibt nur ›trial and error‹ die Auswertung des Kollegenklauschs und abergläubisches Lernen (das heißt, man kopiert sich im Zweifelsfall selbst). Obwohl Idiosynkrasien blühen, driftet Gutachter mit der Zeit in eine bestimmte Richtung, entwickelt und reproduziert sich auf rätselhafte Weise eine identifizierbare Gutachtenkultur« (Wolff 2000: 37).

Um für die Kultur des *Peer Review*-Verfahrens und die Verbesserung innerdisziplinärer Kommunikation einige Orientierungspunkte zu entwickeln, wurden hier aus den Funktionen des Verfahrens sowie aus den Erfahrungen der Kommunikations- und Vermittlungsarbeit zwischen Gutachtenden und AutorInnen fünf Kriterien abgeleitet, die ein qualitativ hochwertiges Gutachten kennzeichnen:

- (1) die einleitende Rekonstruktion der Argumentation;
- (2) Aussagen zur Relevanz der Fragestellung;
- (3) konkrete Überarbeitungshinweise;
- (4) das inhaltliche Einlassen auf die Zielsetzung des Manuskripts;
- (5) ein angemessener Umfang und ein motivierender Stil.

Dies mag manchen als zu hoher Maßstab erscheinen angesichts der Tatsache, dass solche Gutachten für Zeitschriften eine unentgeltliche Service-Leistung darstellen, für die man viele Stunden seiner ständig zu knappen Zeit opfern muss, ohne einen sichtbaren Gewinn zu erzielen. Doch die Analyse aller in den ersten acht Jahren für die ZIB geschriebenen Gutachten zeigt, dass 12% diese Ansprüche durchaus erfüllen und als sehr gute Gutachten bewertet werden konnten. Bei dieser Begutachtung der Gutachten wurde so vorgegangen, dass alle fünf Kriterien gleichgewichtig bewertet wurden. Wenn alle fünf Elemente eines qualitativ hochwertigen Gutachtens in einem

gutachterlichen Text vorhanden waren, galt er als sehr gut. Fehlte eines der fünf Elemente, galt er noch als gut; konnte ein Gutachten nur einem oder gar keinem der fünf Kriterien entsprechen, wurde es als mangelhaft bewertet. Nach dieser Einteilung lagen die meisten Gutachten im qualitativen Mittelfeld (siehe Abbildung 2).

Abbildung 2: Geschlecht der Gutachtenden und Qualität ihrer Gutachten

<i>Qualität des Gutachtens</i>	<i>Gesamtzahl so bewerteter Gutachten</i>	<i>verfasst von Gutachtern</i>	<i>verfasst von Gutachterinnen</i>
sehr gut	54 (11,8%)	46 (11,6%)	8 (14,3%)
gut	106 (23,5%)	89 (22,5%)	17 (30,4%)
befriedigend	125 (27,7%)	107 (27,0%)	18 (32,1%)
ausreichend	97 (21,5%)	88 (22,2%)	9 (16,1%)
mangelhaft	70 (15,5%)	66 (16,7%)	4 (7,1%)
Gesamt	452 (100%)	396 (87,6% aller Gutachten)	56 (12,4% aller Gutachten)

Die bisherige Qualität der Gutachten im ZIB-Review-Verfahren lässt also durchaus Spielraum für Verbesserungen und die damit verbundene Aufwertung der Zeitschrift, ihres Renommes und der von ihr initiierten wissenschaftlichen Kommunikation. Fragt man nach Zusammenhängen zwischen Merkmalen von Gutachtenden und ihrer Gutachten-Qualität, ergeben sich konkretisierte Handlungsfelder für eine veränderte Ausgestaltung des *Peer Review*-Verfahrens (siehe unten Abschnitt 5). Prüft man die möglichen Zusammenhänge zwischen der Qualität von Gutachten und dem Geschlecht der Gutachtenden, ergibt sich ein erstes interessantes Ergebnis: Obwohl nur 12,4% der Gutachten von Frauen verfasst wurden und daher der Vergleich auf einer relativ schmalen Basis vorgenommen werden muss, fällt doch auf, dass Gutachterinnen zur Menge der sehr guten, guten und befriedigenden Gutachten relativ mehr beigetragen haben als Gutachter, deren Anteile bei den ausreichend und mangelhaft bewerteten Gutachten leicht überproportional sind.

Dies könnte allerdings auch – angesichts nur weniger Professorinnen und weiblicher Habilitierter im ZIB-Review-Panel (siehe Schaubild 3) – ein Effekt eines anderen statistischen Zusammenhangs sein: jenes zwischen der Qualität von Gutachten und dem (formalen) Qualifikationsniveau der Gutachtenden. Hier zeigt sich eine recht starke negative Beziehung: Mit steigendem Qualifikationsniveau sinkt die Qualität der Gutachten und umgekehrt; die Qualität der Gutachten niedriger qualifizierter Gutachtender ist im Vergleich besser als die der höher Qualifizierten.³⁹ Ganz eindrücklich lässt sich dies anhand des in Abbildung 3 dargestellten Mittelwerteve-

³⁹ Statistisch ausgedrückt handelt es sich um einen positiven Rangkorrelationskoeffizient von 0,387 zwischen Qualifikationsniveau und Qualität/Benotung des Gutachtens (1-5). Dieses Ergebnis ist zudem hoch signifikant ($p = 0,00$).

gleichs der Qualität der Gutachten, differenziert nach erreichter Qualifikation und Geschlecht der Gutachtenden, veranschaulichen:

Abbildung 3: Gutachtende und die Qualität ihrer Gutachten

Qualifikation und Geschlecht der Gutachtenden	Mittelwert der Qualität der Gutachten	N	Mittelwert der Qualität der Gutachten (N)
unpromovierte Frau	1,67	3	1,44 (9)
unpromovierter Mann	1,33	6	
promovierte Frau	2,33	24	2,41 (123)
promovierter Mann	2,43	99	
habilitierte Frau	3,50	2	2,91 (54)
habilitierter Mann	2,88	52	
Professorin	3,11	27	3,43 (266)
Professor	3,46	239	
Insgesamt	3,05	452	3,05 (452)

Bei den unpromovierten Gutachtenden handelt es sich ausnahmslos um Personen, die zuvor schon als AutorInnen das ZIB-Review-Verfahren erfolgreich durchlaufen hatten⁴⁰ und ihre Gutachten folglich vor dem Hintergrund verfasst haben, jüngst selbst Adressat von Gutachten aus dem Peer Review-Verfahren gewesen zu sein. Mit der Einbeziehung von ZIB-AutorInnen und jüngeren KollegInnen in das Peer Review-Verfahren steigt dessen Qualität (vgl. auch Hellmann/Müller 2003b: 379), ohne dass zu befürchten wäre, DoktorandInnen und Promovierte würden spürbar großzügiger mit der Zuteilung von Veröffentlichungszusagen umgehen.⁴¹ Weder der Mittelwert ihrer Begutachtungsvoten unterscheidet sich deutlich von dem der ProfessorInnen und Habilitierten,⁴² noch beim Vergleich der Bewertungen derselben Manuskripte zeigen sich auffällige Unterschiede zwischen den unterschiedlich qualifizierten Gutachtenden (siehe Abschnitt 3).

Das Peer Review-Verfahren der ZIB ist also vom wissenschaftlichen Nachwuchs besser an- und aufgenommen worden als von den ProfessorInnen. Diese publizieren seltener in der ZIB, verglichen etwa mit dem ProfessorInnen-Anteil unter den PVS-AutorInnen,⁴³ und leisten bescheidenere Beiträge zur Anerkennung des Peer

40 Die Qualifikationsvoraussetzung zur Mitwirkung im *Review-Panel* der ZIB ist die Promotion oder, schon im ZIB-Review-Verfahren erfolgreich gewesen und damit AutorIn eines ZIB-Beitrags zu sein.

41 In verschiedenen Studien wurde eine schärfere Beurteilung durch weniger reputierte Gutachtende festgestellt (Hirschauer 2004: 66; vgl. auch Russett 1978).

42 Unpromovierte 2,33 (n = 9, Standardabweichung 0,500); Promovierte 2,82; Habilitierte 2,74; ProfessorInnen 2,68.

43 Für die letzten zehn Jahrgänge (1994–2003) betrug der ProfessorInnen-Anteil unter den AutorInnen bei der *Politischen Vierteljahresschrift* (PVS) 31,8%, bei der ZIB 24,0%. Berücksichtigt man dabei nur Aufsätze, ergibt sich für die PVS ein Wert von 23,5%, bei der ZIB von 19,7%.

Review-Verfahrens. Man könnte hoffen, dass die dargestellten Zusammenhänge stärker auf die Generationszugehörigkeit zurückzuführen sind als auf den Status innerhalb des Wissenschaftssystems. Dass sich die Gutachten von Habilitierten in ihrer Qualität negativ von denen der Promovierten absetzen (siehe Abbildung 3), lässt allerdings daran zweifeln und die Notwendigkeit entstehen, Initiativen zu ergreifen für eine höhere Anerkennung der im *Peer Review*-Verfahren geleisteten wissenschaftlichen Arbeit und für die Qualitätssteigerung der Gutachten, insbesondere jener aus den Federn der ProfessorInnen.

5. Perspektiven des Peer Review-Verfahrens

Die Gründung der *Zeitschrift für Internationale Beziehungen*, deren Herausgeber ihre Publikationsentscheidungen von Anfang an auf der Grundlage von Gutachten aus einem doppelt anonymen *Peer Review*-Verfahren getroffen haben, hat eine deutliche Zunahme externer Begutachtungen von Zeitschriftenmanuskripten in der deutschen Politikwissenschaft mit sich gebracht. Eine große Zahl von Gutachten ist in diesem Zusammenhang in den vergangenen zehn Jahren geschrieben, AutorInnen um Überarbeitungen gebeten und Manuskripte abgelehnt oder veröffentlicht worden. Viel Arbeit und Zeit ist von einem Teil der politikwissenschaftlichen *community* dafür aufgebracht worden, den so genannten begutachteten Zeitschriften zu erhöhtem Renommee, den darin publizierenden AutorInnen zu besonderer Anerkennung und zugleich zu kompetenten Hilfestellungen für die Verbesserung ihrer Manuskripte zu verhelfen. Diese kollektive Investition in die *Peer Review*-Verfahren von Zeitschriften leistet aber nicht nur einen Beitrag zur Verlässlichkeit der wissenschaftlichen Kommunikation, sondern auch zum gesellschaftlichen Vertrauen in die Verlässlichkeit der Wissenschaft (vgl. Weingart 2001: 286f).⁴⁴ Beides wirkt heute bedroht angesichts aufgedeckter schwerwiegender Betrugsfälle in der Wissenschaft – trotz der *Peer Review*-Verfahren – und der verschwindenden Grenzen zwischen Wissenschaft und Gesellschaft (vgl. Nowotny et al. 2004; Weingart 2001, 2004). Die Anforderungen an das *Peer Review*-Verfahren wachsen. Es gilt, nicht nur die möglichen Auswirkungen der offensichtlichen Schwächen dieses Verfahrens im Auge zu behalten (vgl. Hirschauer 2004), sondern auch auf dem Hintergrund vorliegender Erfahrungen neue Instrumente und Kommunikationsmöglichkeiten zu erproben (vgl. Harnad 2000). Dabei können an dieser Stelle keine Abwägungen von Chancen und Risiken solcher veränderter Verfahren vorgenommen werden, die weitgehend auf Spekulationen beruhen würden. Entsprechende Entscheidungen über die Ausgestaltung zukünftiger *Peer Review*-Verfahren und des damit steigenden oder sinkenden Renommees ihrer Zeitschriften müssen die entsprechenden Her-

44 Dabei dürfen jedoch die von Hirschauer (2004) erwähnten problematischen Effekte nicht übersehen werden, die dann zustande kommen, wenn die Wissenschaftsforschung »ein Fremdstereotyp von Wissenschaft – dass diese sicheres und objektives Wissen generiere – in ihre Selbstbeschreibung« übernimmt und damit »die Erwartungen (und Hoffnungen) der Politik [bestätigt], dass Wissenschaft politikferner sei als sie es tatsächlich ist« (Hirschauer 2004: 80).

ausgeberInnen und Beiräte treffen. Und die Überprüfung der erhofften oder befürchteten Effekte bleibt dann einer zukünftigen *Peer Review*-Forschung überlassen, deren Augenmerk sich wohl noch stärker auf die Kommunikationsprozesse anstatt auf deren statische Ergebnisse in Form gedruckter Aufsätze richten wird (vgl. Hirschauer 2004: Kap. 3.4).

Wenn in diesen Kommunikationsprozessen der größte wissenschaftliche Wert von *Peer Review*-Verfahren liegt, geht es vornehmlich darum, die Qualität der Texte in diesem Prozess – und das sind an zentraler Stelle die Gutachten – zu verbessern.⁴⁵ Dabei ist es zunächst erforderlich, die Motivlagen der Gutachtenden für engagierte und qualitativ hochstehende Gutachten zu verbessern und mehr Transparenz und Öffentlichkeit für jene Kommunikationsprozesse herzustellen, die sich im Rahmen von *Peer Review*-Verfahren abspielen. Drei Problemfelder drängen sich damit auf, in denen Veränderungen und Ergänzungen der bisher praktizierten Verfahren versucht werden könnten, um *Peer Review*-Verfahren weiterzuentwickeln:⁴⁶ Die Qualität der Gutachten sollte nicht in jenem Maße variieren, wie es hier festgestellt wurde; die Anonymität der Gutachtenden ist nicht immer gleich bedeutsam und könnte in eng begrenztem Maße aufgehoben werden; nicht-publizierte Manuskripte und ihre Gutachten verschwinden, ohne greifbare Spuren zu hinterlassen. Gerade die Kombination papierlicher und elektronischer Publizität wissenschaftlicher Texte könnte es ermöglichen, dem *Peer Review*-Verfahren neue Perspektiven für eine stärker auf Kommunikationsprozesse vertrauende Selbstkontrolle der Wissenschaft zu eröffnen:

(1) *Qualität von Gutachten*

- Die Gutachtenden sollten möglichst ausführlich über das *Review*-Verfahren, an dem sie beteiligt waren, informiert werden, und in diesem Zusammenhang die anderen Gutachten und die HerausgeberInnen-Entscheidungen zur Kenntnis bekommen (vgl. dazu Heintz et al. 2004: 3), um Stellenwert und Wirkung ihrer Begutachtung realistisch einschätzen und bewerten zu können.⁴⁷
- HerausgeberInnen sollten sich zur Qualitätskontrolle von Gutachten verpflichten, die Maßstäbe für qualitativ hochstehende Gutachten transparent machen und den VerfasserInnen mangelhafter Gutachten zurückmelden, dass ihr Votum wegen Qualitätsmängeln des Gutachtens nicht berücksichtigt werden konnte.
- Die VerfasserInnen sehr guter Gutachten können beim obligatorischen Dank an alle GutachterInnen des zurückliegenden Jahrgangs gesondert erwähnt werden,

45 Schon allein der Gesamtumfang gutachterlicher Texte ist erheblich und umfasst für die hier untersuchten 452 Gutachten knapp 800 Seiten. Damit ca. 3000 Seiten ZIB-Beiträge veröffentlicht werden konnten, mussten zunächst fast 800 Seiten Gutachten geschrieben werden!

46 Zu verschiedenen Reformvorschlägen für das *Peer Review*-Verfahren vgl. Berger et al. (1997); Meier (1997); Harnad (2000); Ernst/Warwas (2003); Frey (2003); Fröhlich (2003) und Hellmann/Müller (2003b).

47 Bei der ZIB wird den Gutachtenden mit dem Dank für ihre Beteiligung am *Peer Review*-Verfahren zu einem Manuskript das Gesamtergebnis der Begutachtung sowie die Voten der anderen Gutachtenden mitgeteilt.

so dass eine solche Auszeichnung mittelfristig auch ein Element besonderen wissenschaftlichen Qualifikationsnachweises werden könnte.

(2) *Anonymität der Gutachtenden*

Bei Manuskripten, die im *Peer Review*-Verfahren nicht zur Veröffentlichung angenommen werden, bleibt die Anonymität der Gutachtenden ein wesentliches Merkmal substanzialer und kritischer Begutachtung. Werden Manuskripte jedoch publiziert, ist die anhaltende Anonymität der Gutachtenden weit weniger bedeutsam und könnte gelockert werden:

- Gutachtende könnten prinzipiell gefragt werden, ob sie im Falle einer insgesamt positiven Begutachtung und der anstehenden Veröffentlichung eines Manuskripts bereit sind, dann auf ihre Anonymität zu verzichten. Damit könnten die entsprechenden Gutachten mit der Publikation der Beiträge öffentlich zugänglich gemacht werden, was auch den Kommunikationsprozess zwischen den AutorInnen und ihren GutachterInnen befruchten könnte sowie zusätzliche Anreize zur Qualitätssteigerung der Gutachten bereitstellte.
- Bei Zustimmung ihrer VerfasserInnen könnten Gutachten auf der Homepage einer Zeitschrift veröffentlicht werden, um für entsprechende inhaltliche Bezugnahmen und Kontroversen mit den AutorInnen auch zitierfähig zu sein.
- Zugleich würde ermöglicht, dass sich der mit dem Gutachten begonnene Kommunikationsprozess fortsetzt und AutorInnen, die von einem hilfreichen Gutachten für die Endfassung ihres Beitrags sehr profitiert haben, sich auch namentlich bei den Gutachtenden bedanken könnten.
- Sowohl im Hinblick auf die Steigerung des Qualitätsniveaus von Gutachten als auch für individuelle Leistungsprofile könnten transparente Nachweise erbracht werden für besonders qualifizierte Gutachtenden-Tätigkeiten.

(3) *Das Verschwinden nicht-publizierter Manuskripte und der dazu erstellten Gutachten*

Resultiert aus einem *Review*-Verfahren aufgrund seiner oben beschriebenen Kontingenzen die Nicht-Veröffentlichung eines Manuskripts, bleibt es der – zugegebenermaßen manchmal recht zwiespältige – exklusive Gewinn von HerausgeberInnen und ausgewählten Gutachtenden, sich damit auseinanderzusetzen zu können. Bevor Manuskripte, die zur Überarbeitung und Wiedereinreichung zurückgegeben werden, auf ewig verschwinden, samt der zu ihren Stärken und Schwächen erstellten Gutachten,⁴⁸ weil der Überarbeitung unüberwindliche Hindernisse entgegenstanden, könnten sie bei Interesse der AutorInnen und Zustimmung der Gutachtenden über die Homepage einer Zeitschrift zugänglich gemacht werden. Dies diente nicht nur einer vergrößerten Transparenz des *Peer Review*-Verfahrens, sondern auch der Möglichkeit, dass gutachterliche Kommunikationsangebote, statt allein von AutorInnen angenommen oder abgelehnt zu werden, von vielen genutzt werden könnten.

48 Die Argumente für die Veröffentlichung abgelehnter Forschungsanträge finden sich bei Böhme (2000).

Literatur

- Berger, Johannes/Handl, Johann/Hopf, Christel/Hirschauer, Stefan/Streeck, Wolfgang 1997: Editorial, in: Zeitschrift für Soziologie 26: 1, 1-2.
- Böhme, Gernot 2000: Was ich nicht erforschen durfte, in: Gegenworte: Zeitschrift für den Disput über Wissen 5, 40-44.
- Daniel, Hans-Dieter 1993: *Guardians of Science: Fairness and Reliability of Peer Review*, Weinheim.
- Ernst, Sylke/Warwas, Jasmin 2003: Peer Review – Qualitätssicherung bei der Evaluation wissenschaftlicher Leistungen, in: <http://Inhf.gwdg.de/ciwm/dokumente/peerreview.htm>; 20.8.2004.
- Europäische Kommission 2001: Wissenschaftspolitik in der Europäischen Union. Förderung herausragender wissenschaftlicher Leistungen durch Gender Mainstreaming. Bericht der ETAN-Expertinnenarbeitsgruppe »Frauen und Wissenschaft«, Luxemburg.
- Felt, Ulrike 1999: Evaluation im wissenschaftspolitischen Kontext, in: Röbbecke, Martina/Simon, Dagmar (Hrsg.): *Qualitätsförderung durch Evaluation? Ziele, Aufgaben und Verfahren von Forschungsbewertung im Wandel* (WZB Discussion Paper P99-003), Berlin, 11-30.
- Felt, Ulrike/Nowotny, Helga/Taschwer, Klaus 1995: *Wissenschaftsforschung. Eine Einführung*, Frankfurt a. M.
- Finetti, Marco 2000: Die überforderten Türhüter. Auf der Suche nach der fehlenden Zeit, in: Gegenworte: Zeitschrift für den Disput über Wissen 5, 32-35.
- Fleck, Ludwik 1980: Entstehung und Entwicklung einer wissenschaftlichen Tatsache, Frankfurt a. M.
- Frey, Bruno S. 2003: Publishing as Prostitution? – Choosing Between One's Own Ideas and Academic Success, in: *Public Choice* 116, 205-223.
- Fröhlich, Gerhard 2003: Anonyme Kritik. Peer Review auf dem Prüfstand der Wissenschaftsforschung, in: medizin – bibliothek – information 3: 2, 33-39.
- Giles, Micheal W./Patterson, David/Mizell, Francie 1989: Discretion in Editorial Decision-Making: The Case of the Journal of Politics, in: *PS – Political Science and Politics* 22: 1, 58-62.
- Hagemann-White, Carol/Schultz, Dagmar 1986: Die Arbeitssituation von Frauen und Männern im Hochschuldienst aus der Sicht der Betroffenen, in: Clemens, Bärbel/Metz-Göckel, Sigrid/Neusel, Ayla/Port, Barbara (Hrsg.): *Töchter der Alma Mater. Frauen in der Berufs- und Hochschulforschung*, Frankfurt a. M., 99-110.
- Hargens, Lowell L./Herting, Jerald R. 1990: A New Approach to Referees' Assessments of Manuscripts, in: *Social Science Research* 19: 1, 1-16.
- Harnad, Stevan 2000: The Invisible Hand of Peer Review, in: *Exploit Interactive* 5, <http://www.exploit-lib.org/issue5/peer-review/>; 24.8.2004.
- Heintz, Bettina/Kohli, Martin/Münch, Richard/Preisendörfer, Peter/Tyrell, Hartmann 2004: Editorial 2004, in: Zeitschrift für Soziologie 33: 1, 2-4.
- Hellmann, Gunther/Müller, Harald 2002: Editorial, in: Zeitschrift für Internationale Beziehungen 9: 1, 3-8.
- Hellmann, Gunther/Müller, Harald 2003a: Editorial, in: Zeitschrift für Internationale Beziehungen 10: 1, 3-6.
- Hellmann, Gunther/Müller, Harald 2003b: Editing (I)nternational (R)elations: A Changing World, in: *Journal of International Relations and Development* 6: 4, 372-389.
- Héritier, Adrienne 1994: Editorial, in: *Politische Vierteljahresschrift* 35: 1, 1-2.
- Hirschauer, Stefan 2004: Peer Review Verfahren auf dem Prüfstand. Zum Soziologiedefizit der Wissenschaftsevaluation, in: Zeitschrift für Soziologie 33: 1, 62-83.
- Kieser, Alfred 2003: Forschung vom Fließband. Evaluation schafft erst die Wirklichkeit, die zu bewerten sie vorgibt, in: *Die Zeit* 30, 17. Juli 2003, 31.
- Kornhuber, Hans Helmut 1988: Mehr Forschungseffizienz durch objektivere Beurteilung von Forschungsleistungen, in: Daniel, Hans-Dieter/Fisch, Rudolf (Hrsg.): *Evaluation von Forschung: Methoden, Ergebnisse, Stellungnahmen*, Konstanz, 361-382.

- Krais, Beate (Hrsg.) 2000a: Wissenschaftsforschung und Geschlechterordnung. Über die verborgenen Mechanismen männlicher Dominanz in der akademischen Welt, Frankfurt a. M.
- Krais, Beate 2000b: Einleitung: Die Wissenschaft und die Frauen, in: Krais 2000a, 9-29.
- Krais, Beate 2000c: Das soziale Feld Wissenschaft und die Geschlechterverhältnisse. Theoretische Sondierungen, in: Krais 2000a, 31-54.
- Luhmann, Niklas 1994: Die Wissenschaft der Gesellschaft, 2. Auflage, Frankfurt a. M.
- Marsh, Herbert W./Ball, Samuel 1989: The Peer Review Process Used to Evaluate Manuscripts Submitted to Academic Journals: Interjudgmental Reliability, in: Journal of Experimental Education 57: 2, 151-169.
- Meier, Kenneth J. 1997: Reforming the Review Process: Right Problem, Wrong Solution, in: PS: Political Science and Politics 30: 3, 561-563.
- Mittelstraß, Jürgen 1989: Wissenschaft als Kultur, in: Mittelstraß, Jürgen: Der Flug der Eule: Von der Vernunft der Wissenschaft und der Aufgabe der Philosophie, Frankfurt a. M., 13-42.
- Mittelstraß, Jürgen 2000: Exzellenz und Mittelmaß, in: Gegenworte: Zeitschrift für den Disput über Wissen 5, 22-25.
- Neidhardt, Friedhelm 2000: Über die Kunst der Selbstverteidigung, in: Gegenworte: Zeitschrift für den Disput über Wissen 5, 26-29.
- Norris, Pippa/Crewe, Ivor 1993: The Reputation of Political Science Journals: Pluralist and Consensus Views, in: Political Studies 41: 1, 5-23.
- Nowotny, Helga/Scott, Peter/Gibbons, Michael 2004: Wissenschaft neu denken. Wissen und Öffentlichkeit in einem Zeitalter der Ungewissheit, Weilerswist.
- Plümper, Thomas 2003: Publikationsaktivität und Rezeptionserfolg der deutschen Politikwissenschaft in internationalen Fachzeitschriften, 1990–2002. Eine bibliometrische Analyse der Veröffentlichungsleistung deutscher politikwissenschaftlicher Fachbereiche und Institute, in: Politische Vierteljahresschrift 44: 4, 529-544.
- POST (Parliamentary Office of Science and Technology) 2002: Peer Review, in: postnote 182, <http://www.parliament.uk/post/pn182.pdf>; 26.9.2004.
- Prowiss 2004: »Peer Review«, in: Glossar von PRO-WISSL: Rahmenprogramm für die professionelle Unterstützung der wissenschaftlichen Laufbahn von Frauen der Universität Zürich, <http://www.prowiss.unizh.ch/infos/glossar/glossar.html>; 24.8.2004.
- Röbbecke, Martina/Simon, Dagmar 2001: Reflexive Evaluation. Ziele, Verfahren und Instrumente der Bewertung von Forschungsinstituten, Berlin.
- Russett, Bruce 1978: Editorial Comment: On Evaluation Procedures, in: Journal of Conflict Resolution 22: 1, 3-6.
- Seibel, Wolfgang 1998: Editorial, in: Politische Vierteljahresschrift 39: 1, 1.
- Weingart, Peter 2001: Die Stunde der Wahrheit? Zum Verhältnis der Wissenschaft zu Politik, Wirtschaft und Medien in der Wissensgesellschaft, Weilerswist.
- Weingart, Peter 2004: Die Stunde der Wahrheit für die Wissenschaft, in: Das Parlament 1/2, 12.1.2004, 13.
- Weller, Christoph 2002: Die massenmediale Konstruktion der Terroranschläge am 11. September 2001. Eine Analyse der Fernsehberichterstattung und ihre theoretische Grundlage (INEF-Report 63), Duisburg, <http://inef.uni-duisburg.de/page/documents/report63.pdf>; 26.9.2004.
- Wennerås, Christine/Wold, Agnes 1997: Nepotism and Sexism in Peer Review, in: Nature 387: 22, 341-343.
- Wennerås, Christine/Wold, Agnes 2000: Vetternwirtschaft und Sexismus im Gutachterwesen, in: Krais 2000a, 107-120.
- Wolff, Stephan 2000: Freud und Leid eines Gutachters, in: Gegenworte: Zeitschrift für den Disput über Wissen 5, 36-39.