

Christine Wennerås und Agnes Wold

Vetternwirtschaft und Sexismus im Gutachterwesen*

Weltweit brechen Frauen ihre akademischen Karrieren immer noch weitaus häufiger ab als Männer (vgl. WIDNALL 1988).¹ In Schweden zum Beispiel werden 44% der Promotionen in der Biomedizin von Frauen abgelegt, während Postdoc-Stellen nur zu 25% und Professorenstellen sogar nur zu 7% mit Frauen besetzt sind. Sobald es genug Frauen auf der Eingangsstufe gebe, so hieß es, werde automatisch auch die Männerherrschaft in den oberen Rängen der akademischen Forschung schwinden. Aber nichts dergleichen geschah, und immer noch sind die höheren akademischen Positionen in der Biomedizin trotz des seit den 1970er Jahren bedeutenden Zustroms von Frauen überproportional häufig mit Männern besetzt.

1. Gründe für den geringen Erfolg

Warum haben Frauen solche Schwierigkeiten? Manche meinen, Frauen seien in der Regel weniger motiviert und karriereorientiert als Männer und würden sich daher weniger zielstrebig um Stellen und Stipendien bewerben. Andere sind der Ansicht, Frauen seien weniger produktiv als Männer und ihre Arbeit habe deshalb geringeren wissenschaftlichen Wert. Und wieder andere sehen die Frauen als Opfer einer geschlechtsbedingten Diskriminierung. Wir woll-

ten nun untersuchen, ob im Gutachtersystem des schwedischen *Medical Research Council* (MRC), einer der wichtigsten Institution für die Forschungsförderung in der Biomedizin in Schweden, Frauen und Männer auf gleicher Basis evaluiert werden. Auslöser für unsere Untersuchung war, daß die Erfolgsquote von Wissenschaftlerinnen, die sich in den 1990er Jahren beim MRC um ein Habilitationsstipendium beworben hatten, weniger als halb so groß war wie die der männlichen Bewerber.

Unsere Untersuchung deutet entschieden darauf hin, daß Gutachter wissenschaftliche Leistung nicht unabhängig vom Geschlecht beurteilen können. Wie aus den multiplen Regressionsanalysen hervorgeht, mit denen wir das Verhältnis zwischen den bestehenden Parametern für die wissenschaftliche Produktivität und den Kompetenzwerten untersuchten, überbewerteten die Gutachter die Leistungen der Männer und/oder unterbewerteten die Leistungen der Frauen.

Im Gutachtersystem des schwedischen MRC reichen die Bewerberinnen und Bewerber einen Lebenslauf, eine Bibliographie und einen Entwurf für ein Forschungsprojekt ein. Ihre Bewerbung wird von einer der elf Evaluierungskommissionen begutachtet, die jeweils für ein bestimmtes Forschungsgebiet zuständig sind. Jede/r Bewerber/in wird von den fünf Gutachtern der Kommission beurteilt, der er oder sie zugewiesen wurde. Jeder Gutachter gibt der Bewerberin/dem Bewerber einen Punktwert zwischen 0 und 4 für die drei folgenden Parameter: wissenschaftliche Kompetenz, Relevanz des Forschungsprojekts und Qualität der vorgeschlagenen Methodologie. Dann werden die drei von dem Gutachter vergebenen Punktwerte miteinander multipliziert; das Ergebnis ist ein Gesamtwert, der zwischen 0 und 64 liegen kann. Zuletzt wird der Durchschnitt aus den fünf von einer Bewerberin oder einem Bewerber erzielten Gesamtwerten berechnet. Dieser Durchschnitt ergibt den Wert, aufgrund dessen innerhalb jeder Kommission eine Rangfolge unter den Bewerberinnen und Bewerbern festgelegt wird. Zum Abschluß des Verfahrens beschließt dann der MRC-Vorstand, dem die Vorsitzenden der elf Kommissionen angehören, an wen die Stipendien vergeben werden. Gewöhnlich wählt jede

Kommission unter den Bewerbern mit der höchsten Bewertung maximal drei Personen aus.

1995 wurden 20 Habilitationsstipendien angeboten, um die sich 114 Personen bewarben, davon 62 Männer und 52 Frauen. Ihr Durchschnittsalter betrug 36 Jahre, und alle hatten in den fünf davorliegenden Jahren promoviert. Die meisten Frauen hatten zunächst ein Studium in den Naturwissenschaften abgeschlossen (62%), die übrigen in der Medizin (27%) oder in der Krankenpflege (12%); bei den Männern lagen die entsprechenden Zahlen bei 38%, 59% und 3%.

Beurteilungen von Gutachtern sind der Öffentlichkeit traditionell nicht zugänglich, und so verweigerte uns der MRC zunächst auch die Einsicht in die Unterlagen, die mit der Evaluierung der Bewerberinnen und Bewerber zu tun hatten. In Schweden gibt es jedoch ein Pressegesetz, das natürlichen Personen Zugang zu allen Dokumenten von staatlichen oder kommunalen Stellen gewährt. Ausgenommen sind lediglich Dokumente, die unter die Geheimhaltungspflicht fallen, also solche, in denen die Sicherheit des Staates, die auswärtigen Beziehungen oder die Persönlichkeitsrechte eines Staatsbürgers berührt sind. Auf dieser rechtlichen Grundlage legten wir gegen die Weigerung des MRC, uns die Beurteilungen zugänglich zu machen, Einspruch ein. 1995 entschied das Oberste Verwaltungsgericht, die Beurteilungen des MRC seien amtliche Unterlagen. Damit wurden – zum ersten Mal, so weit wir wissen – eine große Kohorte von Bewerberinnen und Bewerber betreffende Original-Unterlagen über die Gutachtertätigkeit der wissenschaftlichen Untersuchung zugänglich gemacht.

Wir stellten fest, daß die von den Gutachtern des MRC vergebenen Werte für alle drei Evaluierungsparameter bei Frauen im Durchschnitt niedriger waren als bei Männern: 0.25 Punkte niedriger bei der wissenschaftlichen Kompetenz (2.21 gegenüber 2.46 Punkten); 0.17 Punkte niedriger bei der Qualität der vorgeschlagenen Methodologie (2.37 gegenüber 2.54 Punkten); und 0.13 Punkte niedriger bei der Relevanz des Forschungsprojekts (2.49 gegenüber 2.62 Punkten). Da diese Werte miteinander multipliziert werden, hatten die Frauen erheblich schlechtere Endwerte als die

Männer (im Durchschnitt 13.8 gegenüber 17.0 Punkten). In diesem Jahr wurden die Forschungsstipendien an vier Frauen und 16 Männer vergeben.

Wie diese Zahlen zeigen, ließ nach Ansicht der Gutachter vor allem die wissenschaftliche Kompetenz der Bewerberinnen zu wünschen übrig. Da dieser Parameter üblicherweise von Zahl und Qualität der wissenschaftlichen Veröffentlichungen abhängt (vgl. COLE et al. 1981; LONG 1992; SONNERT 1995; SONNERT/HALTON 1996), schien die Annahme vernünftig, daß Frauen weniger Punkte als Männer bekamen, weil sie weniger produktiv waren. Wir untersuchten diese Hypothese, indem wir die wissenschaftliche Produktivität aller 114 Bewerber ermittelten und dann die Werte von Bewerberinnen und Bewerbern mit gleicher wissenschaftlicher Produktivität gruppenweise verglichen.

2. Maße für Produktivität

Wir maßen die wissenschaftliche Produktivität der einzelnen Bewerberinnen und Bewerber auf sechs verschiedene Weisen. Erstens bestimmten wir die Gesamtzahl der eigenen wissenschaftlichen Veröffentlichungen, zweitens die Zahl der Veröffentlichungen, in denen die Bewerberin oder der Bewerber als erster Autor genannt wurde. Beide Zahlen wurden den Bibliographien der Bewerber entnommen und anhand der *Medline*-Datenbank überprüft. (Wir nennen diese Maße »Gesamtveröffentlichungen« und »Erstautorveröffentlichungen«).

Um der Tatsache Rechnung zu tragen, daß das Prestige von biomedizinischen Zeitschriften höchst unterschiedlich ist, konstruierten wir Maße auf der Grundlage des Wirkungsfaktors dieser Zeitschriften. Der Wirkungsfaktor einer wissenschaftlichen Zeitschrift ist den *Journal Citation Reports* zu entnehmen, einer von dem unabhängigen *Institute of Scientific Information* aufgestellten Liste, und gibt an, wie oft innerhalb eines Jahres ein in einer bestimmten Zeitschrift veröffentlichter Beitrag durchschnittlich zitiert wird. Für

unser drittes Maß addierten wir die Wirkungsfaktoren aller Zeitschriften, in denen Beiträge der Bewerberin/des Bewerbers erschienen waren, und erhielten damit ein Maß für die »Gesamtwirkung« sämtlicher Veröffentlichungen der betreffenden Person.

Viertens entwickelten wir ein Maß für die »Erstautorwirkung«, indem wir die Wirkungsfaktoren der Zeitschriften addierten, in denen die Beiträge erschienen waren, für die der Bewerber/die Bewerberin als erster Autor genannt wurde. Die Maßeinheit für die Gesamtwirkung und für die Erstautorwirkung sind »Wirkungspunkte«, wobei ein Wirkungspunkt einem in einer Zeitschrift mit dem Wirkungsfaktor 1 veröffentlichten Beitrag entspricht. Fünftens ermittelten wir anhand der *Science Citation Database*, wie oft wissenschaftliche Beiträge der Bewerberin/des Bewerbers im Jahre 1994 zitiert worden waren, und erhielten damit das Maß »Gesamtzitate«. Und sechstens wiederholten wir diesen Vorgang für die Beiträge, bei denen die Bewerberin/der Bewerber der erste Autor war, was das Maß »Erstautorzitate« ergab.

Bekamen nun Männer und Frauen mit gleicher wissenschaftlicher Produktivität von den MRC-Gutachtern dieselben Werte für wissenschaftliche Kompetenz? Nein! Nach unserer Untersuchung erhielten Frauen bei gleichem Niveau wissenschaftlicher Produktivität weniger Punkte als Männer. Ja, die einzige Gruppe von Frauen, die als ebenso kompetent beurteilt wurde wie Männer, wenn auch nur als ebenso kompetent wie die am wenigsten produktive Gruppe der männlichen Bewerber (Männer mit weniger als 20 Punkten Gesamtwirkung), war die produktivste Gruppe der Frauen, nämlich die Bewerberinnen mit 100 und mehr Punkten Gesamtwirkung.

3. Warum Frauen schlechtere Bewertungen erhalten

Obwohl die unterschiedliche Bewertung von Männern und Frauen mit gleicher wissenschaftlicher Produktivität darauf hindeutete, daß hier tatsächlich eine Diskriminierung von Wissenschaftle-

rinnen vorlag, hätten grundsätzlich auch andere Faktoren als das Geschlecht für das schlechtere Abschneiden der Frauen verantwortlich sein können. Wenn die Frauen zum Beispiel hauptsächlich auf Gebieten forschten, die beim MRC eine niedrige Priorität haben, oder wenn sie von weniger renommierten Universitäten kämen oder seltener in Arbeitszusammenhängen mit den Entscheidungsträgern im Wissenschaftsbereich stehen würden, so könnten ihre schlechteren Bewertungen auch auf diese Faktoren zurückzuführen sein statt auf ihr Geschlecht an sich.

Um die Ursache für die schlechteren Bewertungen der Frauen zu bestimmen, führten wir eine multiple Regressionsanalyse durch. Mit dieser lassen sich die Faktoren ermitteln, die wesentlichen Einfluß auf ein bestimmtes Ergebnis (zum Beispiel die Punktwerte für wissenschaftliche Kompetenz) haben, und die Größe dieses Einflusses. Außerdem können mit ihrer Hilfe Faktoren eliminiert werden, deren Einfluß auf ein bestimmtes Ergebnis nur Ausdruck ihrer Abhängigkeit von anderen Faktoren ist.

Bei dieser multiplen Regressionsanalyse nahmen wir an, daß zwischen den Werten für Kompetenz, die die Bewerberinnen und Bewerber erhalten hatten, und ihrer wissenschaftlichen Produktivität eine lineare Beziehung besteht. Wir konstruierten sechs verschiedene Regressionsmodelle, nämlich ein Modell für jede der oben beschriebenen Produktivitätsvariablen. Bei jedem Modell bestimmten wir den Einfluß der folgenden Faktoren auf die Kompetenzwerte: Geschlecht der sich bewerbenden Person; Nationalität (schwedisch/nicht schwedisch); Studienrichtung (Medizin, Naturwissenschaften oder Krankenpflege); Forschungsgebiet; Universitätszugehörigkeit; die für die betreffende Person zuständige Evaluierungskommission; Forschungserfahrung im Ausland nach der Promotion; Empfehlungsschreiben als Bestandteil der Bewerbung; Arbeitszusammenhang der Bewerberin/des Bewerbers mit einem Kommissionsmitglied. Diese letzte Information wird im MRC-Evaluierungsprotokoll festgehalten. Ist ein Arbeitszusammenhang gegeben, darf dieser Gutachter nicht an der Bewertung der betreffenden Person teilnehmen. Ein solcher Zusammenhang war bei Frauen (12%) ebenso häufig wie bei Männern (13%).

Tabelle 1 auf der folgenden Seite zeigt das Ergebnis der multiplen Regressionsanalyse. Drei von sechs Produktivitätsvariablen ergaben statistisch signifikante Modelle, die eine Vorhersage der den Bewerberinnen und Bewerbern zugesprochenen Kompetenzwerte erlaubten: Gesamtwirkung, Erstautorwirkung und Erstautorzitate. Das auf der Gesamtwirkung basierende Modell ergab den deutlichsten Erklärungszusammenhang ($r^2 = 0.47$). Bei allen drei Modellen fanden wir neben der wissenschaftlichen Produktivität zwei weitere Faktoren, die einen signifikanten Einfluß auf die Kompetenzwerte hatten: das Geschlecht der sich bewerbenden Person und ihr Arbeitszusammenhang mit einem Kommissionsmitglied.

Nach dem auf der Gesamtwirkung basierenden Regressionsmodell begannen die Frauen mit einem Kompetenzniveau von 2.09 Punkten (Absolutglied der Kurve der multiplen Regressionsanalyse); für jeden der von ihnen akkumulierten Wirkungspunkte erhielten sie von den Gutachtern zusätzlich 0.0033 Punkte für wissenschaftliche Kompetenz. Männliche Bewerber jedoch erhielten unabhängig von der wissenschaftlichen Produktivität weitere 0.21 Punkte auf dieser Variablen. Damit mußte eine Wissenschaftlerin, um denselben Kompetenzwert zu erhalten wie ein männlicher Kollege, dessen Produktivität um 64 Wirkungspunkte übertreffen (95% Konfidenzintervall: 35-95 Wirkungspunkte). Dies entspricht ungefähr drei zusätzlichen Beiträgen in *Nature* oder *Science* (Wirkungsfaktor 25 bzw. 22) oder 20 zusätzlichen Beiträgen in einer Zeitschrift mit dem Wirkungsfaktor 3, was hervorragenden Fachzeitschriften wie *Atherosclerosis*, *Gut*, *Infection and Immunity*, *Neuroscience* oder *Radiology* entspricht. Da die durchschnittliche Gesamtwirkung in dieser Bewerberkohorte bei 40 Punkten lag, hätte eine Bewerberin also 2.5 mal so produktiv sein müssen wie der durchschnittliche Bewerber, um denselben Kompetenzwert zu erhalten wie er ($((40 + 64)/40 = 2.6)$).

Tabelle 1: Faktoren, die die Bewertung der wissenschaftlichen Kompetenz durch die Gutachter signifikant beeinflussen, nach drei Modellen für die multiple Regressionsanalyse

<i>Wissenschaftliche Produktivität</i>			
Multiple Regression, Modell basierend auf:	r^2	Absolutglied	Kompetenzpunkte/Produktivitätseinheit
Gesamtwirkung	0.47	2.09	0.0033 <0.00005*
Erstautorwirkung	0.44	2.13	0.0094 <0.0001
Erstautorzitate	0.41	2.17	0.0054 0.001
<i>Von den Gutachtern vergebene zusätzliche Punkte für folgende Faktoren</i>			
Multiple Regression, Modell basierend auf:	Geschlecht männlich	Arbeitszus. m. Gutachter	Empfehlungsschreiben
Gesamtwirkung	0.21 <0.00005	0.22 0.0008	0.10 0.04
Erstautorwirkung	0.24 <0.00005	0.20 0.005	NS
Erstautorzitate	0.23 <0.00005	0.23 0.001	NS
<i>Größe des Einflusses von nichtwissenschaftlichen Faktoren in Produktivitätsäquivalenten</i>			
Multiple Regression, Modell basierend auf:	Geschlecht männlich	Arbeitszus. m. Gutachter	Maßeinheit
Gesamtwirkung	64 (35-93) ⁺	67 (29-105)	Wirkungspunkte
Erstautorwirkung	25 (14-36)	21 (6-36)	Wirkungspunkte
Erstautorzitate	42 (23-61)	42 (17-67)	Zitierhäufigkeit 1994

* Kursiv gedruckte Zahlen geben die *P*-Werte für die betreffende Variable an.

⁺ Zahlen in Klammern geben 95% Konfidenzintervall an.

NS: nicht statistisch relevant, *P*-Wert >0.05.

4. Der Freundschaftsbonus

Nach demselben Regressionsmodell lagen die Werte für wissenschaftliche Kompetenzscores der Bewerber, die in einem Arbeitszusammenhang mit einem Kommissionsmitglied standen, um 0.22 Punkte über den Kompetenzwerten von Bewerbern gleichen Geschlechts und gleicher wissenschaftlicher Produktivität, aber ohne einen solchen Arbeitszusammenhang (Tabelle 1). Dieser »Bekanntschafsbonus« war 67 Wirkungspunkte wert (Konfidenzintervall: 29 bis 105 Wirkungspunkte). Obwohl also die Regeln des MCR »befangenen« Gutachtern verbieten, sich an der Bewertung von Bewerberinnen und Bewerbern zu beteiligen, mit denen sie in irgendeinem Arbeitszusammenhang stehen, war diese Regel unzureichend, da die »neutralen« Kommissionsmitglieder zum Ausgleich höhere Bewertungen vergaben, wenn sie Bewerber beurteilten, die mit einem ihrer Kollegen in einem Arbeitszusammenhang standen.

Weil der Bekanntschafsbonus die gleiche Größenordnung hatte wie der Bonus »männlich«, konnte eine Frau für ihr Geschlecht (- 0.21 Kompetenzpunkte) einen Ausgleich durch den Arbeitszusammenhang mit einem der Gutachter schaffen (+ 0.22 Kompetenzpunkte). Auf der anderen Seite mußte eine Frau (- 0.21 Kompetenzpunkte) ohne persönliche Beziehungen in der Kommission (- 0.22 Kompetenzpunkte) den MRC-Gutachtern 131 zusätzliche Wirkungspunkte präsentieren können, um denselben Kompetenzwert zu erreichen wie ein männlicher Bewerber, der mit einem der Gutachter in einem Arbeitszusammenhang stand. Ein solches Produktivitätsniveau erreichten nur 3 der 114 Bewerberinnen und Bewerber, ein Mann und zwei Frauen. Damit stellten die Zugehörigkeit zum weiblichen Geschlecht und das Fehlen von persönlichen Arbeitszusammenhängen ein doppeltes Handicap von solcher Schwere dar, daß es durch wissenschaftliche Produktivität allein kaum auszugleichen war.

Die beiden anderen, auf der Erstautorwirkung bzw. den Erstautorzitaten basierenden Regressionsmodelle ergaben, was die Wirkung von Geschlecht und Arbeitszusammenhang angeht, annä-

hernd gleiche Resultate wie das erste Modell (Tabelle 1). Da sowohl die Gesamtwirkung und die Erstautorwirkung mit $r = 0.63$ als auch die Gesamtwirkung und die Erstautorzitate mit $r = 0.62$ nur gering korreliert waren, ist diese Übereinstimmung kein statistisches Artefakt, das auf einen höheren Interrelationsgrad zwischen den drei Produktivitätsvariablen zurückzuführen wäre. Wir sind daher der Ansicht, daß das männliche Geschlecht und der Arbeitszusammenhang mit einem Kommissionsmitglied in den Augen der MRC-Gutachter echte Determinanten der wissenschaftlichen Kompetenz waren.

Nationalität, Ausbildung, Forschungsgebiet oder Forschungserfahrung nach der Promotion hatten bei keinem Modell Einfluß auf die Werte für wissenschaftliche Kompetenz. Bei dem auf der Gesamtwirkung basierenden Modell hatte das Vorhandensein eines Empfehlungsschreibens positive Auswirkungen auf diesen Wert, nicht aber bei den beiden anderen Modellen (Tabelle 1). Durchaus einen Einfluß auf die Kompetenzwerte hatte dagegen die für die Beurteilung des Bewerbers zuständige Evaluierungskommission, denn manche Kommissionen waren bei der Kompetenzevaluierung »strenger« als andere. Da die Stipendien jedoch nicht aufgrund absoluter Punktwerte vergeben wurden, sondern aufgrund der Rangposition, die der Bewerberin oder dem Bewerber innerhalb der zuständigen Kommission gegeben worden war, hatten Bewerberinnen und Bewerber, die von einer »strengen« Kommission begutachtet wurden, die gleichen Chancen wie alle anderen, ein Stipendium zu erhalten.

5. Das System verändern

Die Begutachtung durch *peers*, auch als »Kernstück des modernen wissenschaftlichen Gutachterwesens« (GLANTZ/BERO 1994) bezeichnet, ist aus vielerlei Gründen kritisiert worden, unter anderem wegen der geringen Übereinstimmung zwischen den Bewertungen der verschiedenen Gutachter und wegen der Möglichkeit,

daß Gutachter Projekte favorisieren, die mit ihren eigenen Ansichten übereinstimmen (ERNST et al. 1992). Unsere Untersuchung ist die erste Analyse, der echte Gutachterdaten zugrunde liegen, und sie liefert direkte Anhaltspunkte für die Vetternwirtschaft, die, wie beiläufig bereits vermutet (vgl. FORSDYKE 1993; CALZA/GERBISA 1995; PEREZ- ENRICO 1995), das Gutachterwesen beeinträchtigt.

Nun könnte man argumentieren, daß junge Wissenschaftlerinnen und Wissenschaftler, die in einem Arbeitszusammenhang mit Gutachtern stehen, Teil einer besser ausgebildeten wissenschaftlichen Elite und daher kompetenter als durchschnittliche Bewerberinnen und Bewerber sind. Tatsächlich hatten Bewerberinnen und Bewerber, die in Arbeitszusammenhängen mit Gutachtern standen, im Durchschnitt höhere Werte bei der Gesamtwirkung als Bewerber ohne solche Beziehungen. Damit profitierten Bewerberinnen und Bewerber mit persönlichen Beziehungen aufgrund ihrer höheren wissenschaftlichen Produktivität zu Recht von höheren Werten bei der wissenschaftlichen Kompetenz. Sie erhielten jedoch obendrein zusätzliche Kompetenzpunkte, die nicht durch die wissenschaftliche Produktivität gerechtfertigt waren. Wir sehen keinen Grund dafür, warum Bewerberinnen und Bewerber, die trotz fehlender Anbindung an eine hochangesehene Forschergruppe zu qualitativ hochwertiger Forschung imstande sind, nicht in gleicher Weise belohnt werden sollten.

Verschiedene Studien haben gezeigt, daß Frauen wie Männer die Qualität der Arbeit von Männern höher bewerten als die Qualität der Arbeit von Frauen, wenn ihnen das Geschlecht der zu evaluierenden Person bekannt ist, nicht aber, wenn sie zwar dieselbe Person bewerten, aber deren Geschlecht nicht kennen (vgl. etwa NIEVA/GUTER 1980). Es ist einigermaßen überraschend, daß die Ergebnisse dieser Studien die *scientific community* nicht davon abhalten, sich weiterhin auf Evaluierungssysteme zu verlassen, die anfällig für Gutachtervorurteile sind.

Interessant ist auch eine andere Frage, der wir hier allerdings nicht nachgehen konnten, nämlich ob die strengere Evaluierung von Wissenschaftlerinnen nicht auf die geringe Zahl der Frauen unter den Gutachtern zurückzuführen ist. Eben diese geringe Zahl

aber (5 von 55) sowie die ungleichmäßige Verteilung der Frauen auf die MRC-Kommissionen machten eine statistische Analyse ihres Bewertungsverhaltens unmöglich. Andere Studien lassen allerdings vermuten, daß weibliche Gutachter bei der Bewertung der Leistung von Frauen objektiver sein könnten als ihre männlichen Kollegen (vgl. FRIEZE 1975). Wir sind jedoch nicht davon überzeugt, daß das Problem der geschlechtsbedingten Diskriminierung durch die bloße Erhöhung des Frauenanteils unter den Gutachtern zu lösen wäre.

Wenn eine geschlechtsbedingte Diskriminierung in der von uns beobachteten Größenordnung auch im Gutachtersystem anderer Forschungsräte und Forschungsförderungsorganisationen sowie in anderen Ländern als Schweden wirksam wäre, könnte dies allein schon genügen, um die im Vergleich zu Männern geringere Erfolgsquote von Frauen beim Erreichen der höheren akademischen Ränge zu erklären. Die *Vereinten Nationen* haben vor kurzem Schweden als das im Hinblick auf die Chancengleichheit von Männern und Frauen weltweit führende Land bezeichnet, so daß die Annahme, daß geschlechtsbedingte Diskriminierung auch anderswo auftreten dürfte, nicht allzuweit hergeholt ist. Es wäre deshalb überaus wichtig, Studien wie diese auch in anderen Ländern und für andere Forschungsgebiete durchzuführen.

Eine gründliche Analyse anderer Gutachtersysteme ist nur dann zu erreichen, wenn die Politik der Geheimhaltung aufgegeben wird. Wir selber konnten unsere Studie nur dank dem schwedischen Pressegesetz durchführen. Es wird oft behauptet, die Geheimhaltung von Gutachten schütze die Gutachter vor unstatthafter Beeinflussung. Unsere Ergebnisse lassen jedoch Zweifel an dieser Behauptung aufkommen. Weiter heißt es, die Rekrutierung von hoch qualifizierten Gutachtern würde behindert, wenn ihnen keine Anonymität zugesichert werden könne. Solche Befürchtungen scheinen übertrieben, denn obwohl die Evaluierungsergebnisse von Gutachtern seit dem Gerichtsurteil von 1995 jedermann zugänglich sind, ist es nicht zum allgemeinen Auszug der Gutachter aus den Auswahlkommissionen gekommen.

Viel wichtiger ist, daß in den Augen der Öffentlichkeit die

Glaubwürdigkeit des akademischen Systems untergraben wird, wenn es nicht zuläßt, daß sein eigenes Evaluierungssystem wissenschaftlich evaluiert wird. Wir sind der festen Überzeugung, daß Wissenschaftlerinnen und Wissenschaftler am besten in der Lage sind, wissenschaftliche Leistungen zu evaluieren. Man muß jedoch auch sehen, daß diese Wissenschaftlerinnen und Wissenschaftler nicht weniger anfällig für die Wirkungen von Vorurteil und Kameraderie sind als andere Menschen. Die Entwicklung von Gutachtersystemen mit einer gewissen eingebauten Resistenz gegenüber den Schwächen der menschlichen Natur hat deshalb hohe Priorität. Erfolgt sie nicht, werden große, vielversprechende Begabungsreserven ungenutzt bleiben.²

Anmerkungen

- * Dieser Beitrag wurde zuerst veröffentlicht unter dem Titel »Nepotism and sexism in peer-review« in: *Nature* 387, Mai 1997, S. 341–343. Die Übersetzung aus dem Englischen besorgte Hella Beister.
- 1 Wir danken der Dokumentarin Ann-Marie Holst für ihre Unterstützung und Maria Wold-Troell, Svante Wold und Christer Andersson für Beratung in statistischen Fragen. Die Studie wurde mit einem Stipendium des schwedischen Bildungsministeriums gefördert.
- 2 Nachtrag von *Nature*: Dieser Beitrag wurde von drei Männern begutachtet.

Literatur

- CALZA, Laura/GERBISA, Spiridione (1995): Italian Professorships. In: *Nature* 374, S. 492
- COLE, Stephen/COLE, Jonathan R./SIMON, Gary A. (1981): Chances and Consensus in Peer Review. In: *Science* 214, S. 881–886
- ERNST, E./RESCH, K. L./UHER, E. M. (1992): Reviewer Bias. In: *Annals of Internal Medicine* 116, S. 958
- FORSDYKE, D. R. (1993): On Giraffes and Peer Review. In: *Journal of the Federation of American Societies for Experimental Biology* 7, S. 619–621

- FRIEZE, Irene Hanson (1975): Women's Expectations for and Casual Attributions of Success and Failure. In: MEDNICK, Martha Tamara Shuch/TANGRI, Sandra Schwartz/HOFFMAN, Lois Wladis (eds.): *Women and Achievement: Social and Motivational Analyses*. Washington D.C.: Hemisphere, S. 158-171
- GLANTZ, S. A./BERO, L. A. (1994): Inappropriate and Appropriate Selection of Peers in Grant Review. In: *Journal of the American Medical Association* 272, S. 114-116
- LONG, J. Scott (1992): Measures of Sex Differences in Scientific Productivity. In: *Social Forces* 71, S. 159-178
- NIEVA, V. F./GUTER, B. A. (1980), in: *Academic Management Review* 5, S. 267-276
- PEREZ-ENCISO, Miguel (1995): Spanish Practices. In: *Nature* 378, S. 760
- SONNERT, Gerhard (1995): What Makes a Good Scientist - Determinants of Peer Evaluation Among Biologists. In: *Social Studies of Science* 25, S. 35-55
- SONNERT, Gerhard/HOLTON, Gerald (1996): Career Patterns of Women and Men in the Sciences. In: *American Scientist* 84, S. 63-71
- WIDNALL, Sheila E. (1988): AAAS Presidential Lecture: Voices from the Pipeline. In: *Science* 241, S. 1740-1745