

Disagreement in Peer Review

von Anne K. Krüger

Workshop an der Humboldt-Universität zu Berlin,
19.-20. Mai 2016

Dass sich über Geschmack nicht streiten lässt, ist bekannt. Doch gilt das auch für die Wissenschaft? Meinungsverschiedenheiten sind ein alltägliches Phänomen, und auch in der Wissenschaft ist der Dissens die Triebfeder forscherschen Handelns. Es stellt sich jedoch die Frage, wie solche Differenzen mit dem Anspruch einer vermeintlich objektiv bewertbaren wissenschaftlichen Qualität zu vereinbaren sind. Welche Konsequenzen haben Meinungsverschiedenheiten im Prozess des Peer Reviews, das es doch ermöglichen soll, ein eindeutiges wissenschaftliches Qualitätsurteil zu treffen? Diese Fragen bildeten den Ausgangspunkt für den interdisziplinären Workshop „Disagreement in Peer Review“, der von Geert Keil (HU Berlin - Lehrbereich philosophische Anthropologie), Martin Reinhart (HU Berlin - Lehrbereich Wissenschaftsforschung, Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW)) und Ralf Poscher (Freiburg - Institut für Staatswissenschaft und Rechtsphilosophie) organisiert wurde.

GEERT KEIL (Berlin) formulierte einleitend die Frage, wie es überhaupt möglich sei, dass zwei ExpertInnen desselben Fachgebiets auf der Grundlage derselben Informationen zu unterschiedlichen Bewertungen wissenschaftlicher Erkenntnis kommen können, wenn Wissenschaft doch als Suche nach objektiver Wahrheit und der Weg dorthin durch Rationalität gekennzeichnet seien. Als zentrales Problem für eine philosophische Auseinandersetzung mit dem Peer Review benannte er vor diesem Hintergrund die bei der Suche nach wissenschaftlicher Wahrheit gegebene Notwendigkeit, trotz einer solchen „tiefen Meinungsverschiedenheit“¹ eine übereinstimmende Entscheidung über wissenschaftliche Qualität treffen zu müssen.

In Zeiten steigender Drittmittelraten und wachsenden Publikationsdrucks ist die Frage nach der Beurteilung von wissenschaftlicher Qualität besonders relevant geworden. Denn Peer Review ist derzeit das Nadelöhr, das nicht nur über wissenschaftlichen Erkenntnisfortschritt, sondern über ganze Wissenschaftskarrieren entscheidet. Doch verwundert es gerade vor diesem Hintergrund, dass diese Form der Begutachtung bislang eine *black box* geblieben ist. Jenseits von konkreten Verfahrensregeln ist wenig bekannt über die sozialen Prozesse, die zur Annahme bzw. Ablehnung von Forschungsanträgen oder von wissenschaftlichen Publikationen führen. Diese Erkenntnis hat dazu geführt, dass in

der soziologischen Wissenschaftsforschung zunehmend einerseits die formalen Prozesse und andererseits deren praktische Umsetzung in den Fokus rücken. STEFAN HORNBOSTEL (Berlin), MARTIN REINHART (Berlin), STEFAN HIRSCHAUER (Mainz) und MICHÈLE LAMONT (Harvard) haben dazu bereits einschlägige Forschung vorgelegt², deren Weiterführung sie im Rahmen des Workshops präsentierten.

Hornbostel konstatierte, Peer Review habe sich zu einer Institution der wissenschaftlichen Qualitätssicherung entwickelt und bis heute erhalten, obwohl es immer schon breiter Kritik ausgesetzt gewesen sei. Dies habe nicht zuletzt damit zu tun, dass Peer Review einerseits als „part of the job“ und andererseits immer auch als Chance, Lobbyismus für die eigene Forschungsrichtung zu betreiben, betrachtet werde. Die Herausforderung bestehe deshalb vor allem in der Auswahl der richtigen GutachterInnen. Jedoch stelle die gestiegene Nachfrage nach Gutachten und die sich gleichzeitig durch andere Anforderungen wie Drittmittelwerbung und Publikationstätigkeit verknappende Zeit ein wesentliches Problem für das Peer Review und die daraus folgende wissenschaftliche Qualität dar.

Reinhart betonte, die Soziologie solle weniger die Ergebnisse von Peer-Review-Verfahren als vielmehr die dahinterstehenden Prozesse ins Visier der Forschung nehmen. Dazu entwickelte er eine Typologie von Faktoren, die Peer-Review-Prozesse beeinflussen. Diese setzte bei der Frage nach der (wechselseitigen) Sichtbarkeit von GutachterInnen innerhalb des Reviewprozesses an, die es ermögliche, grundlegend zwischen extern geschriebenen Reviews wie im Fall von Zeitschriften und interne Diskussionsrunden wie im Fall von Förderanträgen zu unterscheiden. Auch Faktoren wie einer homogenen versus einer diversifizierten Fachkultur rechnete Reinhart einen wesentlichen Einfluss auf den Verlauf des Begutachtungsprozesses zu. Zudem widmete er sich der Frage, welche Rolle Emotionen im Peer-Review-Prozess spielen. Obwohl Wissenschaft als affektiv neutral gedacht werde, sei Peer Review emotional aufgeladen. Insbesondere die Aufrechterhaltung von Vertrauen (confidence) sei die grundlegende Voraussetzung für ein Peer-Review-Verfahren. Denn nur das Vertrauen darauf, dass Peer Review das halte, was es verspreche, ermögliche den Fortbestand dieser Maßnahme der Qualitätssicherung.

Auch Hirschauer verwies darauf, dass es in der Forschung zu Peer Review weniger um die Ergebnisse selbst als vielmehr um den Reviewprozess als „Kommunikationssituation“ gehen müsse, in der die Allokation von Reputation sowohl in der Begutachtung als auch in der Herstellung dieser Begutachtungen vorgenommen werde. Er skizzierte die unterschiedlichen Sequenzen eines Reviewprozesses von der Einreichung bis zum Gutachten, in denen einerseits die Aufmerksamkeit für die Publikation und deren Forschungsgegenstand und andererseits die Meinungsverschiedenheiten zwischen den Begutachtenden sowie die Disziplinierung von AutorInnen durch den Zwang zur Einhaltung von formalen und inhaltlichen Regeln kommunikativ intensiviert werde. Hierbei verdeutlichte er insbesondere die „panoptische Organisation“ des Prozesses, die dadurch gegeben

sei, dass die HerausgeberInnen nicht nur über die eingereichte Publikation, sondern auch über deren Gutachten befinden.

In ihrem Abendvortrag präsentierte Lamont einen weiteren Ansatz, um die „black box peer review“ zu öffnen und nach grundlegenden Prozessen der Urteilsbildung zu fragen. Die Schwierigkeit bestehe jedoch darin, dass die Regeln, Normen und Standards zur Bewertung von Qualität nicht festgeschrieben seien, sondern innerhalb einer Disziplin jeweils erlernt werden. Es sei demnach der Kontext, in dem eine Bewertung stattfindet, der für die Art und Weise der Begutachtung entscheidend sei. Unter Rückgriff auf eine vergleichende Studie zu drei Fällen von Begutachtung (belletristische Buchkritiken, wissenschaftliche Projektförderung und Unternehmensberatung) verdeutlichte Lamont die „customary rules“, die in jedem dieser drei Bereiche die Verfahren lenken und die Begutachtungsobjekte, -agenten und -adressaten in ein charakteristisches Verhältnis zueinander setzen.

Nach diesen Versuchen, mit soziologischem Instrumentarium etwas Licht ins Dunkel zu bringen, folgten am zweiten Tag Einblicke aus der Begutachtungspraxis selbst. ELKE BRENDEL (Bonn, Mitglied im Fachkollegium Philosophie der DFG), THOMAS BALDWIN (York, Chefherausgeber von *Mind* 2005–2015) und MICHÈLE GARFINKEL (Heidelberg, European Organisation of Molecular Biology, EMBO) stellten die Verfahrensregeln dar, anhand derer in ihrem jeweiligen fachlichen Kontext über die finanzielle Förderung von Forschungsprojekten oder die Veröffentlichung von Forschungsergebnissen entschieden wird. Da sie aufgrund ihrer Position Zugang zu Gutachten haben, konnten sie darüber hinaus Einsichten über die Rolle von Meinungsverschiedenheiten und den Umgang damit beisteuern. Brendel berichtete, Meinungsverschiedenheiten hinsichtlich der Annahme oder Ablehnung eines Projekts würden oftmals nicht in Bezug auf denselben Punkt auftreten, sondern aufgrund einer unterschiedlichen Gewichtung der einzelnen Kriterien, die wiederum durch unterschiedliche Expertise oder disziplinäre Hintergründe zustande kämen. Auch Baldwin betonte, dass Meinungsverschiedenheiten zumeist gerade nicht aus einer wie von Keil dargestellten Ausgangssituation von gleicher Expertise in demselben Fachgebiet hervorgingen. Er verwies zudem auf die Rolle der HerausgeberInnen als letzte Instanz des Begutachtungsverfahrens, die nicht nur über die Qualität der wissenschaftlichen Leistung einer Publikation, sondern auch über jene der GutachterInnen befinden müssten. Damit verdeutlichte er die von Hirschauer angesprochene panoptische Kommunikationssituation des seinerseits unter Beobachtung stehenden Peer Reviews.

Besonders hervorzuheben unter den Berichten aus der Begutachtungspraxis ist der Beitrag von Garfinkel. Während Baldwin herausstellte, dass das Journal *Mind* sogar auf einen triple-blind-Prozess setzt, in dem auch die verantwortlichen HerausgeberInnen bei ihrer Entscheidung nicht wissen, wer die AutorInnen sind, verwies Garfinkel auf ihren Versuch, größtmögliche Transparenz zu schaffen. Zwar ist das Begutachtungsverfahren der von EMBO herausgegebenen Zeitschriften ebenfalls so organisiert, dass die AutorInnen nicht die Namen der Begutachtenden

kennen. Jedoch würden die Einschätzungen der GutachterInnen auch den jeweils anderen GutachterInnen zugänglich gemacht, so dass jede(r) auch Bezug auf andere Meinungen nehmen könne. Darüber hinaus seien bei Annahme eines Artikels zur Veröffentlichung sowohl eine Dokumentation des Ablaufs des gesamten Reviewprozesses als auch bei Einwilligung aller Beteiligten die Gutachten sowie die gesamte Kommunikation dazu öffentlich einsehbar. Transparenz anstelle von *blackboxing* empfahl Garfinkel als Weg zur Qualitätssicherung des Peer Reviews und damit der wissenschaftlichen Qualität insgesamt.

Neue methodische Entwicklungen des Peer Reviews waren auch Gegenstand des Vortrags von VERENA GRAF (Berlin). Sie sprach über Post-Publication Peer Review als eine neue Form der Begutachtung, die damit den Begutachtungsprozess über die Veröffentlichung hinaus verlängert. Zudem habe die gesamte Fachcommunity durch die öffentliche Diskussion bereits publizierter und dadurch allen zugänglicher Artikel auf Foren wie beispielsweise PubPeer <https://pubpeer.com/> die Möglichkeit, selbst über die Kriterien für wissenschaftliche Legitimität und Autorität zu entscheiden und dabei auch die Verfahrensregeln mitzugestalten. Die Offenheit dieser Form des Peer Reviews, das auch anonyme Kommentare oder Pseudonyme zulässt, berge in sich zwar immer auch die Gefahr des Missbrauchs. Jedoch habe sie auch zur Aufdeckung von wissenschaftlichem Fehlverhalten beigetragen, was zur nachträglichen Rücknahme einer Vielzahl von (naturwissenschaftlichen und medizinischen) Publikationen in renommierten Fachzeitschriften geführt habe. Post-Publication Peer Review stelle demnach den Versuch dar, dem Qualitätsproblem des herkömmlichen Peer Reviews durch die Herstellung einer öffentlichen Peer-Diskussion zu begegnen.

Anhand der von Garfinkel und Graf dargestellten Entwicklungen wird ersichtlich, dass die Diskussion der Qualität von Peer Review nicht mehr bei der Feststellung stehenbleibt, dass - in Anlehnung an ein Zitat von Winston Churchill - „Peer Review das schlechteste aller Bewertungsverfahren ist, ausgenommen alle anderen“. Stattdessen werden sowohl aus der Fachcommunity als auch vonseiten der ZeitschriftenherausgeberInnen methodische Konsequenzen gezogen, die ein starres Festhalten an bisherigen Verfahren in Frage stellen. Garfinkel sieht vor diesem Hintergrund die Zukunft in pre-print-Verfahren, bei denen ein Text bereits vor der Einreichung öffentlich kommentiert werden kann, um anschließend in einen transparenten Peer-Review-Prozess überführt zu werden.

Ergänzt wurden diese philosophischen, soziologischen und praxisbezogenen Einblicke durch juristische Perspektiven. KLAUS F. GÄRDITZ (Bonn, stellv. Mitglied des Verfassungsgerichtshofs für das Land Nordrhein-Westfalen), der als externer Gutachter für das Plagiatsverfahren gegen Bundesbildungsministerin Annette Schavan eingesetzt worden war, ging auf die Frage ein, inwieweit wissenschaftlicher Dissens ein Rechtsproblem sei. Er legte dar, dass das Recht nicht dafür zuständig sei, die Qualität in der Wissenschaft zu überprüfen, da es über keine eigene wissenschaftliche Erkenntnistheorie verfüge. Das Recht könne lediglich dazu herangezogen werden, anhand der Überprüfung, ob ein Verstoß

gegen Verfahrensregeln vorliegt, festzustellen, ob Forschung den Ansprüchen an die Wissenschaftlichkeit genügt.

WOLFGANG LÖWER (Bonn, Ombudsmann für die deutsche Wissenschaft) verwies darauf, dass bereits die Existenz von Begutachtungsprozessen in der Wissenschaft entscheidend dafür sei, dass die zu Begutachtenden auf Qualität bei der Einreichung von Förderanträgen oder Publikationen achten. Zwar stelle sich immer die Frage, wie Verfahren zu optimieren wären. Aber die Voraussetzung dafür sei die Legitimität der Verfahren, die beispielsweise durch Ombudspersonen gewährleistet werde. Eine Möglichkeit zur Steigerung dieser Legitimität sah Löwer in einer gesetzlichen Regelung zur Transparenz von Begutachtungsverfahren. So stellte er dar, dass es in der Schweiz per Gesetz möglich ist, Einsicht in die entsprechenden Unterlagen zu Förderentscheiden beim Schweizerischen Nationalfond zu nehmen. In Deutschland gibt es dagegen bislang kein entsprechendes Gesetz, das eine solche Einsichtnahme in die Begutachtungsprozesse der Deutschen Forschungsgemeinschaft erlauben würde.

Vortragsübergreifend erwies sich die fehlende Transparenz im Peer Review-Prozess als das Kernproblem, das von allen Teilnehmenden – einerseits als Gegenstand und andererseits als Problem von Forschung – thematisiert wurde. Leider konnte eine philosophische Herangehensweise an das Thema nicht weiter vertieft werden, da die beiden hierzu geplanten Vorträge entfallen mussten. Doch es wurde deutlich, dass sich etwas tut. Nicht nur wird zunehmend versucht, mehr Licht in die „black box peer review“ hineinscheinen zu lassen. Zudem wird über Möglichkeiten eines teilweisen oder gar vollständigen Umbaus von Begutachtungsverfahren nachgedacht. Die Forschung ist somit gefordert, die Organisation von Peer-Review-Prozessen als entscheidende Legitimierungsinstanz von wissenschaftlicher Erkenntnis sowie die dahinterstehenden Vorstellungen von Qualität noch stärker und – im Hinblick auf die neuen methodischen Entwicklungen – insbesondere auch vergleichend in den Blick zu nehmen. Die dadurch möglichen Befunde sollten nicht zuletzt für die Praxis von großem Interesse sein.

Konferenzübersicht:

Geert Keil (Berlin), Introduction

Martin Reinhart (Berlin), Reliable Disagreement. The (Non-)Issue of Interrater Reliability in Peer Review Research

Stefan Hirschauer (Mainz), Communication Benefits in Peer Review. Towards a Theory of an Evaluation Practice

Michèle Lamont (Harvard), The Blackboxing of Expert Judgment

Klaus F. Gärditz (Bonn), Wissenschaftlicher Dissens als Rechtsproblem

Michèle S. Garfinkel (Heidelberg), Revealing Reviewer Agreement and Disagreement: A Transparent Black Box

Thomas Baldwin (York), Editorial Judgment and Peer Review

Verena Graf (Berlin), Peer Disagreement in Post-Publication Peer Review

Elke Brendel (Bonn), Peer Disagreement and Third-Party Research Funding

Holm Tetens (Berlin), Is There an Expert's Dilemma in Philosophy? (entfallen)

Cornelis Menke (Bielefeld), Peer Disagreement and Scientific Controversies (entfallen)

Wolfgang Löwer (Bonn), Begutachtung von Forschungsanträgen im Rahmen der Regeln guter wissenschaftlicher Praxis

Stefan Hornbostel (Berlin), Taking stock

Fußnoten

1 Die Idee zum Workshop ist aus dem Forschungsprojekt „Tiefe Meinungsverschiedenheiten“ an den Universitäten Berlin (HU), Freiburg und Kiel hervorgegangen.

2 Susan Böhmer / Jörg Neufeld / Sybille Hinze / Christian Klode / Stefan Hornbostel (Hrsg.), Wissenschaftler-Befragung 2010: Forschungsbedingungen von Professorinnen und Professoren an deutschen Universitäten. iFQ-Working Paper No. 8 (2011); Martin Reinhart, Soziologie und Epistemologie des Peer Review, Baden-Baden 2012; Stefan Hirschauer, Peer Review Verfahren auf dem Prüfstand. Zum Soziologiedefizit der Wissenschaftsevaluation, in: Zeitschrift für Soziologie 33 (2004), 1, S. 62-83; Michèle Lamont, How professors think. Inside the curious world of academic judgment, Cambridge, MA, 2009.