

Begleitwort des DGPs-Vorstands zur Veröffentlichung des Berichts der Kommission Anreizsystem, Machtmissbrauch und wissenschaftliches Fehlverhalten

In den vergangenen Jahren wurden immer wieder Fälle von Machtmissbrauch in der Wissenschaft und wissenschaftlichem Fehlverhalten öffentlich. Diese Fälle traten und treten in allen Wissenschaftsdisziplinen auf, auch in der Psychologie. Für die Deutsche Gesellschaft für Psychologie (DGPs) als Vertretung der wissenschaftlich arbeitenden Psychologinnen*Psychologen ist eine kritische Auseinandersetzung mit der Frage, in welcher Hinsicht das Wissenschaftssystem solche Verhaltensweisen ermöglicht oder unter Umständen sogar durch (Fehl-)Anreize fördert, zwingend erforderlich. Unsere Fachgesellschaft hat sich in den vergangenen Jahren bereits an verschiedenen Stellen mit Fragen der Qualitätssicherung in Forschung und Lehre sowie mit der Verbesserung von Arbeitsbedingungen in allen Karrierephasen beschäftigt. Zusätzlich wurden von engagierten DGPs-Mitgliedern und dem DGPs-Vorstand konkrete Aktivitäten innerhalb der akademischen Psychologie angestoßen.

Eine solche Aktivität bestand in der Einsetzung des DGPs-Ombudsgremiums im Jahr 2020, einem ergänzenden Angebot zu den Anlaufstellen der Deutschen Forschungsgemeinschaft (DFG) und der lokalen Universitäten und Forschungsinstitute. Das Ombudsgremium beschäftigte sich neben der Beratung von Betroffenen mit Fragen der guten wissenschaftlichen Praxis, wissenschaftlichem Fehlverhalten, sowie Fehlverhalten am Arbeits- und Studienplatz an psychologischen Instituten in Deutschland. Das Gremium führte daher im Juli 2020 erstmalig eine deutschlandweite Befragung von Mitarbeitenden psychologischer Institute und Wissenschaftseinrichtungen durch, um eine Bestandsaufnahme von Verstößen gegen die wissenschaftliche Integrität in der akademischen Psychologie vorzunehmen. Die Umfrage ergab, dass Machtmissbrauch und bewusstes wissenschaftliches Fehlverhalten in der Psychologie keine Einzelfälle sind – auch wenn die Ergebnisse aufgrund der freiwilligen Teilnahme der kontaktierten Personen kein repräsentatives Bild der Lage in Bezug auf unethisches Verhalten liefern kann. Die Mehrheit der Befragten gab an, Fehlverhalten in einem frühen Karrierestadium erlebt zu haben – also in dem Stadium der wissenschaftlichen Karriere, in dem die Abhängigkeiten von Vorgesetzten und Betreuer*innen am größten sind. Die Umfrage soll nun in regelmäßigen Abständen und mit verbesserter Methodik wiederholt werden, um Veränderungen in der Häufigkeit wissenschaftlichen Fehlverhaltens sowie die Wirksamkeit angestoßener Maßnahmen zu evaluieren.

Ende 2021 wurde außerdem die DGPs-Kommission „Anreizsystem, Machtmissbrauch und wissenschaftliches Fehlverhalten“ (AMWF) eingerichtet. Ihr Arbeitsauftrag besteht darin, systemische Faktoren zu identifizieren, die zu Machtmissbrauch und wissenschaftlichem Fehlverhalten im Wissenschaftssystem beitragen können, und Lösungsvorschläge zur Veränderung der Strukturen zu erarbeiten. Unter der Leitung von Prof. Dr. Daniel Leising (TU Dresden) und mit Beteiligung aller wissenschaftlichen Statusgruppen hat die Kommission jetzt mit ihrem differenzierten Bericht eine theoretische Einordnung und Ursachenbeschreibung vorgelegt sowie umfangreiche Möglichkeiten des Gegensteuerns und der Abhilfe aufgezeigt. Die dem Bericht zugrundeliegende systemische Betrachtung von Machtmissbrauch und wissenschaftlichem Fehlverhalten zeigt, dass die dargestellten Probleme das gesamte Wissenschaftssystem auf verschiedensten Ebenen und nicht nur die Psychologie betreffen.

Aus Sicht des DGPs-Vorstands liefert der Bericht wichtige Diskussionsimpulse und fundierte Ansatzpunkte für Veränderungen. Die Kommission hat insgesamt 12 Problembereiche identifiziert, zu denen unter anderem bedenkliche Anreizsysteme in Berufungs- und Bewertungsverfahren, ein starkes Machtgefälle sowie die prekäre Situation von Wissenschaftler*innen in frühen Karrierephasen gehören.

Der DGPs-Vorstand wird sich mit den im Bericht entwickelten Vorschlägen beschäftigen und dabei auch die beabsichtigten Wirkungen sowie potenzielle Nebenwirkungen sorgfältig abwägen. Dort, wo die DGPs als Vertretung der akademischen Psychologie aktiv werden kann, wird der Vorstand gemeinsam mit den Kommissionen und Fachgruppen der DGPs Maßnahmen anstoßen.

Auch die bereits initiierten Maßnahmen werden weitergeführt, wie zum Beispiel eine erneute Befragung der Beschäftigten in der akademischen Psychologie durch das DGPs-Ombudsgremium. Diese erneute Befragung wurde mit fachlicher Beratung durch das Leibniz-Institut für Sozialwissenschaften (GESIS) erstellt; sie wird konkrete Hinweise auf die Beschaffenheit und das Ausmaß der Probleme sowie auf mögliche Veränderungen seit der ersten Befragung geben. Der Fakultätentag Psychologie (FTP) führt außerdem aktuell eine Umfrage zur Förderung guter wissenschaftlicher Praktiken (GWP) an den Instituten durch, mit der das Angebot sowie der Bedarf an Trainings- und Fortbildungsmöglichkeiten für Wissenschaftler*innen aller Karrierephasen ermittelt werden soll.

Machtmissbrauch und wissenschaftliches Fehlverhalten sind Phänomene, die systemisch betrachtet werden müssen. Natürlich sind auch personale Faktoren beteiligt – einige Personen lassen sich leichter zu solchen Verhaltensweisen verleiten als andere. Aber der Verweis auf das fragwürdige Verhalten Einzelner greift zu kurz, wenn man die Probleme umfassend erklären und adressieren möchte, da er die Rolle situativer Faktoren unterschlägt, die es solchen Personen ermöglichen, in bestimmten Kontexten ungestört wirken zu können. Machtmissbrauch und wissenschaftliches Fehlverhalten treten nicht isoliert auf: Sie werden durch Anreizsysteme begünstigt, und zwar über alle Wissenschaftsdisziplinen hinweg, von der Grundlagen- bis hin zur Anwendungsforschung. Um nachhaltige Veränderungen anzustoßen, ist es daher auch erforderlich, bei der Analyse das wissenschaftliche System in seiner Gesamtheit zu betrachten. Im Wissenschaftssystem wiederum können Änderungen nur dann erfolgreich sein, wenn sie von vielen Akteurinnen*Akteuren mitgetragen werden. Daher richtet sich mit der Veröffentlichung dieses Berichts auch ein Appell an andere Fachgesellschaften, Hochschulen und Forschungseinrichtungen sowie politische Entscheidungsträger*innen, sich an der Auseinandersetzung zu beteiligen, sich gemeinsam für bessere Arbeitsbedingungen am Wissenschaftsstandort Deutschland einzusetzen sowie gemeinsame Grundsätze für die Bekämpfung von Machtmissbrauch und Fehlverhalten über alle Wissenschaftsdisziplinen hinweg zu erarbeiten.

„An allem Unfug, der passiert, sind nicht etwa nur die schuld, die ihn tun, sondern auch die,
die ihn nicht verhindern.“

Erich Kästner

Anreizsystem, Machtmissbrauch und Wissenschaftliches Fehlverhalten

Eine Analyse zum funktionalen Zusammenhang zwischen strukturellen Bedingungen
und unethischem Verhalten in der Wissenschaft

DGPs-Kommission

“Anreizsystem, Machtmissbrauch und Wissenschaftliches Fehlverhalten”

Nicole Bössel (Universitätsmedizin Greifswald)

Annette Kluge (Ruhr-Universität Bochum)

Daniel Leising (Vorsitz) (TU Dresden)

Dorothee Mischkowski (MPI zur Erforschung von Gemeinschaftsgütern)

Le Vy Phan (Universität Bielefeld)

Manfred Schmitt (Rheinland-Pfälzische Technische Universität Kaiserslautern-Landau)

Jutta Stahl (Universität zu Köln)

Inhaltsverzeichnis

Einleitung	3
Beratungshergang	4
Grundlegende konzeptuelle Vorstellungen	5
Multiple Situiertheit	8
Zusammenfassende Auflistung potenzieller Maßnahmen	11
Differenzierte Analyse der identifizierten Problembereiche	18
Problembereich 1. Nachlässiges oder betrügerisches Forschungshandeln (P1)	19
Problembereich 2. Machtmissbrauch (P2)	23
Problembereich 3. Unzureichende Betreuung von ECRs (P3)	25
Problembereich 4. Mangelhaft Qualität der Lehre (P4)	28
Problembereich 5. Fehlanreize (P5)	30
Problembereich 6. Überfrachtung der Professur mit Aufgaben (P6)	33
Problembereich 7. Befristungspraxis für Stellen unterhalb der Professur (P7)	36
Problembereich 8. Unnötig starkes Machtgefälle (P8)	38
Problembereich 9. Versagen des Peer-Review-Systems (P9)	41
Problembereich 10. Mangelnde Qualität von Berufungsverfahren (P10)	44
Problembereich 11. Mangelnde Klarheit und Verbindlichkeit ethischer Normen (P11)	49
Problembereich 12. Schwache Kontroll- und Sanktionsinstanzen (P12)	51
Literatur	54

Einleitung

Die Kommission "Anreizsystem, Machtmissbrauch und wissenschaftliches Fehlverhalten" (AMWF) wurde vom Vorstand der Deutschen Gesellschaft für Psychologie (DGPs) im Januar 2022 eingesetzt. Ihr Auftrag besteht in der Analyse derjenigen Parameter des Wissenschaftssystems, die unethisches Verhalten im Wissenschaftssystem ermöglichen oder fördern, sowie der Erarbeitung von konkreten Vorschlägen zur Verbesserung der Situation.

Unter *unethischem Verhalten* versteht die Kommission Verhaltensweisen, deren kurz- und langfristige Konsequenzen im Widerspruch zu weithin akzeptierten ethischen Imperativen stehen (z.B. definiert in UN-Charta, Grundgesetz, DFG-Leitlinien zur Sicherung Guter Wissenschaftlicher Praxis, fachspezifischen Ethikcodizes), insbesondere der Vorstellung, dass wissenschaftliche Arbeit dem Gemeinwohl dienen soll und dass Schädigungen einzelner Personen zu vermeiden sind.

Anlass der Einsetzung der Kommission war der sich konsolidierende Eindruck, dass ethisch problematisches Verhalten im Wissenschaftskontext einerseits nicht selten und zum Teil gravierend ist, andererseits aber oft unentdeckt oder unsanktioniert bleibt. Dieser Eindruck speist sich aus in der Publikumspresse erscheinenden Berichten über mehr oder weniger drastische, konkrete Fälle, ebenso wie aus Befunden systematischer Forschung zum Thema (z.B. Elson, Fiedler, Kirsch & Stahl, 2021). Unethisches Verhalten ist kein spezifisches Problem der Psychologie, sondern auch in vielen anderen akademischen Fächern zu beobachten. Da die Psychologie jedoch eine besondere Expertise im Bereich des menschlichen Verhaltens und Fehlverhaltens hat, sehen der DGPs-Vorstand und die Kommissionsmitglieder das Fach in einer besonderen Verantwortung, sich proaktiv und aus wissenschaftlicher Perspektive mit dem Thema auseinanderzusetzen und Gegenmaßnahmen zu erarbeiten. Ein Gegensteuern scheint notwendig, um die Glaubwürdigkeit der Forschung zu stärken und um Machtmissbrauch im Wissenschaftsbetrieb einzudämmen.

Der Fokus der Kommissionsarbeit lag und liegt bei der Erarbeitung von präventiven Maßnahmen zur Vermeidung unethischen Verhaltens. Zusätzlich werden intervenierende Maßnahmen nach Auftreten von Fehlverhalten vorgestellt. Da unethisches Verhalten sowohl bei reinen Forschungstätigkeiten als auch in der akademischen Ausbildung (d.h. Lehre im Bachelor-/Masterstudium bis hin zur Betreuung von Promovierenden und Postdocs) zu beobachten ist, werden beide Aspekte betrachtet.

Beratungshergang

Die Mitglieder der Kommission "Anreizsystem, Machtmissbrauch und wissenschaftliches Fehlverhalten" haben sich bis zum DGPs Kongress 2022 in Hildesheim zu insgesamt 21 regulären Sitzungen getroffen. Die konstituierende Sitzung fand am 21.01.2022 statt.

Die Vorgehensweise der Kommission bestand in

1. Auswertung von einzelnen besonders problematischer Fallbeispielen ebenso wie von Best-Practice-Beispielen
2. Intensiven Diskussionen persönlicher Erfahrungen
3. Rezeption relevanter Lektüre
4. Konsultation externer Expert:innen zu einzelnen Schwerpunktthemen
 - a. *Machtgefälle, Anstellungsverhältnisse und Befristungen* (Expert:innen: Rolf Nichelmann, Tilman Reitz, Tobias Rosefeldt, Jule Specht, Patrick Wöhrle)
 - b. *Kontroll- und Sanktionsinstanzen, Ombudswesen* (Expert:innen: Peter Kirsch, Malte Elson und Jutta Stahl vom Ombudsgremium der DGPs¹; Eva Korus und Sabine Chai von der Österreichischen Agentur für wissenschaftliche Integrität; Stephan Rixen, Hjördis Czesnick vom Gremium Ombudsman für die Wissenschaft)
 - c. *Personalauswahl* (Expert:innen: Gerhard Blickle, Sibylle Detel, Uwe Kanning, Martin Kersting, Martin Kleinmann, Stefan Krumm, Bernd Markus, Patrick Mussel, Tuulia Ortner und Dominik Schwarzinger).

Für weitere Beratung, Feedback und Unterstützung in Einzelfragen bedankt sich die Kommission bei Maximilian Frank, Kathrin Fücke, Tobias Heed, Gina Henze, Jana Lasser, Seraina Lerch, Tuulia Ortner, Julia Rohrer, Sonja Utz, Erich Witte und Joachim Wittkowski, sowie bei den Mitgliedern des Netzwerks gegen Machtmissbrauch in der Wissenschaft.

Aufgrund der Komplexität der identifizierten Problembereiche und der Diversität der vorgeschlagenen Lösungsansätze muss die Erarbeitung und (Weiter-)Entwicklung sämtlicher Empfehlungen als ein kontinuierlicher und noch nicht abgeschlossener Prozess betrachtet werden. Es ist geplant, dass die Kommission nach der Fertigstellung dieses ersten Berichts weiter an der Thematik arbeitet.

Zwei Themen von großer Wichtigkeit konnten in der ersten Phase der Kommissionsarbeit nicht ausreichend gründlich bearbeitet werden. Sie wurden daher im vorliegenden Bericht weitgehend ausgespart, sollten aber zeitnah angegangen werden: Das erste dieser Themen ist die weiterhin mangelhafte Vereinbarkeit von Beruf und Familie in der Wissenschaft. Das zweite Thema ist die Projektförderung, mit besonderem Augenmerk auf Verbundprojekte, Qualitätssicherung, und die in diesem Bereich bestehenden (Fehl-)Anreize.

¹ Da die Arbeit der AMWF-Kommission auf den Befunden des DGPs-Ombudsgremiums aufbaut, wurde Jutta Stahl seitens des Präsidiums bewusst nominiert, um diese Expertise zu übertragen und eine Vernetzung der Gremien zu erleichtern.

Grundlegende konzeptuelle Vorstellungen

Die Analyse wurde von in Forschung und Lehre aktiven Psycholog:innen vorgenommen. Sie bedient sich gängiger theoretischer Konzepte aus der Psychologie und adressiert auf dieser Basis die innerhalb der akademischen Psychologie beobachteten Probleme. Gleichwohl kann vermutet werden, dass viele der hier gemachten Aussagen sich auf den Wissenschaftsbetrieb insgesamt – also auch auf andere Fächer und Länder – übertragen lassen.

Abbildung 1 enthält ein Modell der Verhaltensregulation mit Variablen, die von der Kommission im gegebenen Kontext als besonders relevant identifiziert wurden: Das Verhalten von Individuen in akademischen Institutionen wird in diesem Modell zunächst von Variablen (“Motivatoren”) aus zwei zentralen Motivbereichen beeinflusst:

- *Akademische Ideale* wie Wissenszuwachs und Wissensvermittlung, Integrität und Gemeinwohl. Menschen können bestimmte Verhaltensweisen (z.B. Fehlerkontrolle) zeigen, um diesen Idealen zu entsprechen. Die *Existenzberechtigung* akademischer Institutionen speist sich vor allem aus dem Ausmaß, in dem das Verhalten ihrer Mitglieder diesen Idealen dient.
- Aussicht auf *persönliche Vorteile* wie Existenzsicherung, Profit, Prestige oder Reproduktion eines tradierten akademischen Status. Menschen können bestimmte Verhaltensweisen zeigen (z.B. Vertuschung von Fehlern), die primär im Dienst dieser Motive stehen.

Die Kommission geht davon aus, dass das Verhalten der meisten Personen, die in akademischen Institutionen tätig sind, von Motivatoren aus *beiden* Bereichen beeinflusst wird, auch wenn sich die durchschnittliche Stärke der jeweiligen Einflüsse wahrscheinlich zwischen Individuen unterscheidet. Die Einflüsse von Motivatoren aus den beiden Bereichen können sich im Prinzip gegenseitig unterstützen (wenn z.B. der Wunsch nach Anerkennung in der Community dazu führt, dass eine Forscherin sich ernsthaft um besonders aussagekräftige Daten und Analysen bemüht), aber auch in dysfunktionaler Konkurrenz zueinander stehen (wenn z.B. fragwürdige wissenschaftliche Praktiken zum Einsatz kommen, um schnell die eigene Publikationsliste zu verlängern und damit die Chancen auf eine Festanstellung zu steigern).

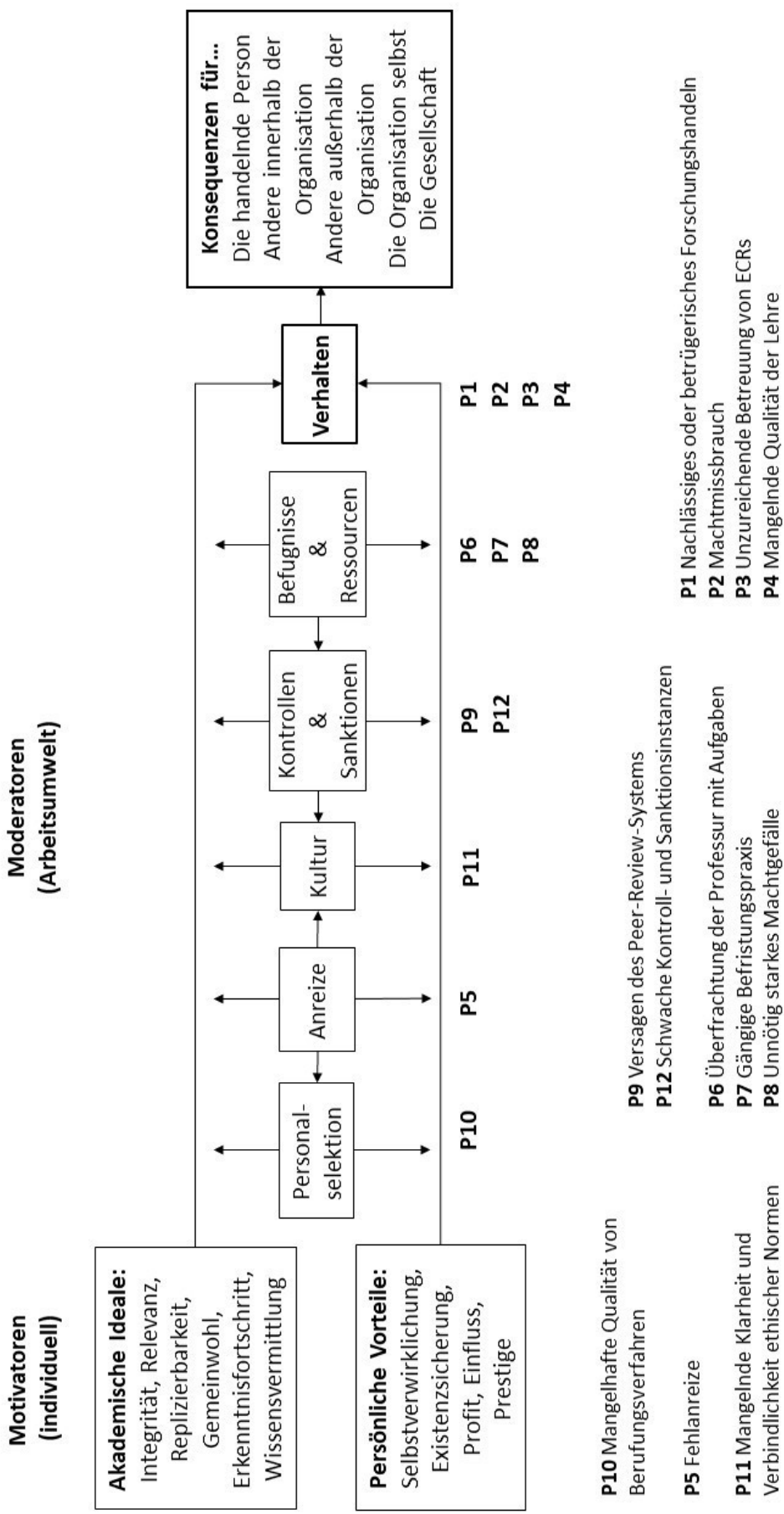


Abbildung 1. Modell der Verhaltensregulation mit zwei Typen von Motivatoren auf der individuellen Seite und fünf Moderatoren auf Seiten der Arbeitsumwelt. Bezug zu zwölf im vorliegenden Bericht identifizierten Problembereichen (P1 bis P12; Details s. Text); ECRs = Early Career Researchers.

Verhalten hängt aber nicht nur von Motiven ab, sondern ganz entscheidend auch von Umgebungsbedingungen. Diese im Modell als "Moderatoren" bezeichneten Variablen sind Klassen von Variablen, die sich zum Teil auch untereinander beeinflussen können (Abb. 1). Das Ausmaß, in dem das Verhalten von Personen in einer akademischen Organisation eher dem einen oder dem anderen Motivbereich dient, hängt unserer Ansicht nach entscheidend von den folgenden Moderatoren ab:

- (1) welche Personen (mit welchen Eigenschaften und Fähigkeiten) in der Organisation eingestellt und weiterbeschäftigt werden (Personalselektion),
- (2) welches Verhalten dort (direkt oder indirekt) belohnt wird (Anreize),
- (3) welches Verhalten *andere* Organisationsmitglieder in der Regel zeigen (Kultur),
- (4) ob und wie "unethisches" Verhalten entdeckt und ggfs. sanktioniert wird (Kontroll- und Sanktionsmechanismen), und
- (5) welchen Handlungsspielraum (Befugnisse und Ressourcen) die Personen in der Organisation haben.

Die vorliegende Analyse stellt dar, wie die im Wissenschaftssystem derzeit geltenden Umgebungsbedingungen dazu führen können, dass das Verhalten der darin arbeitenden Personen in nicht mehr vertretbarem Ausmaß von dem abweicht, was akademische Ideale gebieten würden. Dadurch wird u.a. die Legitimation wissenschaftlicher Arbeit gefährdet.

Zwölf in dieser Hinsicht zentrale Problembereiche wurden im Verlauf der Kommissionsarbeit identifiziert. Sie sind in Abbildung 1 aufgeführt und werden im zweiten Teil des vorliegenden Berichts im Detail diskutiert. Dabei ist zu beachten, dass die Analyse nicht zwischen einzelnen akademischen Institutionen unterscheidet, sondern sich vielmehr auf typische institutionelle Bedingungen bezieht.

Das aus dem Zusammenspiel von Motivatoren und Moderatoren resultierende Verhalten ist dann als problematisch zu betrachten, wenn es zu ethisch inakzeptablen Konsequenzen führt. Ob dies der Fall ist, lässt sich vor allem anhand eines Abgleichs zwischen den langfristigen Konsequenzen herleiten, die das Verhalten für verschiedene Akteur:innen hat (s. Abbildung 1). So ist zum Beispiel nachlässiges oder betrügerisches Forschungshandeln auch deshalb als ethisch problematisch zu betrachten, weil es öffentliche Gelder nutzt, ohne für diejenigen, die diese Gelder beisteuern, einen Mehrwert (etwa im Sinne eines relevanten Wissenszuwachses) zu erbringen.

Multiple Situiertheit

Für die vorliegende Analyse ist relevant, dass das Verhalten von Personen in akademischen Institutionen *multiple situier*t ist. Damit ist gemeint, dass das Verhalten individueller Personen im Wissenschaftsbetrieb in verschiedene institutionelle und rechtliche Kontexte (z.B. das *Wissenschaftszeitvertragsgesetz*) sowie hochschulspezifische Regelwerke, Ordnungen und Richtlinien (z.B. Berufungsordnungen) eingebettet ist. Diese Kontexte sind hierarchisch organisiert, wie in Abbildung 2 dargestellt.

Diese multiple Situiertheit ist aus mindestens zwei Gründen relevant: Erstens wird es nötig sein, für die hier beschriebenen Verbesserungsvorschläge die jeweils relevante Ebene zu identifizieren, auf der die Interventionen stattfinden müssen. Zweitens können Änderungen auf einer Ebene mit impliziten oder expliziten Regularien anderer Ebenen in Konflikt geraten. Ein offensichtliches Beispiel hierfür sind die Evaluationskriterien für akademische Leistungen im Rahmen von Berufungsverfahren. Sollte sich eine Berufungskommission entscheiden, andere als die gängigen Kriterien (wie h-Index und eingeworbene Drittmittel) zu nutzen, würde sie unmittelbar in Widerspruch mit den Interessen der Fakultät und der Hochschule geraten, welche selbst anhand solcher Kriterien evaluiert werden.

Zu den relevanten Akteur:innen gehören neben den Institutionen vor allem drei Gruppen von Personen: (1) Personen, denen hauptverantwortlich Aufgaben im Rahmen von Forschung, Lehre und Nachwuchsförderung übertragen wurden (i.d.R. *Professor:innen*), (2) Personen, die sich aktuell auf eine spätere Tätigkeit in Forschung und Lehre vorbereiten (sie werden im Folgenden als *Early Career Researchers* (ECRs) bezeichnet, und (3) *Studierende*. Sie alle interagieren tagtäglich vor dem Hintergrund der komplexen institutionellen Bedingungen miteinander.

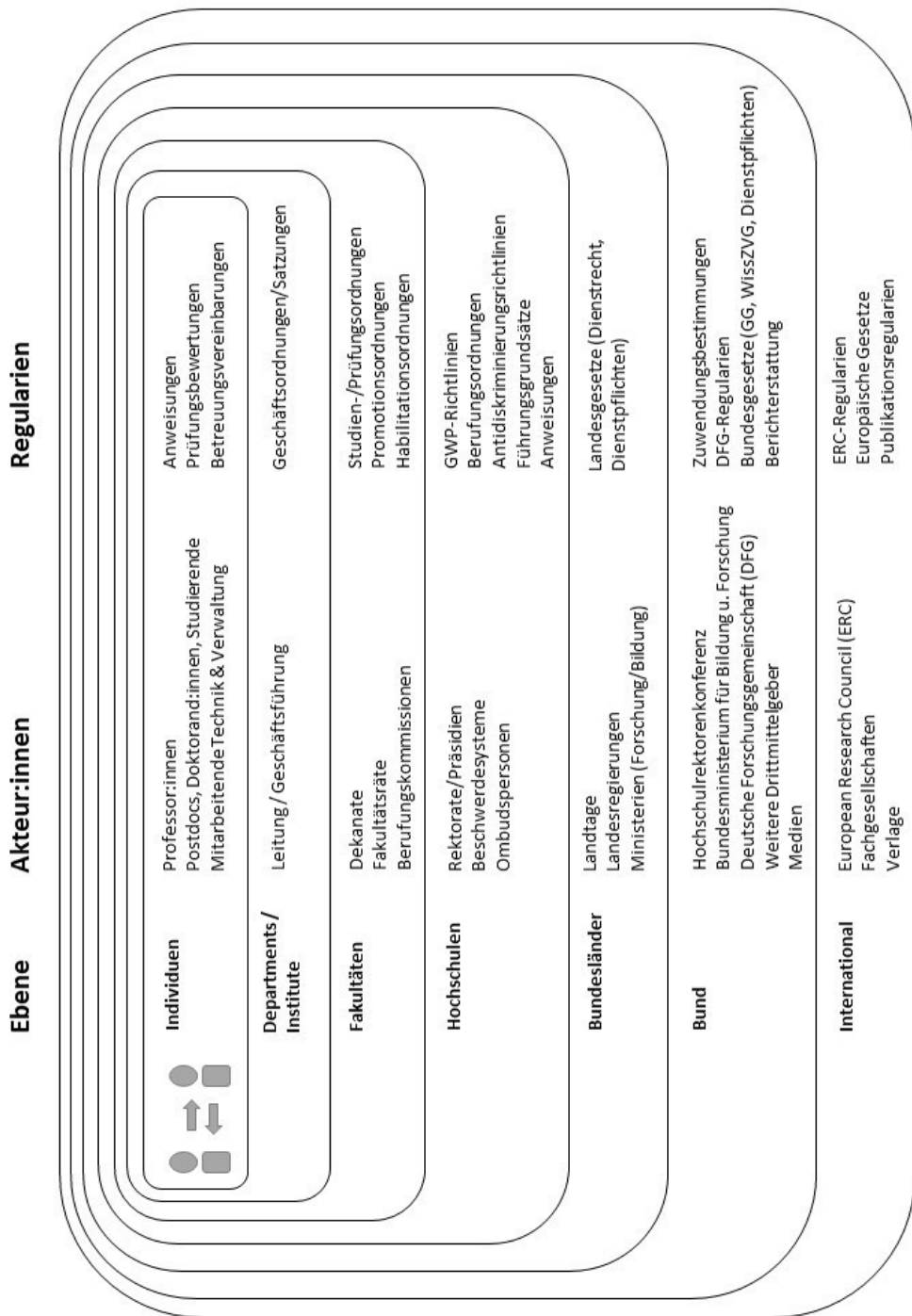


Abbildung 2. Multiple Situiertheit maßgeblicher Akteur:innen und Institutionen im Wissenschaftssystem. GWP = Gute Wissenschaftliche Praxis, GG = Grundgesetz, WissZVG = Wissenschaftszeitvertragsgesetz.

Vor allem zwischen den Mitgliedern der erstgenannten Gruppe und denen der zweiten und dritten Gruppe besteht dabei ein erhebliches Machtgefälle, da Professor:innen qua Amt über eine ganze Reihe verschiedener Machtquellen verfügen (z.B. Bewertungsmacht, Belohnungsmacht, Bestrafungsmacht, Informationsmacht, Personalmacht).

Das Verhalten der Akteur:innen findet statt vor dem Hintergrund von Regularien der Fakultäten und Hochschulen (z.B. Prüfungs-, Promotions-, und Berufungsordnungen, Entfristungsregelungen, Evaluationsrichtlinien für Juniorprofessuren). Diese Institutionen sind Teil des Hoheitsbereichs der Ministerien des jeweiligen Bundeslandes, zugleich aber auch in das übergeordnete (nationale und internationale) Wissenschaftssystem eingebettet.

Professor:innen werden durch ihre Stellung - etwa im Rahmen des Beamt:innenrechts - und durch die jeweilige Dienststelle (Hochschule) Aufgaben und Verantwortungen in der Forschung und in der Lehre übertragen. Ihre beruflichen Handlungsfreiheiten werden einerseits durch das Grundgesetz (Freiheit von Forschung und Lehre) legitimiert und andererseits durch Strafgesetzgebung (Bund), Dienstrecht (Land) etc. limitiert.

Das *Wissenschaftssystem* bietet auf den unterschiedlichen Ebenen verschiedene Anreize für das Verhalten eines Individuums (vgl. Leising et al., 2022a): Zu den eher *extrinsischen* Anreizen gehört die vergleichsweise sehr hohe Arbeitsplatzsicherheit für Personen, die eine Professur oder eine der wenigen unbefristeten Positionen (z.B. Akademische Ratsstellen) erreichen, ebenso wie alle finanziellen Belohnungen für als wünschenswert erachtete Leistungen. Die Mittelvergabe der Ministerien an Hochschulen orientiert sich häufig an Parametern wie Studierenden- und Absolvent:innenzahlen, Drittmittelinwerbungen und/oder Publikationszahlen. Insbesondere Drittmittelinwerbungen werden dann wiederum von Hochschulen und Fakultäten als Bewertungskriterium auf der individuellen Ebene verwendet und zum Beispiel im Rahmen von Berufungsverfahren, Zwischenevaluationen (Juniorprofessuren) oder Verhandlungen über Leistungszulagen herangezogen. Als eher extrinsischer Anreiz ist darüber hinaus das mit einer festen Stelle in der Wissenschaft verbundene, relativ hohe gesellschaftliche Ansehen zu betrachten, ebenso wie das Verfügen über die bereits genannten Machtquellen für Professor:innen und Arbeitsgruppenleitungen.

Zu den eher *intrinsischen* Anreizen einer Tätigkeit in der Wissenschaft gehören der sehr große Handlungsspielraum, die hohe Komplexität der Arbeitsaufgaben sowie die Möglichkeit, sich mit herausfordernden, ungelösten Fragen zu beschäftigen, diese im besten Fall einer Lösung näherzubringen, und damit eventuell sogar zum Gemeinwohl beizutragen. Auch die Möglichkeit, im Rahmen von Lehre und Forschung in ständigem Austausch mit jungen Talenten (Studierende und ECRs) zu stehen und diese fördern zu können, wird von vielen Wissenschaftler:innen als großes Privileg empfunden.

Die Psychologie versteht sich als empirische Wissenschaft. Wissenschaftliche Forschung kann mit Mitteln der Professur (sog. Haushaltsmittel) oder mit zusätzlich eingeworbenen Drittmitteln (z.B. DFG, BMBF, sowie privatwirtschaftliche Mittel etc.) durchgeführt werden. Beim Versuch der Einwerbung von Drittmitteln für die Forschung wird die Qualität des

jeweiligen Vorhabens durch eine Anzahl von Gutachter:innen geprüft und dann in einem übergeordneten Gremium (z.B. Fachkollegium Psychologie in der DFG) kompetitiv im Vergleich zu anderen Anträgen bewertet und eventuell zur Förderung empfohlen. Fördergeldgebende wie DFG oder BMBF haben dabei prinzipiell die Möglichkeit, durch Setzung von Kriterien (z.B. Open Science / Gute Wissenschaftliche Praxis) direkten Einfluss auf das individuelle oder kollektive Verhalten potenzieller Drittmittelempfänger:innen zu nehmen. Auch die Grundfinanzierung der Hochschulen durch die Bundesländer wird regelmäßig an die Erreichung konkreter Zielvorgaben gekoppelt (z.B. hinsichtlich Gleichstellungsmaßnahmen).

Die Veröffentlichung von Forschungsergebnissen in der wissenschaftlichen Psychologie erfolgt bisher überwiegend über die Zeitschriften kommerzieller Verlage und ein damit verbundenes Peer-Review-Verfahren. Erklärtes Ziel desselben ist die Qualitätssicherung: Durch konstruktives Feedback und Hinweise zu Verbesserungsmöglichkeiten soll eine möglichst hohe Qualität der Berichterstattung über Forschungsergebnisse gewährleistet werden. Allerdings verfolgen Wissenschaftsverlage immer auch eigene wirtschaftliche Interessen, welche mit dem Ziel der wissenschaftlichen Qualitätsmaximierung teilweise inkompatibel sind. So können strenge Begutachtungen dazu führen, dass das Volumen publizierbarer Manuskripte sinkt, was Umsatzeinbußen der Verlage zur Folge haben kann.

Im Rahmen ihres Marketings verwenden Wissenschaftsverlage Metriken (wie z.B. den Journal Impact Factor), die kenntlich machen sollen, wie relevant, anspruchsvoll und angesehen eine Fachzeitschrift innerhalb der Scientific Community ist. Solche Metriken spiegeln einerseits direkt das (Einreichungs- und Zitations-)Verhalten der Forschenden wider und spielen andererseits oft eine wichtige Rolle im Rahmen von individuellen Evaluationen (z.B. Berufungsverfahren).

Zusammenfassende Auflistung potenzieller Maßnahmen

Das Ziel der Kommissionsarbeit bestand in der Erarbeitung von konkreten Lösungsvorschlägen. Auf den nächsten Seiten werden zunächst alle Empfehlungen in kondensierter Form nacheinander aufgelistet². Im zweiten Teil des Berichts finden sich die differenzierten Analysen der einzelnen Problembereiche, welche die Grundlage für diese Empfehlungen waren, sowie detaillierte Beschreibungen der einzelnen Empfehlungen.

Nicht alle Empfehlungen werden von allen Mitgliedern der Kommission als gleichermaßen hilfreich eingeschätzt. Dies liegt daran, dass die Konsequenzen mancher Veränderungen in einem derart komplexen System wie der Wissenschaft schwer absehbar sind. In einigen Fällen erscheint nur eine Kombination von mehreren Maßnahmen sinnvoll, und einzeln

² Auch zur Vermeidung der verschiedenen Formen des Machtmissbrauchs (P2) werden Maßnahmen vorgeschlagen. Da diese sich mit Maßnahmen aus anderen Kapiteln überschneiden, werden sie hier nicht separat aufgeführt. Im zweiten Teil des Berichts wird in Kapitel P2 auf die jeweiligen Maßnahmenvorschläge in anderen Kapiteln verwiesen.

herausgenommene Maßnahmen könnten sogar zu einer Verschlechterung führen. Da die Umsetzung der einzelnen Maßnahmen unterschiedlich viel Zeit in Anspruch nehmen wird (z.B. Gesetzesänderungen vs. Einführen von Betreuungsvereinbarungen), müssen deren wahrscheinliche Auswirkungen sowohl einzeln als auch in Kombination miteinander abgeschätzt werden, was immer mit einer gewissen Unsicherheit behaftet ist. Im schlimmsten Fall führen eigentlich als Verbesserungen gedachte Maßnahmen zu neuen Problemen. Die Kommission schlägt daher vor, systematische, begleitende *Forschung* zu den angeregten, dynamischen Veränderungsprozessen von Anfang an mitzudenken und bereits jetzt zu initiieren.

Förderung Guter Wissenschaftlicher Praxis (GWP)

- (1.1.)** Beginn einer Kulturentwicklung durch die Vergabe eines Gütesiegels ähnlich dem “great place to work“-Award - aber spezifisch bezogen auf Universitäten, Hochschulen und außeruniversitäre Forschungseinrichtungen.
- (1.2.)** Bereitstellung und routinemäßige Nutzung ausgereifter Infrastrukturen für die Implementierung Guter Wissenschaftlicher Praktiken durch Universitäten, Hochschulen und Forschungseinrichtungen, sowie durch externe Akteure.
- (1.3.)** Explizite Anerkennung und Berücksichtigung des Mehraufwandes, der durch Einhaltung höherer Forschungsstandards entsteht, im Rahmen von Verfahren zur Einwerbung von Drittmitteln.
- (1.4.)** Ausbau des Trainings in GWP für individuelle Forschende.
- (1.5.)** Theoretische Vermittlung und Anwendung von GWP und Open Science als fester Bestandteil der Lehre.
- (1.6.)** Stärkung des Psychologiestudiums bezüglich der Themen Theoriebildung, Theorie-spezifikation und Theorietestung.
- (1.7.)** Initiierung systematischer Prozesse mit dem Ziel einer Ordnung der bislang chaotischen Konstrukt- und Begriffslandschaft in der Psychologie.
- (1.8.)** Entwicklung und Spezifikation von Elementen einer konstruktiven Fehlerkultur.
- (1.9.)** Gründung einer DGPs-Fachgruppe “Theoretische Psychologie”.
- (1.10.)** Weitere Bekanntmachung des “Negativkatalogs” der DFG zu Beiträgen, die keine Co-Autorenschaft an einem wissenschaftlichen Artikel rechtfertigen.
- (1.11.)** Standardmäßige Erklärung zu den tatsächlichen Beiträgen aller genannten Autor:innen (“authors’ contributions”) zu Veröffentlichungen.

Strukturelle Änderungen in der Zusammenarbeit zwischen Professor:innen und ECRs

- (3.1.)** Öffentliche Erklärung (der Hochschulen, der DGPs, der die Forschung fördernden Einrichtungen) zu einer plausiblen Maximalzahl von ECRs, die gleichzeitig mit derselben betreuenden Person wissenschaftlich zusammenarbeiten können.
- (3.2.)** Eine große Anzahl von (gleichzeitig) Promovierenden pro Betreuer:in sollte nur unter besonderen Bedingungen zugelassen und die Qualität der Betreuung dann regelmäßig und sorgfältig überprüft werden (z.B. durch Befragungen der betroffenen ECRs).
- (3.3.)** Formulierung eines Leitbildes zur Zusammenarbeit von ECRs und Professor:innen.
- (3.4.)** Verstärkte Nutzung von Betreuungsvereinbarungen, um die Berechenbarkeit und Steuerbarkeit der wissenschaftlichen Zusammenarbeit für ECRs zu erhöhen.
- (3.5.)** Regelmäßiges, strukturiertes Feedback an Professor:innen bzgl. der Qualität der Zusammenarbeit mit ECRs durch Letztere.
- (3.6.)** Explizite, rechtsverbindliche Regelung des Verfahrens zum Betreuungswechsel in der Promotionsordnung.
- (3.7.)** Regelmäßige Befragung der ECRs zu ihren Betreuungserfahrungen durch eine unabhängige Instanz (z.B. Graduiertenschulen) ab dem Beginn der Zusammenarbeit.
- (3.8.)** Einführung von Thesis Advisory Committees (TACs) als Regelmodell der Promotionsbetreuung.

Wertschätzung für gute Lehre und Entwicklung validerer Methoden der Lehrevaluation

- (4.1.)** Schaffung expliziter, wirksamer (auch finanzieller) Anreize für Engagement und überdurchschnittliche Qualität in der Lehre.
- (4.2.)** Explizit stärkere Gewichtung von qualitativ guter Lehre in Berufungsverfahren und im Rahmen von Evaluationen von Fakultäten und Universitäten.
- (4.3.)** Ausbau der hochschuldidaktischen Weiterbildung für Dozierende und Ermunterung oder sogar Verpflichtung zur regelmäßigen Teilnahme.
- (4.4.)** Proaktive Ansprache individueller Lehrender durch die Institution, verbunden mit dem Angebot einer freiwilligen externen Evaluation und Beratung durch Expert:innen bei der Gestaltung der Lehre.
- (4.5.)** Ermunterung der Dozierenden zu stärkerer inhaltlicher Abstimmung untereinander und ggfs. zum Austausch von Lehrmaterialien.
- (4.6.)** Das derzeitige Vorgehen bei der Lehrevaluation (v.a. durch Studierende) sollte kritisch hinterfragt und optimiert erweitert werden.
- (4.7.)** Die Frage nach angemessenen Prüfungsformen in der Psychologie sollte kritisch reflektiert werden.

Etablierung gegenstandsadäquater Metriken für Forschungsproduktivität

- (5.1.)** Weitere Sensibilisierung der Öffentlichkeit (v.a. der Politik) für die mangelnde Validität gängiger Metriken und die Möglichkeit der Entwicklung validerer Metriken.
- (5.2.)** Öffentliche und verbindliche Distanzierung von der Nutzung wenig valider Metriken in der Wissenschaftsevaluation.
- (5.3.)** Spezifikation validerer Metriken.
- (5.4.)** Verwendung validerer Metriken als Evaluationskriterien in Zielvereinbarungen zwischen Professor:innen, Fakultäten und Uni-Leitungen.
- (5.5.)** Verwendung validerer Metriken in Rankings von Forschungseinrichtungen.
- (5.6.)** Ersatzlose Streichung der reinen *Anzahl* abgeschlossener Promotionen als Evaluationskriterium in allen zuvor genannten Zusammenhängen.
- (5.7.)** Verwendung validerer Metriken als Kriterien in Berufungsverfahren.
- (5.8.)** Öffentliche Selbstverpflichtung von Forschenden auf eine (niedrige) Maximalzahl von Erst- und Co-Autorenschaften pro Jahr.

Entlastung der Professor:innen und Refokussierung auf ihre Kernaufgaben

- (6.1.)** Verstärkte Einführung spezialisierter Stellen in den Fakultäten, die Professor:innen von delegierbaren Aufgaben (z.B. Studiengangskoordination, Lab Management) entlasten.
- (6.2.)** Zulassen einer flexibleren Profilbildung hinsichtlich der verschiedenen Aufgaben zwischen verschiedenen Professuren, sowie innerhalb von Professuren über die berufliche Laufbahn hinweg. Dies beinhaltet auch eine Flexibilisierung des Lehrdeputats.
- (6.3.)** Ermöglichung einer Besetzung von Professuren in Teilzeit, z.B. durch Jobsharing.
- (6.4.)** Sichtbarmachung der Überlastung der Professur durch öffentliche Benennung als Problem sowie durch systematische Forschung zum Thema.
- (6.5.)** Erhöhung der Grundfinanzierung der Hochschulen, um eine angemessene Personalausstattung zu ermöglichen.

Verbesserung der Beschäftigungsverhältnisse für ECRs

- (7.1.)** Tenure-Track-Stellen als Regelfall der postdoktoralen Weiterqualifikation.
- (7.2.)** Erhöhung der Zahl der Dauerstellen mit vollem oder erhöhtem Lehrdeputat für Promovierte.
- (7.3.)** Vergabe von Promotionsstellen für in der Regel 4 Jahre.
- (7.4.)** Die Datenerhebung im Rahmen von Forschungsprojekten sollte in der Regel von Wissenschaftlichen Hilfskräften geleistet werden.

Reduktion des Machtgefälles

- (8.1.)** Abgabe der Personalverantwortung für einzelne Mitarbeiter:innen an die Fakultät. Personalentscheidungen sollten von einem Gremium mit regelmäßig wechselnder Zusammensetzung getroffen werden.
- (8.2.)** Öffentliche Sichtbarmachung vorbildlicher Lösungen in dieser Hinsicht als Orientierung für andere Fakultäten.
- (8.3.)** Fakultätsinterne Entscheidungen über die Allokation von Forschungsressourcen (z.B. Laborkapazitäten) durch demokratisch legitimierte Gremien unter Beteiligung von ECRs.
- (8.4.)** Abgabe (eines Teils) der Coachingfunktion der Professur an Graduiertenschulen.
- (8.5.)** Verpflichtende externe Bewertung wissenschaftlicher Arbeiten als Regelfall.
- (8.6.)** Ein einziges bei einem wissenschaftlichen Journal zur Begutachtung eingereichtes Manuskript sollte als Promotionsgrundlage ausreichen.
- (8.7.)** Externe Begutachtung dieses Manuskripts (vgl. 8.6), unabhängig vom Peer-Review-Verfahren des Journals.
- (8.8.)** Abschaffung der Promotionsnoten.
- (8.9.)** Nachdenken über eine paritätischere Besetzung von Gremien der akademischen Selbstverwaltung mit Vertreter:innen der verschiedenen Statusgruppen.
- (8.10.)** Bei Einstellung: Feste Zusage eines individuellen Budgets für Kongressreisen, über das ECRs eigenständig verfügen können.
- (8.11.)** Einrichtung unabhängiger juristischer Beratung für ECRs und Finanzierung derselben über die Institution (z.B. Graduiertenschulen).
- (8.12)** Erarbeitung und Veröffentlichung einer juristisch abgesicherten Leitlinie zum Thema "Rechte an Forschungsdaten" durch die DGPs (und/oder die DFG).

Verbesserung der Peer-Review-Verfahren

- (9.1.)** Klare öffentliche Positionierung von Fachgesellschaften und Forschungsinstitutionen zugunsten einer deutlichen Reduktion des rein quantitativen Publikationsoutputs.
- (9.2.)** Reduktion der Menge zu begutachtender Manuskripte durch eine deutliche Erhöhung der methodischen Mindestanforderungen und "Desk Rejection" bei Nichterfüllung derselben.
- (9.3.)** Schnellstmöglicher Übergang zu Not-for-Profit-Publishing als Regelfall.
- (9.4.)** Öffentliche Erklärung dieses Wandels zu einem zentralen Entwicklungsziel.
- (9.5.)** Die Inhalte von Peer-Reviews sollten im Regelfall öffentlich gemacht werden.

(9.6.) Einsatz spezialisierter Expert:innen, die - zusätzlich zum inhaltlichen Review - alle eingereichten Arbeiten im Hinblick auf dieselben technischen Aspekte (z.B. Einhaltung von Prä-Registrierungen; Reproduzierbarkeit von Ergebnissen) bewerten.

(9.7.) Peer-Reviews sollten, wenn keine gewichtigen Gründe entgegenstehen, von den reviewenden Personen namentlich unterzeichnet werden.

(9.8.) Beobachtung möglicher unerwünschter Effekte einer solchen Entblindung.

(9.9.) Peer-Reviews sollten als eigenständiges Publikationsformat anerkannt, mit DOIs versehen und damit zitierbar werden.

(9.10.) Gezielte Förderung (z.B. durch Fast-Tracking) von Registered Reports.

Verbesserung der Berufungsverfahren

(10.1.) Sicherstellung des Vorliegens diagnostischer Expertise bei Mitgliedern von Berufungskommissionen.

(10.2.) Entlastungen in der Lehre oder Bewilligung zusätzlicher Personal- und Sachmittel, um gründliche Kommissionsarbeit zu ermöglichen und zu belohnen.

(10.3.) Verwendung von Kriterien Guter Wissenschaftlicher Praxis bei der Evaluation von Bewerber:innen im Rahmen von Berufungsverfahren.

(10.4.) Weiterentwicklung der Verfahren zur Erfassung von Lehrkompetenz im Rahmen von Berufungsverfahren.

(10.5.) Explizite Aufnahme von Verfahren zur Erfassung relevanter Persönlichkeitsmerkmale (z.B. Integrität) und überfachlicher Kompetenzen (z.B. Führung, Konfliktmanagement) in die Eignungsbeurteilung.

(10.6.) Perspektivisch: Implementierung des in der Wirtschaft bewährten 360°- Feedbacks als Maßnahme vor bedeutenden Karriereschritten.

(10.7.) Deutliche Erhöhung des Anteils externer Mitglieder in Berufungskommissionen und Auswahl derselben unabhängig von den Präferenzen der fakultätsinternen Mitglieder.

(10.8.) Ermittlung des wissenschaftlichen Forschungsstandes (und möglicher Forschungsbedarfe) zur Erfassung der persönlichen Integrität im Bewerbungskontext und Bereitstellung der Ergebnisse als Handreichung für die Community.

Schaffung bzw. Stärkung von Bewusstheit für Machtmissbrauch und Fehlverhalten

(11.1.) Angebot von Informationsveranstaltungen bzw. laufenden Fortbildungen zu Themen mit ethischen Bezug.

(11.2.) Diese Veranstaltungen sollten dezidiert psychoedukativen Charakter haben. Die Bearbeitung konkreter Fallbeispiele erscheint als Methode besonders geeignet.

(11.3.) Der Adressat:innenkreis für solche Maßnahmen umfasst *alle* Personen, die am Wissenschafts- und Lehrbetrieb beteiligt sind.

(11.4.) Der effizienteste Umgang mit dieser Aufgabe könnte darin bestehen, in Zusammenarbeit zwischen Kolleg:innen an verschiedenen Standorten gemeinsam ein entsprechendes Curriculum zu entwickeln.

(11.5.) Ausbildung von Multiplikator:innen, die (v.a. ihren jeweiligen Peers) explizit als Ansprechpersonen für ethische Belange zur Verfügung stehen.

Stärkung von Kontroll- und Sanktionsinstanzen

(12.1.) Spezifikation hochschulinterner Meldeprozesse und Bekanntmachung derselben auf leicht auffindbaren Webseiten.

(12.2.) Stärkung zuständiger Stellen durch angemessene personelle und finanzielle Ausstattung sowie ernstzunehmende Ermittlungs- und ggfs. auch Sanktionsbefugnisse.

(12.3.) Einrichtung institutionsexterner Instanzen auf Länder- oder Bundesebene.

(12.4.) Standardmäßige Beteiligung einer unabhängigen *externen* Aufsichtsperson zu Beginn jedes Beschwerdeverfahrens.

(12.5.) Ablösung des bisher verbreiteten Prinzips, interne Ombudsstellen mit einzelnen Personen zu besetzen, durch *Ombudsgremien* als Regelfall.

(12.6.) *Regelmäßige, proaktive Nachfragen* der Institutionen bei ihren Mitgliedern hinsichtlich Entwicklungsbedarf und Verbesserungsmöglichkeiten in Strukturen und Prozessen.

(12.7.) Erprobung der Möglichkeit *anonymer* Meldungen von Fehlverhalten.

(12.8.) Evaluationsforschung zur Effektivität bestehender Beschwerdeverfahren.

Differenzierte Analyse der identifizierten Problembereiche

Im Folgenden werden die zwölf im Laufe der Kommissionsarbeit identifizierten Problembereiche detaillierter beschrieben und analysiert. Dies beginnt mit der Darstellung einiger besonders problematischer Verhaltensweisen und deren Konsequenzen (P1 bis P4). Ihre Existenz stellt gewissermaßen den Ausgangspunkt aller weiteren Analysen dar. Es folgen Analysen bedeutsamer Charakteristika des akademischen Arbeitsumfeldes (P5 bis P13), die das Auftreten der vier zuvor genannten Probleme begünstigen (s. Abbildung 1). Aus diesen Analysen werden dann Empfehlungen abgeleitet, welche konkreten Maßnahmen jeweils zu einer Verbesserung der Situation führen könnten. Die meisten der relevanten Variablen sind mehr oder weniger eng miteinander verwoben. In vielen Fällen kann daher eine empfohlene Maßnahme möglicherweise zu Verbesserungen in gleich *mehreren* Bereichen führen - dies wird durch entsprechende Verweise sichtbar gemacht.

Problembereich 1. Nachlässiges oder betrügerisches Forschungshandeln (P1)

Das Problem / die Betroffenen

Forschung zielt auf die Erzeugung belastbaren, relevanten Wissens zum Wohle der Gesellschaft ab, die für diese Arbeit (u.a. Personalkosten, Materialien und Geräte) in erheblichem Umfang zahlt. Empirische Untersuchungen zeigen jedoch auf, dass es vielen veröffentlichten Studien an methodischer Qualität mangelt und dass Forschungsergebnisse oft nur bedingt oder gar nicht replizierbar sind (Open Science Collaboration, 2015). Das Problem betrifft auch andere Fächer (z.B. Prinz, Schlange & Asadullah, 2011), aber die Psychologie ist definitiv in erheblichem Ausmaß betroffen.

Die Ursache hierfür liegt vor allem im konkreten alltäglichen Verhalten der Forschenden. Dabei können natürlich wie bei jedem menschlichen Handeln unbeabsichtigte Fehler auftreten. Im aktuellen Kontext wichtiger sind aber vorsätzliche, sogenannte *Questionable Research Practices* (QRP) wie etwa das Beschönigen von Forschungsbefunden (z.B. das Ausgeben von Zufallsbefunden als aus der Theorie vorhergesagte Effekte). Die Grenze zwischen QRPs und Forschungsbetrug ist fließend. Explizit betrügerisches Verhalten wie etwa Datenerfindung scheint unter Wissenschaftler:innen deren eigenen Aussagen zufolge eher selten vorzukommen (John, Loewenstein & Prelec, 2012). Eindeutige Zahlen sind u.a. aufgrund der Brisanz des Themas aber schwierig zu bestimmen, und entsprechend ist die Dunkelziffer möglicherweise hoch. Um die Integrität der Befundlage in Frage zu stellen, reicht die weitverbreitete Anwendung von QRPs allerdings völlig aus.

Solche Probleme, die eher die Qualität der Daten und damit die Interpretierbarkeit der Ergebnisse beeinträchtigen, sind kein Spezifikum der Psychologie, sondern treten auch in vielen anderen Fächern auf. Die Psychologie hat darüber hinaus aber auch noch weitere qualitätsmindernde Schwachstellen wie etwa die mangelnde Präzisierung (Formalisierung) vieler ihrer Theorien und eine sehr unübersichtliche Konstruktlandschaft, in der es an verbindlichen Definitionen mangelt, derselbe Begriff für verschiedene Dinge verwendet wird (Jingle-Fallacy), oder verschiedene Begriffe für dasselbe Ding (Jangle-Fallacy) (Block, 1995; Flake & Fried, 2020). Aus dieser konzeptionellen Unschärfe resultiert eine weitgehende Ahistorizität: Inhaltliche Fragen, die das Fach bereits seit Jahren oder Jahrzehnten beschäftigen, werden unter Verwendung neuer Begrifflichkeiten immer wieder als "neu" eingeführt und aufgegriffen.

Unklare Konzepte und methodisch unzulängliche Forschungsergebnisse können für die Gesellschaft wenig Erkenntnisgewinn bei dringenden Fragen liefern. Außerdem beeinträchtigen diese Mängel die Qualität der Arbeit anderer Forscher:innen dadurch, dass diese ihre eigene Forschung auf unsoliden Vorbefunden aufbauen müssen und damit weitere (finanzielle, zeitliche und menschliche) Ressourcen nicht sinnvoll genutzt werden. Letztendlich leidet dann auch das öffentliche Vertrauen in die Psychologie und die empirischen Wissenschaften insgesamt, sowie die Identifikation der nächsten Generation von Forschenden mit dem eigenen Fach (Brachem et al. 2022).

Ursachen / beitragende Faktoren

Nachlässiges und betrügerisches Forschungshandeln wird aus Sicht der Kommission durch eine ganze Reihe problematischer Systemfaktoren begünstigt:

1) Es mangelt an einer konstruktiven Fehlerkultur. Unbeabsichtigte Fehler gehören zu jedem menschlichen Handeln und Lernen dazu. Um schnell und effizient aus Fehlern (z.B. im Experimentalprogramm oder bei der Datenverarbeitung) lernen zu können, ist jedoch eine offene Kommunikation darüber notwendig. Die Angst vor öffentlicher Blamage kann dieser wünschenswerten Offenheit im Weg stehen und zum Verheimlichen von Fehlern oder sogar zu absichtlicher Täuschung über geschehene Fehler führen.

2) Ein hoch relevanter Faktor ist in den Anreizstrukturen (P5) im Wissenschaftssystem zu sehen: Sowohl die Karrieren einzelner Forschender als auch die Evaluationen akademischer Institutionen orientieren sich derzeit stark an primär quantitativen Metriken wie Anzahl von Autorenschaften und Zitationen, Impact Faktoren, Drittmittelsummen und h-Indizes. Die *Validität* dieser Metriken als Maße für Forschungsproduktivität (insbesondere im Sinne der Generierung eines belastbaren Wissenszuwachses) ist jedoch zweifelhaft (Chapman et al., 2019; Leising et al., 2022a). Aufgrund ihrer ungebrochen hohen Relevanz im Rahmen von Berufungsverhandlungen (P10) ist es derzeit jedoch rational, das eigene Verhalten stark danach auszurichten, um die eigenen Metrik-Werte entsprechend zu maximieren. Mit einem solchen Fokus auf der Optimierung ungeeigneter Metriken leidet aber fast zwangsläufig die Qualität. Insbesondere für ECRs gilt es hier strukturelle Alternativen zu schaffen.

3) Die *Kosten* dafür, dass so viele wissenschaftlich wenig nützliche Befunde veröffentlicht werden, tragen nicht primär diejenigen, die sie veröffentlichen, sondern andere Forschende und die Öffentlichkeit, die den Forschungsbetrieb zum großen Teil finanziert.

4) Das Risiko einer Entdeckung ist vergleichsweise gering, da das Peer-Review eine Aufdeckung in der Regel nicht leisten kann (P9) und da die vorgesehenen Kontroll- und Sanktionsmechanismen bei Verstößen zu wenig wirksam sind (P12).

5) Die Überfrachtung der Professur mit Aufgaben (P6) ist ein weiterer Faktor, der zum Auftreten von QRPs beiträgt: Vielen Forschenden fehlen in Anbetracht ihrer institutionell angelegten chronischen Mehrfachbelastung schlicht die zeitlichen Ressourcen um qualitätsfördernde methodische Schritte zu unternehmen (z.B. Vorstudien zur Abschätzung von Effektgrößen, ausreichend spezifische Prä-Registrierungen, Replikationsversuche, gründliche theoretische Vorüberlegungen oder auch nur Literaturrecherche).

6) Das Auftreten bzw. Beibehalten von QRPs kann ebenso wie deren Nichtentdeckung auch durch die bestehenden Machtverhältnisse in Forschungsteams (P8) erheblich gefördert werden. So wird Fehlverhalten oft deshalb nicht gemeldet, weil potenzielle Hinweisgeber:innen Vergeltungsmaßnahmen z.B. in Form einer Nichtverlängerung befristeter Stellen befürchten (P7). In der Befragung von Elson et al. (2021) wurde sogar von Vorgesetzten berichtet, die Mitarbeitende explizit zu Datenmanipulationen auffordern.

Empfehlungen

- (1.1.)** Beginn einer Kulturentwicklung durch die Vergabe eines Gütesiegels ähnlich dem “great place to work“-Award - aber spezifisch bezogen auf Universitäten, Hochschulen und außeruniversitäre Forschungseinrichtungen (z.B. “great place to do research”). Wichtig ist, hier eine wirklich unabhängige Evaluation zuzulassen. Die relevanten Kriterien müssen ebenso wie der Prozess der Vergabe transparent und nachvollziehbar sein, kontinuierlich hinterfragt und ggfs. weiterentwickelt werden.
- (1.2.)** Bereitstellung und routinemäßige Nutzung ausgereifter Infrastrukturen für die Implementierung Guter Wissenschaftlicher Praktiken (wie etwa Prä-Registrierung und Open Data) durch Universitäten, Hochschulen und Forschungseinrichtungen, aber auch durch externe Akteure wie Forschungsinfrastrukturanbieter (z.B. ZPID, Open Science Framework).
- (1.3.)** Explizite Anerkennung und Berücksichtigung des Mehraufwandes, der durch Einhaltung höherer Forschungsstandards entsteht (z.B. größere Stichproben, Einplanen von Replikationsversuchen, dadurch längere Projektzeiten) im Rahmen von Verfahren zur Einwerbung von Drittmitteln (DFG, BMBF).
- (1.4.)** Ausbau des Trainings in GWP für individuelle Forschende, etwa als Routinebestandteil der Ausbildung von ECRs und als verpflichtende Weiterbildung für Professor:innen.
- (1.5.)** GWP und Open Science sollten fester Bestandteil der Lehre sein, nicht nur in der theoretischen Vermittlung, sondern auch in der Anwendung (z.B. im Rahmen von Experimentalpraktika und Abschlussarbeiten). Siehe hierzu auch die entsprechenden Forderungen der Psychologie-Fachschafftenkonferenz (PsyFaKo, 2018).
- (1.6.)** Das Psychologiestudium muss im Hinblick auf die Themen Theoriebildung, Theorie-spezifikation (Formalisierung) und Theorietestung deutlich gestärkt werden. Entsprechende Lehrveranstaltungen gehören ins Curriculum einer jeden Fakultät. Die Kommission empfiehlt, die hierfür notwendige Expertise ggfs. durch Kooperation mit anderen Fachbereichen (Mathematik, Philosophie, Physik, Informatik) zu erschließen.
- (1.7.)** Das Problem der chaotischen Begriffs- und Konstruktlandschaft in der Psychologie kann nur im Wege konzertierter Bemühungen um Dokumentation und allmähliche Konsensbildung in diesem Bereich gelöst werden. Aufgrund des hohen Komplexitätsgrades dieser Aufgabe (Leising et al., 2022a, 2022b) scheint es geboten, sie seitens Journals und/oder Fachgesellschaften explizit in den Blick zu nehmen, zu forcieren, ggfs. auch zu mandatieren und massiv zu inzentivieren (z.B. durch Schaffung hoher öffentlicher Sichtbarkeit).
- (1.8.)** Entwicklung und Spezifikation von Elementen einer konstruktiven Fehlerkultur. Hier kann man sich an bestehenden Vorbildern, etwa aus der Medizin, orientieren (<https://jederfehler-zaehlt.de>).
- (1.9.)** Gründung einer DGPs-Fachgruppe “Theoretische Psychologie”. Dabei sollte von Anfang an auf eine nachhaltige Vernetzung derselben mit anderen Fachgruppen geachtet werden: die neu zu gründende Fachgruppe soll nicht abgekoppelt von den inhaltlichen

Fragen arbeiten, sondern hilfreiche Beiträge zur stärker inhaltlichen Arbeit der anderen Fachgruppen liefern. Ein solides Fundament von Wissenschaftstheorie, Logik und Mathematik ist dafür notwendig, was eine regelmäßige Beteiligung entsprechend qualifizierter Kolleg:innen erfordert. Alternativ zur Gründung einer neuen Fachgruppe wäre an eine deutliche Stärkung dieses Themenbereichs innerhalb der Fachgruppe Methodenlehre zu denken (durch gezielte Ansprache und Einbezug von daran interessierten Kolleg:innen).

(1.10.) Weitere Bekanntmachung des "Negativkatalogs" der DFG zu Beiträgen, die keine Co-Autorenschaft an einem wissenschaftlichen Artikel rechtfertigen:

<https://wissenschaftliche-integritaet.de/kommentare/kriterien-der-autorschaft-negativkatalog/>

(1.11.) Standardmäßige Erklärung zu den tatsächlichen Beiträgen aller genannten Autor:innen ("authors' contributions") zu Veröffentlichungen (Allen, Scott, Brand, Hlava & Altman, 2014; <https://credit.niso.org/>). Dies einzufordern liegt im Verantwortungsbereich der Herausgeber:innen wissenschaftlicher Zeitschriften.

Öffentliche Selbstverpflichtung der Forschenden auf eine (niedrige) Maximalzahl von Erst- bzw. Co-Autorenschaften pro Jahr (siehe **5.8.**). Dies sollte helfen, der Qualität ein größeres Gewicht gegenüber der reinen Quantität des wissenschaftlichen Outputs zu verschaffen.

Problembereich 2. Machtmissbrauch (P2)

Das Problem / die Betroffenen

Befragungen an deutschen Hochschulen und Forschungsinstituten (Arcudi et al., 2019; Beadle et al., 2020; Elson et al., 2021; Olsthoorn et al., 2020; Peukert et al., 2020; Regler et al., 2019; Scherpenberg et al., 2021; Schraudner et al., 2019) sowie Falldarstellungen in Publikumsmedien zeigen, dass Machtmissbrauch von bereits im System etablierten Wissenschaftler:innen gegenüber ECRs, Studierenden und Verwaltungspersonal keine Ausnahme ist, sondern durchaus verbreitet. Bestimmte Gruppen (z.B. mit Migrationshintergrund, erste Generation Akademiker:in) sind dabei besonders gefährdet (Niemann et al., 2020).

Machtmissbrauch kann viele verschiedene Formen in unterschiedlichen Schweregraden annehmen: (a) Fälschliche Aneignung geistigen Eigentums der abhängigen Person, inklusive "Gast-" und "Ehrenautorenschaften", (b) Erpressen von Mehrarbeit (z.B. Übernahme von Lehre), (c) Einfordern problematischer Forschungspraktiken und Abschieben der Verantwortung für selbige auf die abhängige Person im Fall einer Entdeckung, (d) Demütigung und Erniedrigung, (e) Körperliche Misshandlung, sowie (f) sexuelle Belästigung. Neben den unmittelbar Betroffenen können unter solchen Vorkommnissen auch Kolleg:innen leiden (wenn sie ihre eigene Forschung z.B. auf fragwürdige Vorbefunde stützen müssen), sowie - bei Bekanntwerden - auch die Reputation einzelner akademischer Institutionen oder gar der Wissenschaft insgesamt.

Ursachen / beitragende Faktoren

Machtmissbrauch im akademischen Kontext ergibt sich häufig aus einem *Zusammenspiel* mehrerer Faktoren:

1) Persönliche Dispositionen (insbesondere solche der "dunklen Triade": Narzissmus, Machiavellismus und Psychopathie; Paulhus & Williams, 2002) können Machtmissbrauch wahrscheinlicher machen. Die Forschung zu Führungspersönlichkeiten im Wirtschaftskontext zeigt klar die Relevanz dieser Merkmale (Palmer et al., 2020); ihre Bedeutung im Hochschulkontext wird hingegen erst seit kurzem systematisch betrachtet (Cohen & Baruch, 2022). Standardverfahren der Personalauswahl für akademisches Führungspersonal erfassen diesen Bereich bisher nicht systematisch (P10).

2) Strukturelle Eigenschaften des Wissenschaftssystems (vor allem in Kombination mit den o.g. Personenmerkmalen) können Machtmissbrauch begünstigen (Bramlage et al., 2019; Cohen & Baruch, 2022; Lasser et al., 2021). Ein zentraler Grund dafür ist die Bündelung unnötig vieler Entscheidungsbefugnisse bei den Professor:innen (P8). Die daraus resultierende, erhebliche Entscheidungsmacht macht es für potenzielle Beschwerdeführer:innen riskant, auf Missstände hinzuweisen, denn sie müssen Vergeltungsmaßnahmen wie "schlechte Bewertung" oder "ausbleibende Vertragsverlängerung" von Seiten der beschuldigten Person fürchten (Horbach et al., 2020).

3) Es gibt bislang kaum effektive Kontrollinstanzen (P12) im Fall von missbräuchlichem Machteinsatz, da die bestehenden Instanzen personell und materiell schlecht ausgestattet sind, über geringe Befugnisse und kaum Sanktionsgewalt verfügen, und den potenziellen Beschwerdeführer:innen zudem oft gar nicht bekannt sind. Auch zeigen Befragungen, dass zumindest den *internen* Beschwerdewegen wenig Vertrauen entgegengebracht wird, und dass einmal angestoßene Verfahren selten Ergebnisse bringen, die von den Beschwerdeführenden als zufriedenstellend erlebt werden (Elson et al., 2021; Höbel, Durglishvili, Reinold & Leising, 2022). Verfehlungen bleiben daher für die Verantwortlichen oft konsequenzlos. Das Wissen um die geringe Wirksamkeit der bestehenden Kontrollmechanismen kann dazu führen, dass manche Personen es eher wagen, ihre Macht zu missbrauchen.

4) Ein weiterer wichtiger Faktor, der Machtmissbrauch begünstigt, ist in der mangelnden Bewusstheit für den Problembereich zu sehen (P11). Im Laufe der Jahrzehnte scheint sich die Anschauung etabliert zu haben, dass Professor:innen ihren Mitarbeiter:innen in nahezu jeder Hinsicht sehr viel abverlangen dürfen, da der Weg zur Professur “nun einmal steinig” sei (“das war schon immer so”, “da mussten wir alle mal durch”). Die Grenze zwischen legitimem Machtgebrauch und Machtmissbrauch kann so weiter schwimmen.

Empfehlungen

Regelmäßige Befragungen (der Pre- und Postdocs) zum Führungsverhalten von Vorgesetzten und differenziertes Feedback dazu an selbige (siehe **3.5.**, **3.7.**).

Verstärkte Nutzung von Betreuungsvereinbarungen (siehe **3.4.**).

Geänderter Umgang mit Befristungen (siehe **7.1.**, **7.2.**, **7.3.**).

Verteilung der bei der Professur gebündelten Funktionen (siehe **8.1.** bis **8.5.**, **8.9.**, **8.10.**).

Einrichtung kostenloser, unabhängiger juristischer Beratung für ECRs (siehe **8.11.**).

Erarbeitung und Veröffentlichung einer juristisch abgesicherten Leitlinie zum Thema “Rechte an Forschungsdaten” durch die DGPs (und/oder die DFG) (siehe **8.12.**).

Anerkennung der wichtigen Rolle von Persönlichkeitsmerkmalen (Integrität) und Führungsqualifikationen in der Personalauswahl und systematische Erfassung dieser Merkmale durch geeignete Methoden in Berufungsverfahren (siehe **10.5.**, **10.6.**).

Schaffung von “Awareness” für die Themen Macht und Machtmissbrauch auf allen Systemebenen (siehe **11.1.** bis **11.5.**).

Stärkung von Kontroll- und Sanktionsinstanzen durch Zuweisung von mehr Ressourcen, durch Ausweitung der Befugnisse und durch verstärkte Beteiligung externer Akteur:innen (siehe **12.1.** bis **12.8.**).

Problembereich 3. Unzureichende Betreuung von ECRs (P3)

Das Problem / die Betroffenen

Manche ECRs erleben einen gravierenden Mangel an kompetenter und verantwortungsvoller Betreuung (Mackie & Bates, 2018). Dies kann sowohl fachliche Aspekte betreffen (z.B. Gedrängtwerden zu schlechter wissenschaftlicher Praxis) als auch soziale und solche der Arbeitsgestaltung (Planlosigkeit in der Projektdurchführung, chronische Unerreichbarkeit der betreuenden Person, mangelnder Zugang zu benötigten Ressourcen, Desinteresse an den Forschungsthemen der ECRs oder ihrer persönlichen und fachlichen Entwicklung).

Schlechte Betreuung schädigt zunächst die betroffenen ECRs in fachlicher und/oder gesundheitlicher Hinsicht. Langfristig schädigt sie potenziell aber auch den wissenschaftlichen Prozess, weil mangelhafte Anleitung zu ineffizienter und schlechter Arbeit führen kann. Das Bekanntwerden schlechter Betreuung schadet zudem der Reputation von Professuren, Fakultäten und Hochschulen. Negative Erfahrungen in der Betreuung können schließlich auch dazu beitragen, dass sich talentierte ECRs dazu entscheiden, die Wissenschaft zu verlassen (Kis et al., 2022; Christian et al., 2021), sodass deren wissenschaftliche Qualifikation der Wissenschaft nicht mehr zur Verfügung steht.

Ursachen / beitragende Faktoren

- 1) Die Überfrachtung der Professur mit Aufgaben (P6) spielt eine wichtige Rolle: Etablierte Wissenschaftler:innen sind oft mit so vielen Dingen gleichzeitig beschäftigt, dass eine qualitativ hochwertige Betreuung durch sie nur bei einer kleinen Zahl von ECRs möglich ist.
- 2) Es wird im geltenden Anreizsystem (P5) belohnt, *möglichst viele* Doktorand:innen zum Abschluss zu bringen, ohne dass dabei die Frage nach *Betreuungsqualität* gestellt würde.
- 3) Das enorme Machtgefälle zwischen etablierten Forschenden und ECRs (P8) kann dazu führen, dass ECRs Kritik an unzureichender Betreuung - wenn überhaupt - nur zurückhaltend äußern.
- 4) Eine relativ große Unklarheit besteht darüber, was überhaupt "gute Betreuung" ausmacht (siehe P11). Es fehlen einheitliche Standards und ECRs haben in der Regel keine klare Vorstellung davon, welche Quantität und Qualität von Betreuung sie erwarten dürfen.
- 5) Personalführungs- und Betreuungskompetenzen spielen in der Ausbildung auf dem Weg zur Professur und bei der Beurteilung von Bewerbungen im Rahmen von Berufungsverfahren (P10) bisher keine nennenswerte Rolle.
- 6) Oft ist unklar, an wen sich Betroffene wenden können, wenn sie mit der eigenen Betreuung unzufrieden sind, und wie der Prozess aussähe, der in einer solchen Situation für Abhilfe sorgen könnte. Auch sind Verstöße durch die Institution schwer zu ahnden, da es dafür keine explizit vorgesehenen Maßnahmen gibt.

Empfehlungen

(3.1.) Öffentliche Erklärung (der Hochschulen, der DGPs, der die Forschung fördernden Einrichtungen, z.B. DFG, BMBF) zu einer plausiblen Maximalzahl von ECRs, die gleichzeitig mit derselben betreuenden Person wissenschaftlich zusammenarbeiten können.

(3.2.) Die Kommission hält es für unrealistisch, dass eine einzelne Person für viele ECRs gleichzeitig eine angemessene Betreuung gewährleisten kann. Deshalb sollte z.B. eine große Anzahl von Promovierenden nur unter besonderen Bedingungen (z.B. Ko-Betreuung) zugelassen und die Qualität der Betreuung dann regelmäßig und sorgfältig überprüft werden (z.B. durch Befragung der Betroffenen, siehe **3.7.**).

(3.3.) Formulierung eines Leitbildes zur Zusammenarbeit von ECRs und Professor:innen.

(3.4.) Verstärkte Nutzung von Betreuungsvereinbarungen, um die Berechenbarkeit und Steuerbarkeit der wissenschaftlichen Zusammenarbeit für ECRs zu erhöhen. Darin sollten wesentliche Parameter der Zusammenarbeit (vorläufiges Forschungsthema, Modalitäten von Themenänderung und Betreuungswechsel, Zugang zu Ressourcen, Nutzungsrechte an Daten, Autorenschaften, Arbeitsberichte, Feedback) festgehalten werden, ebenso wie die zuständigen Stellen bei Konflikten über diese Punkte. Da Betreuungsvereinbarungen selbst auch als Machtmittel missbraucht werden können, sollte ihre Verwendung jedoch kritisch begleitet und kontrolliert werden.

(3.5.) Regelmäßiges Feedback an Professor:innen bzgl. der Qualität der Zusammenarbeit mit ECRs durch Letztere (z.B. im Rahmen von halbjährlichen Mitarbeitendengesprächen oder als Ergebnis von systematischen Befragungen).

(3.6.) Explizite, rechtsverbindliche Regelung des Verfahrens zum Betreuungswechsel in der Promotionsordnung.

(3.7.) Regelmäßige Befragung der ECRs zu ihren Betreuungserfahrungen durch eine unabhängige Instanz (hier bieten sich Graduiertenschulen an), ab dem Beginn der Zusammenarbeit. Dies sollte auch Fragen zu den Gründen einer evtl. abgebrochenen Promotion beinhalten.

(3.8.) Einführung von Thesis Advisory Committees (TACs) als Regelmodell der Promotionsbetreuung. Hier begleiten neben der hauptamtlich betreuenden Person mehrere weitere Personen den inhaltlichen Fortschritt eines Promotionsprojekts im Rahmen regelmäßiger Austausche darüber. Vorausgesetzt, dass eine ausreichende Verbindlichkeit dieses Vorgehens durch die Betreuungsvereinbarung (s.o.) sichergestellt ist, sollte dies zu einer Erhöhung der Betreuungsdichte und der inhaltlichen Qualität beitragen. Wenn TACs Mitglieder von anderen Universitäten und/oder Fachrichtungen umfassen (was wünschenswert erscheint), sollten dadurch auch die überregionale Vernetzung und die Interdisziplinarität gefördert werden können. Ein positiver Nebeneffekt könnte die Begrenzung von Machtmissbrauch (siehe P8) infolge erhöhter Transparenz (Mehraugenprinzip) sein.

Abschaffung der reinen *Anzahl* abgeschlossener Promotionen als Evaluationskriterium, da diese Praxis vor allem niedrige wissenschaftliche Maßstäbe incentiviert (siehe **5.6.**).

Entlastung der Professur durch Abgabe bzw. Verteilung von Funktionen (siehe **8.1.** bis **8.5.**). Dies sollte auch Ressourcen für eine bessere Betreuung von ECRs durch Professor:innen freisetzen helfen.

Übertragung eines Teils der Betreuungsaufgaben, die nicht fachspezifisch sind (z.B. Ausbildung von Schlüsselqualifikationen, strategische Karriereberatung) auf dafür besonders qualifizierte Personen, eventuell koordiniert durch Graduiertenschulen (siehe **8.4.**).

Die Bewertung der Arbeitsergebnisse von ECRs sollte durch Dritte (möglichst extern) erfolgen, also unabhängig von der wissenschaftlichen Zusammenarbeit (siehe **8.5.** bis **8.7.**). Dies reduziert das Machtgefälle und schafft für Betreuer:innen einen Anreiz, in die Projekte der ECRs, die mit ihnen zusammenarbeiten, zu investieren.

Stärkere Beachtung von Betreuungskompetenz und Betreuungsbereitschaft als Kriterien in Berufungsverfahren z.B. durch Einholen von Referenzen oder Befragungen (siehe **10.5.**, **10.6.**).

Explikation des Beschwerdeweges und möglicher Sanktionen bei schlechter Betreuung sowie Bekanntmachung bei allen Beteiligten (siehe **12.1.**).

Problembereich 4. Mangelhafte Qualität der Lehre (P4)

Das Problem / die Betroffenen

Lehre ist eine der zentralen Aufgaben von Hochschulen. Jedoch sind Klagen über eine mangelhafte Qualität der akademischen Lehre nicht selten. Bemängelt werden unter anderem (a) mangelndes Engagement des Lehrpersonals, (b) veraltete Inhalte, (c) Inhalte von fraglicher oder offenkundig geringer Relevanz (z.B. im Hinblick auf eine spätere Berufstätigkeit), (d) wenig lernförderliche Darbietungsformate, (e) Redundanzen sowie (f) ungeeignete Prüfungsformate.

Leidtragende dieser Defizite sind vor allem die Studierenden, die nicht nur mit unnötigen Frustrationen und Hindernissen im Studienverlauf zu kämpfen haben, sondern auch mit dem Gefühl, in ihrer Rolle und Position nicht wertgeschätzt zu werden. Nicht zuletzt kann so der Lernerfolg hinter dem zurückbleiben, was wünschenswert wäre. Leidtragende sind aber auch diejenigen Dozent:innen, die sich mehr als notwendig in der Lehre engagieren, da sie für ihre Bemühungen im Rahmen der aktuell geltenden Anreizstrukturen bislang kaum oder gar nicht belohnt werden.

Ursachen / beitragende Faktoren

1) Qualitativ hochwertige Lehre wird im Hochschulbetrieb bislang kaum incentiviert. Zwar sind Lehrevaluationen inzwischen fester Bestandteil des Lehrbetriebs, kritische Evaluationen haben in aller Regel aber keine ernstzunehmenden Konsequenzen. Dem Lehrpersonal bleibt es vielmehr selbst überlassen, ob bzw. welche Schlüsse es jeweils aus den Rückmeldungen ziehen möchte. Die einfachste Variante ist, keine Schlüsse daraus zu ziehen und weiterzumachen wie bisher.

2) Lehrproben spielen in den meisten Berufungsverfahren eine untergeordnete Rolle. Dies kann als Widerspiegelung der geltenden Anreizstrukturen angesehen werden, in denen Publikationen, Zitationen und Drittmittel eine viel größere Bedeutung beigemessen wird.

3) Ein weiterer problematischer Aspekt besteht darin, *wie* die Lehrqualität bestimmt wird. Lehrqualität wird bislang nicht an der Erreichung langfristiger Lernziele gemessen, welches die eigentliche Aufgabe der Lehre ist, sondern an modulbezogenen Prüfungsleistungen (z.B. Bestehensraten) und Studierendenevaluationen. Solche Maßstäbe sind fehleranfällig:

Prüfungsgegenstände und Bewertungsmaßstäbe werden in der Regel von denjenigen Personen festgelegt, die auch die Lehrveranstaltungen abhalten. Diese Verquickung erlaubt es den Prüfenden, Leistungsüberprüfungen nach eigenem Ermessen einfacher oder schwerer zu gestalten. Hierdurch kann eine Situation entstehen, in der Dozierende hohe Bestehensraten, gute Prüfungsnoten, gute Lehrevaluationen und eine geringe Zahl von Beschwerden dadurch erreichen können, dass sie die Maßstäbe und Anforderungen entsprechend absenken (Prinzip "Gute Noten gegen schlechte Lehre").

4) Ein letzter wichtiger Faktor ist die Überfrachtung der Professur mit Aufgaben (P6). In Anbetracht der Menge an Aufgaben, die Professor:innen regelmäßig übertragen werden, ist es wenig überraschend, wenn diese ihr Engagement an derjenigen Stelle reduzieren, an der es aufgrund der Anreizstrukturen (P5) am wenigsten unangenehme Konsequenzen für sie selbst haben wird, das heißt, u.a. in der Lehre.

Empfehlungen

- (4.1.)** Schaffung expliziter, wirksamer (auch finanzieller) Anreize für Engagement und überdurchschnittliche Qualität in der Lehre.
- (4.2.)** Explizit stärkere Gewichtung von qualitativ guter Lehre in Berufungsverfahren und im Rahmen der Evaluation von Fakultäten und Universitäten.
- (4.3.)** Ausbau der hochschuldidaktischen Weiterbildung für Dozierende, inklusive solcher in der ECR-Phase. Ermunterung oder sogar Verpflichtung zur regelmäßigen Teilnahme.
- (4.4.)** Proaktive Ansprache individueller Lehrender durch die Institution, verbunden mit dem Angebot einer freiwilligen externen Evaluation und Beratung durch Expert:innen.
- (4.5.)** Ermunterung der Dozierenden im selben Fachgebiet zu stärkerer inhaltlicher Abstimmung und ggfs. zum Austausch von Lehrmaterialien untereinander. Solcher Austausch könnte qualitätsfördernd wirken und sogar bis zur Entwicklung eines gemeinsamen Kerncurriculums und von Open Educational Resources (d.h. frei verfügbaren Lehrmaterialien) führen. Hier könnte die DGPs eine koordinierende Rolle einnehmen.
- (4.6.)** Das Vorgehen bei der Lehrevaluation (v.a. durch Studierende) sollte kritisch hinterfragt und optimiert werden. Entsprechende Vorschläge wurden z.B. von Kreitzer und Sweet-Cushman (2021) gemacht.
- (4.7.)** Auch die Frage nach angemessenen Prüfungsformen in der Psychologie sollte kritisch reflektiert werden. Sehr regelmäßig sind z.B. Klagen über "Referate-Seminare" und Multiple-Choice-Klausuren zu vernehmen, ohne dass diesen ausreichend spezifizierte Verbesserungsmöglichkeiten gegenübergestellt würden.

Die Kommission empfiehlt die Bearbeitung der beiden letztgenannten Fragen durch die Kommission "Studium und Lehre", die hier über besondere Expertise verfügt.

Entlastung der Professor:innen von Aufgaben, die nicht zum Kern ihres Stellenprofils (Forschung und Lehre) gehören (siehe **6.1.**).

Zulassen einer stärkeren Profilbildung der Professor:innen durch Flexibilisierung der individuellen Gewichtung von Forschung vs. Lehre (siehe **6.2.**).

Erhöhung der Lehrkapazität der Fakultäten durch Schaffung von mehr Dauerstellen oberhalb der Promotion mit vollem oder erhöhtem Lehrdeputat (siehe **7.2.**).

Problembereich 5. Fehlanreize (P5)

Das Problem / die Betroffenen

Wissenschaftler:innen konkurrieren regelmäßig miteinander um begrenzte Ressourcen wie Stellen und Fördermittel. Um dabei möglichst objektive Entscheidungen über die Vergabe zu ermöglichen, hat sich der Einsatz sogenannter Metriken von Forschungsproduktivität fest etabliert. Den aktuell gängigen Formen der Metrisierung anhand primär *quantitativer* Indizes wie Publikations- und Zitationszahlen, Drittmittelsummen, Impact Faktoren und h-Indizes mangelt es jedoch an Validität, sodass ihr Einsatz den eigentlichen Zielen der Wissenschaft oft sogar eher schadet (Abele-Brehm & Bühner, 2016a; vgl. auch EU-Direktive 2019/1937, 2019; Leising et al., 2022a, b). Es ist durchaus so, dass inhaltlich starke Beiträge zum Wissensstand eines Faches gute Chancen haben, "hochrangig" publiziert und dann auch viele Male zitiert zu werden. Leider werden sehr hochrangige Metrik-Ergebnisse aber auch von inhaltlich sehr schwachen Arbeiten erzielt. Das liegt auch daran, dass die Qualitätssicherung durch Peer-Review bislang nicht ausreichend funktioniert (P9). Zu den Folgen dieser Situation gehören (in der Psychologie ebenso wie in vielen anderen Fächern) eine zweifelhafte wissenschaftliche Befundlage (P1) sowie Verzerrungen zu Ungunsten gewissenhafter Forscher:innen in akademischen Personalauswahlprozessen (P10).

Viele der gängigen Metriken sind zudem sehr leicht manipulierbar und werden auch fleißig manipuliert (Chapman et al., 2019; Falagas & Alexiou, 2008). Beispielsweise ist Umfragen unter Forschenden zufolge der Erwerb von sogenannten Co-Autorenschaften im Wege gegenseitiger Gefälligkeiten oder bedingt durch institutionelle Abhängigkeiten eine häufige Form wissenschaftlichen Fehlverhaltens (Fong & Wilhite, 2017; Gasparyan, Ayyazyan & Kitas, 2013; Pruschak & Hopp, 2022). Der *individuelle* Nutzen dieser Praxis ist vor dem Hintergrund der bestehenden Anreize unbestreitbar. Der tatsächliche Beitrag, der sich hinter einer solchen Co-Autorenschaft verbirgt, kann aber zwischen mehrjähriger intensiver Arbeit und exakt null variieren. Die *Kosten*, inklusive des Entdeckungs- und Sanktionsrisikos (P12), liegen dabei für diejenigen, die so genannte Gast- oder Ehrenautorenschaften einfordern oder gewähren, ebenfalls nahe null. Der Indikator "Anzahl von (Co-)Autorenschaften" ist daher als sehr problematisch zu betrachten.

Fehlanreize gibt es auch im Bereich der Drittmittelförderung. Drittmittelerfolg scheint an manchen Standorten das einzige Kriterium zu sein, an dem die Forschungsleistung von Professor:innen (etwa im Rahmen von Zielvereinbarungen) gemessen wird. Aus den eingeworbenen Drittmittelsummen ergeben sich anteilig Overhead-Beträge für die Universitäten und Hochschulen, die damit ihre - oft als deutlich zu knapp beschriebene - Grundfinanzierung aufstocken können. Hochschulen und Forschungseinrichtungen haben daher ein Interesse daran, *möglichst große finanzielle Ressourcen für die Forschung zu verbrauchen*. Entsprechend werden Forschende belohnt (z.B. mit Gehaltszuschlägen), die den Hochschulen helfen, dieses Ziel zu erreichen. Besonders leicht zu erreichen ist dies mit Forschung, die *per se* schon teurer ist, etwa weil sie z.B. den Einsatz teurer Geräte erfordert.

Die derzeitige Anreizstruktur orientiert sich primär an der Quantität und misst dem erzielten inhaltlichen Fortschritt eine zu geringe Bedeutung bei. Tatsächlich lässt sich darstellen, wie diese Form der Metrisierung der Qualität der Forschung systematisch schadet. Dies ist zum Beispiel der Fall, wenn Professor:innen, Fakultäten und Hochschulen für die reine *Anzahl* der bei ihnen abgeschlossenen Promotionen oder Habilitationen belohnt werden (etwa durch bessere Ranking-Ergebnisse), die Entscheidung über die Qualitätskriterien hierfür aber allein bei eben jenen Professor:innen, Fakultäten und Hochschulen liegt. Dadurch entsteht eine hohe Motivation zur Aufwands- und Anspruchsminimierung.

Dies geht auf Kosten (a) der Allgemeinheit, die die Arbeit der Forschenden finanziert, deren Anspruch auf einen inhaltlich soliden "Return-on-Investment" aber nicht eingelöst wird, und (b) derjenigen Forschenden, die höhere inhaltliche Ansprüche an die eigene Arbeit stellen, und/oder deren Forschung *weniger* Ressourcen verbraucht, da diese geringere Werte in den bisher gängigen Metriken erreichen.

Ursachen / beitragende Faktoren

1) Die Ursachen für die Verwendung der aktuellen, weitgehend dysfunktionalen Metriken sind vielfältig: Eingeführt wurden sie einmal in bester Absicht, nämlich mit dem Ziel einer stärkeren *Objektivierung* bei der Erfassung wissenschaftlicher Produktivität. Die Schwäche des Systems besteht darin, dass zu großes Vertrauen in die Redlichkeit der Forschenden und das Funktionieren des Peer-Review-Systems und weiterer Kontroll- und Sanktionsmechanismen gesetzt wird. Es wurde offenkundig unterschätzt, wie leicht die gängigen Metriken manipuliert werden können, und wie bereitwillig viele genau dies auch tun würden (Chapman et al., 2019; Fong & Wilhite, 2017; John et al., 2012; Pruschak & Hopp, 2022).

2) Nicht selten rechtfertigen die in diesem System arbeitenden Wissenschaftler:innen ihre Beteiligung damit, dass sie mehr oder weniger machtlos seien, die geltende Form der Metrisierung zu durchbrechen, obwohl sie die lauter werdende Kritik daran eigentlich teilen. Professor:innen verweisen auf Fakultäts- oder Hochschulleitungen, und diese wiederum auf Ranking-Agenturen. An dieser Stelle wird die "multiple Situiertheit" des Verhaltens Einzelner im akademischen System besonders gut sichtbar.

3) Ein weiterer Faktor besteht in den kommerziellen Interessen der Wissenschaftsverlage, die der wissenschaftlichen Qualität der von ihnen veröffentlichten Artikel möglicherweise geringeres Gewicht beimessen als dem Verkauf eines möglichst hohen Volumens von Artikeln (z.B. an Universitätsbibliotheken).

Empfehlungen

- (5.1.)** Sensibilisierung der Öffentlichkeit - und insbesondere der Politik - für die mangelnde Validität gängiger Metriken und die Möglichkeit, validere Metriken zu entwickeln und zu verwenden. Diese Sensibilisierung könnte z.B. durch Wortmeldungen in Publikumsmedien erfolgen, aber auch durch Eingaben bei Parlamenten.
- (5.2.)** Öffentliche und verbindliche Distanzierung von der Nutzung invalider Metriken in der Wissenschaftsevaluation (Akteur:innen: einzelne Professor:innen, Uni-Leitungen, DFG, Ranking-Agenturen).
- (5.3.)** Spezifikation validerer Metriken (Leising et al., 2022a, 2022b; Schönbrodt et al., 2022). Bezüglich der Umsetzung erscheint die standardisierte Kenntlichmachung bestimmter methodischer Gütekriterien auf einzelnen Forschungsartikeln am zielführendsten. "Badges" für Open Data, Prä-Registrierung etc. gibt es bereits, aber weitere werden benötigt - etwa für repräsentative Daten, ausreichende Power, und durchgeführte Reproduzierbarkeits-Checks. Kennwerte einzelner Artikel können dann bei Bedarf über Autor:innen oder Institutionen aggregiert werden.
- (5.4.)** Verwendung validerer Metriken als Evaluationskriterien in Zielvereinbarungen zwischen Professor:innen, Fakultäten und Uni-Leitungen.
- (5.5.)** Verwendung validerer Metriken als Evaluationskriterien in Rankings von Forschungseinrichtungen.
- (5.6.)** Ersatzlose Streichung der reinen *Anzahl* abgeschlossener Promotionen als Evaluationskriterium in allen zuvor genannten Zusammenhängen.
- (5.7.)** Verwendung validerer Metriken als Kriterien in Berufungsverfahren (siehe auch **10.3.**).
- (5.8.)** Öffentliche Selbstverpflichtung von Forschenden auf eine (niedrige) Maximalzahl von Erst- und Co-Autorenschaften pro Jahr (Frith, 2020). Dies wäre ein wirksames Mittel zur Durchbrechung des Drucks zu immer noch mehr Publikationsbeteiligungen und würde für Forschende einen Anreiz schaffen, sich auf die inhaltlich vielversprechendsten Projekte zu konzentrieren. Zudem würde so sichtbar gemacht, welche Menge qualitativ hochwertigen Outputs pro Jahr realistisch ist, und die Menge zu begutachtender Manuskripte würde sinken, was wiederum gründlichere Peer-Reviews ermöglichen sollte (siehe P9).

Problembereich 6. Überfrachtung der Professur mit Aufgaben (P6)

Das Problem / die Betroffenen

Die aktuell gängigen Erwartungen an die Leistungsfähigkeit von Professor:innen werden von vielen als *deutlich zu hoch* erlebt: Professor:innen sollen in der Regel (a) qualitativ hochwertige Lehrleistungen im Umfang von 8 oder 9 SWS pro Semester erbringen. Dazu gehört es, Lehrmaterialien aktuell zu halten und sie ggf. auch parallel im digitalen Format anzubieten, sowie die eigene Lehre evaluieren zu lassen und dabei erkannte Verbesserungsmöglichkeiten umzusetzen. Die mit Lehrveranstaltungen verbundenen Prüfungen für eine große Anzahl Studierender pro Jahr bedeuten einen hohen Arbeitsaufwand (Vorbereitung und Korrektur bei Klausuren; Prüfungszeit bei mündlichen Einzelprüfungen). Teil der Lehre ist außerdem die Anleitung von studentischen Abschlussarbeiten, was regelmäßige Treffen mit Studierenden ebenso bedeutet wie das Gegenlesen und Kommentieren von Entwürfen und das abschließende Lesen und Bewerten der fertigen Arbeiten. Auch gehört ein nicht zu vernachlässigender administrativer Anteil zur Lehre (z.B. Noten verbuchen, archivieren).

Professor:innen sollen sich zudem (b) in der akademischen Selbstverwaltung engagieren, was regelmäßige Teilnahmen an Sitzungen diverser Gremien (Fakultätsrat, Senat, Lehrkommission, Prüfungsausschuss, Berufungskommission, Strukturkommission, Bibliothekskommission, Haushaltskommission, Promotionsausschuss, Ethikkommission, Open Science Kommission, Raumkommission, Forschungskommission, Zulassungskommission) sowie das Studium und ggfs. die Überarbeitung von umfangreichen Dokumenten wie etwa Studienordnungen oder Evaluationsberichten bedeutet.

Weiterhin sind Professor:innen (c) sowohl die hauptamtlichen Personal- als auch die Budgetverantwortlichen ihres Arbeitsbereiches. Dies erfordert kontinuierliches Monitoring der zur Verfügung stehenden Gelder, das Anschieben und Genehmigen von Beschaffungsvorgängen, die Inventarisierung und die Verwaltung des Zugangs zu räumlichen und materiellen Ressourcen sowie die Ausschreibung und Besetzung von Stellen. Letzteres wiederum impliziert das Studium von Bewerbungsunterlagen, das Führen von Einstellungsinterviews, eventuell das Einholen weiterer Referenzen, das Treffen von Entscheidungen und eine Rückmeldung derselben an Kandidat:innen.

Die Personalverantwortung der Professor:innen erstreckt sich auf die Führung aller wissenschaftlichen und evtl. nicht-wissenschaftlichen Mitarbeiter:innen (durch Zielsetzungen, Evaluationen, Förderung, Karriereberatung sowie evtl. das Erkennen und das Management von Konflikten). Sie bedeutet zudem eine kontinuierliche Beschäftigung mit Formalia von Anstellungsverhältnissen (Aufstockung, Umsetzung, Teilzeit, Verlängerung, Elternzeit) und - nach dem Ausscheiden von Angestellten aus der Institution - das Abfassen von Arbeitszeugnissen, oft in Abstimmung mit weiteren Akteur:innen aus der Verwaltung.

Über die Erfüllung all dieser Erwartungen hinaus sollen Professor:innen zudem (d) exzellente und international konkurrenzfähige Forschungsleistungen erbringen. Diese erfordern langfristige (oft mehrjährige), intensive inhaltliche Arbeit auf hohem intellektuellen Niveau

und in ständiger Abstimmung mit wechselnden Teams von Kolleg:innen innerhalb oder außerhalb der jeweiligen Institution. Die wissenschaftliche Befundlage muss rezipiert, Projektskizzen müssen verfasst und bei Drittmittelgebern eingereicht werden. Forschungsvorhaben von Mitgliedern der eigenen Arbeitsgruppe müssen besprochen und laufend optimiert werden. Zunehmend spielen auch Stellungnahmen von Ethik-Gremien eine Rolle. Um diese zu erhalten, müssen neben den Projektskizzen oft weitere Unterlagen erstellt und eingereicht werden. Datenerhebungen dauern oft Monate oder Jahre und müssen initiiert und monitoriert werden. Projektgelder müssen verausgabt und korrekt abgerechnet werden. Daten müssen ausgewertet und interpretiert, Fehler dabei entdeckt und korrigiert werden. Ergebnisse müssen auf Kongressen vorgestellt und in Form von wissenschaftlichen Aufsätzen bei Zeitschriften eingereicht werden.

Wissenschaftliches Publizieren erfordert in der Regel umfangreiche (oft mehrfache) Überarbeitungen von Manuskripten im Prozess des Peer-Reviews. Die Abfassung und Optimierung eines einzigen, hochwertigen Manuskriptes bis zur Publikationsreife bedeutet meist *Hunderte* Stunden Arbeit. Die derzeit zu beobachtende - und absolut wünschenswerte - Entwicklung hin zu Open Science hat die hier relevanten Anforderungen noch einmal deutlich erhöht. Selbstverständlich treten Professor:innen regelmäßig auch auf der "anderen Seite" des wissenschaftlichen Prozesses in Erscheinung, nämlich als Gutachter:innen für Manuskripte, Drittmittelanträge und Qualifikationsarbeiten sowie im Rahmen von Berufungskommissionen und Preiskomitees. Begutachtungen dieser Art erfordern regelmäßig den Einsatz ganzer Arbeitstage. Ebenso gehört es dazu, die eigene wissenschaftliche Arbeit sichtbar zu machen, etwa durch Einträge in entsprechende Datenbanken oder direkte Kommunikation mit der Öffentlichkeit.

Schließlich sollen Professor:innen (e) gute Botschafter:innen ihrer jeweiligen Institution und der Wissenschaft insgesamt sein, indem sie bei z.B. sozialen "Events" wie Alumni-Treffen oder Career Days in Erscheinung treten. Quasi ehrenamtlich engagieren sich viele Professor:innen zudem in erheblichem Umfang, wenn sie Gutachten über besonders förderungswürdige Studierende im Rahmen von Vergabeverfahren für Stipendien verfassen, oder in Leitungsgremien und Kommissionen von Fachgesellschaften mitwirken.

Leidtragende dieser Überfrachtung mit Aufgaben sind zunächst die Professor:innen selbst, von denen viele konstant an der Grenze der eigenen Leistungsfähigkeit arbeiten, oft begleitet von dem Gefühl, Abstriche bei der Qualität machen zu müssen, um allen Anforderungen einigermaßen gerecht werden zu können. Betroffen sind auch ECRs und andere Kolleg:innen der Professor:innen, da letztere für eine ernsthafte wissenschaftliche Zusammenarbeit mit ihnen schlicht keine Zeit und Energie mehr aufbringen können. Die Überfrachtung der Professur trägt zudem auch direkt zur mangelnden Gründlichkeit von Peer-Reviews (P9) und in der Arbeit von Berufungskommissionen (P10) bei. Betroffene sind schließlich Studierende, da konstant überlastete Professor:innen sich nicht auch noch für hochwertige Lehre und Prüfungen engagieren wollen oder können (P4), insbesondere wenn dafür so gut wie keine Anreize bestehen (P5).

Ursachen / beitragende Faktoren

- 1) Die meist anspruchsvollen Aufgaben von Professor:innen sind aufgrund mangelnder finanzieller und personeller Ressourcen auf zu wenige Schultern verteilt.
- 2) Aufgrund einer geringen formalen Flexibilität in der Arbeitsteilung bzw. Aufgabenverteilung (z.B. gleich hohes Lehrdeputat für alle Professor:innen) kommt es zu einer ineffizienten Allokation von Ressourcen.

Empfehlungen

Die Hauptaufgaben von Professor:innen liegen in den Bereichen Forschung und Lehre. Die Kommission empfiehlt, die Professur zu entlasten, um wieder mehr Zeit für diese Aufgaben zu gewinnen.

(6.1.) Verstärkte Einführung spezialisierter Stellen, die Professor:innen von delegierbaren Aufgaben entlasten (Beispiele: Lehrkoordinator:in, Studiengangsentwickler:in, Open Science Koordinator:in; Lab-Manager:in, Data Stewards). Die Schaffung solcher Stellen beugt auch dem Missbrauch von ECRs für die Erfüllung solcher Aufgaben "nebenbei" vor.

(6.2.) Zulassen einer flexibleren Profilbildung hinsichtlich der verschiedenen Aufgaben (Forschung, Lehre, Verwaltung, „Third Mission“) zwischen verschiedenen Professuren und innerhalb einer Professur über die berufliche Laufbahn hinweg. Dies beinhaltet auch eine Flexibilisierung des Lehrdeputats, was Gesetzesänderungen erforderlich machen kann.

(6.3.) Ermöglichung einer Besetzung von Professuren in Teilzeit, z.B. durch Job Sharing. Auch dies kann Gesetzesänderungen erforderlich machen.

(6.4.) Sichtbarmachung der Überlastung der Professur durch öffentliche Benennung als Problem, sowie durch systematische Forschung zum Thema (z.B. Befragung von Professurinhaber:innen).

(6.5.) Erhöhung der Grundfinanzierung der Hochschulen, um eine angemessene Personalausstattung zu ermöglichen (siehe auch **6.1.**).

Mehr überregionale Kooperation zwischen Professor:innen in der Lehre: Zusammenstellung gemeinsamer Lehrmaterialien zu einem Themengebiet, wobei Teilthemen jeweils von denjenigen beigesteuert werden, die dort über die größte Expertise verfügen (siehe **4.5.**).

Erhöhung der konzeptionellen und methodischen Schwellen für die Übergabe eingereicherter Manuskripte ins Peer-Review (siehe **9.2.**). Dies sollte die Zahl zu begutachtender Manuskripte - und damit die Arbeitslast der Professor:innen - senken helfen.

Entlastung von Professor:innen (z.B. in der Lehre oder der Verwaltung) als Kompensation für Engagement in der Kommissionsarbeit (siehe **10.2.**).

Problembereich 7. Befristungspraxis für Stellen unterhalb der Professur (P7)

Das Problem / die Betroffenen

Befristungen von Arbeitsverträgen in der Wissenschaft haben durchaus ihren Sinn, da die meisten Forschungsprojekte aus inhaltlichen Gründen zeitlich begrenzt sind und eine gezielte Einstellung für die Laufzeit eines Projektes möglich sein sollte. Außerdem wird die Promotionszeit als Selektionsphase sowie als Entscheidungsphase im Hinblick auf einen weiteren Verbleib im Wissenschaftssystem angesehen. Die neuesten Zahlen zeigen, dass 81% der wissenschaftlichen Mitarbeiter:innen befristet beschäftigt sind (2013 waren es noch 83%). Das geht aus der Evaluation der Wirkungen des WissZVG hervor, die das HIS-Institut für Hochschulentwicklung (HIS-HE) und die Interval GmbH im Auftrag des Bundesministeriums für Bildung und Forschung (2022) veröffentlicht haben. Betrug 2015 die mittlere Vertragslaufzeit an den Hochschulen noch rund 15 Monate bei Nichtpromovierten bzw. 17 Monate bei Promovierten, so erreichte sie im Jahr 2017 21 bis 22 Monate. Ein wichtiger Faktor bei dieser positiven Entwicklung war die Zunahme von Arbeitsverträgen mit einer dreijährigen Laufzeit. Nach dem Höchstwert im Jahr 2017 haben sich die Laufzeiten in den Jahren 2018 und 2019 auf einem Niveau von rund 20 Monaten eingependelt.

Laut Evaluation hat an den Hochschulen aber immer noch ungefähr ein Drittel der in einem Kalenderjahr abgeschlossenen Arbeitsverträge eine Laufzeit von weniger als zwölf Monaten. In einer Befragung von rund 6.000 Personen auf Qualifikationsstellen gab zudem jede zweite befristet beschäftigte Person an, ihre Vertragslaufzeit für zu kurz bemessen zu halten, um das angestrebte Qualifikationsziel zu erreichen.

Die Kommission sieht vor allem drei Nachteile der aktuellen Situation: (1) Ständige Jobunsicherheit beeinträchtigt wahrscheinlich die Qualität der wissenschaftlichen Arbeit, da die wissenschaftlich arbeitenden Personen sich neben ihren eigentlichen Aufgaben auch permanent mit der eigenen beruflichen Zukunft beschäftigen (und z.B. Bewerbungen schreiben) müssen. (2) In Berichten von Opfern akademischen Machtmissbrauchs (P2) wird die weit verbreitete Abhängigkeit von immer neuen Vertragsverlängerungen regelmäßig als zentrales Instrument genannt, mit dessen Hilfe Vorgesetzte ihre Untergebenen unter Druck setzen. Hierzu bedarf es in der Regel nicht mehr als einer Entscheidung der vorgesetzten Person, einen Vertrag begründungslos nicht zu verlängern. (3) Prekäre und unstete Beschäftigungsverhältnisse auch *nach* der Promotion machen einen langfristigen Verbleib in der deutschsprachigen Wissenschaft (im Vergleich zur Wissenschaft in anderen Ländern ebenso wie im Vergleich zu einer Tätigkeit in der Wirtschaft) vergleichsweise unattraktiv. Insbesondere die nicht gegebene Vereinbarkeit einer solchen Beschäftigungssituation mit der Familienplanung ist in dieser Hinsicht relevant. Sie kann dazu führen (vgl. Kuhnt et al., 2022), dass insbesondere Frauen sich nach der Promotion gegen einen Verbleib in der Wissenschaft entscheiden. Als Ergebnis verlassen viele hochqualifizierte Personen das Berufsfeld nach einiger Zeit und es geht viel intellektuelles Potenzial verloren ("Brain Drain").

Ursachen / beitragende Faktoren

- 1) Der deutlich gestiegene Anteil der Drittmittelfinanzierung an den Hochschulbudgets führt zu verminderter Planbarkeit und zwingt die Hochschulen geradezu, sich durch befristete Einstellungen möglichst viel Flexibilität zu bewahren.
- 2) Das WissZVG wurde ursprünglich einmal mit der Absicht eingeführt, die Hochschulen zu mehr Entfristungen zu bringen (durch Verbot weiterer befristeter Arbeitsverhältnisse nach 12 Jahren Zugehörigkeit zum Wissenschaftsbetrieb). Es hatte jedoch ziemlich genau den gegenteiligen Effekt: Nun werden jüngere Forschende regelmäßig aus dem Wissenschaftsbetrieb entfernt, wenn die 12-Jahres-Grenze naht.
- 3) In der aktuellen Situation, in der Professuren regelmäßig mit Personalverantwortung ausgestattet sind, versuchen Hochschulleitungen durch Fortsetzung der derzeitigen Befristungspraxis für Bewerber:innen auf Professuren attraktiv zu bleiben, indem Stellen durch diese "frei besetzbar" gehalten werden.
- 4) Gelegentlich wird auch ins Feld geführt, dass Entfristungen zu einer Reduktion der wissenschaftlichen Produktivität führen könnten - ohne, dass dies allerdings bisher empirisch belegt worden wäre. Dasselbe Argument wäre konsequenterweise auch gegen dauerhafte Einstellungen von Professor:innen zu verwenden.

Empfehlungen

- (7.1.)** Tenure-Track-Stellen als Regelfall der postdoktoralen wissenschaftlichen Weiterqualifikation. Einführung dieser Stelle in einem "rollierenden Verfahren" (vgl. Forderung der DGPs, 2022). Eine positive Evaluation sollte zu einer Entfristung führen, entweder mit oder ohne Professor:innentitel.
- (7.2.)** Erhöhung der Zahl der Dauerstellen mit vollem oder sogar erhöhtem Lehrdeputat. Voraussetzung für Bewerbungen auf solche Stellen sollte die abgeschlossene Promotion sein. Auch weitere Daueraufgaben sollten diesen Stellen fest zugeteilt werden können, was zu einer Entlastung der Professur (siehe P6) beitragen dürfte.
- (7.3.)** Vergabe von Promotionsstellen für in der Regel 4 Jahre.
- (7.4.)** Die reine *Datenerhebung* im Rahmen von Forschungsprojekten kann und sollte primär von Wissenschaftlichen Hilfskräften geleistet werden. Dadurch würden Promovierende von Aufgaben entlastet, für die sie in den allermeisten Fällen überqualifiziert sind, und damit Zeit gewinnen, sich auf ihre eigentlichen Aufgaben (Rezeption der Literatur, Konzipierung von Studien, Datenanalysen, Manuskripterstellung) zu konzentrieren. Auch wären die Hilfskräfte, die Datenerhebungen im Rahmen von Studien durchführen, durch die Projektleitung weniger leicht unter Druck zu setzen, da das Druckmittel des Promotionsvorhabens entfielen. Schließlich könnten auf diese Weise Kosten gespart werden.

Problembereich 8. Unnötig starkes Machtgefälle (P8)

Das Problem / die Betroffenen

Eine Tätigkeit in der Wissenschaft bietet je nach Position unterschiedlich stark die Möglichkeit, Macht über andere auszuüben. Dabei ist das Machtgefälle zwischen Personen, die eine Professur innehaben, und den meisten anderen Personen im System (v.a. ECRs, Studierende) sehr groß (Scherpenberg et al., 2021). In der Promotions- und der Postdoc-Phase bündeln sich in den Aufgaben und Verantwortlichkeiten der Professur die Entscheidungsmacht über Ressourcen (z.B. Laborzeiten, finanzielle Mittel für Versuchsdurchführung oder Kongressbesuche), die Bewertungsmacht (Benotung der Dissertation, Arbeitszeugnisse) und die Personalmacht (Arbeitsverträge: Vertragsdauer und Verlängerung). Professor:innen entscheiden zudem als größte Gruppe in Berufungskommissionen (P10) darüber, wer eine Professur erhält, und als Gutachter:innen darüber, wessen Arbeiten wie publiziert werden (P9) und wessen Drittmittelprojekte förderungswürdig sind.

Betroffen davon sind vor allem die im Wissenschaftsbetrieb abhängig Beschäftigten und wissenschaftlich Betreuten (v.a. ECRs, sowie Verwaltungsangestellte), die dadurch in Gefahr geraten, Opfer von Machtmissbrauch (P2) zu werden. Die Bündelung dieser verschiedenen Machtquellen in der Professur kann missbräuchlich genutzt werden, um einen Einsatz fragwürdiger wissenschaftlicher Praktiken (QRPs) einzufordern (P1). Auch leidet die Qualität des wissenschaftlichen Diskurses, wenn - aufgrund des Bestehens einer "Angstkultur" - zwischen Professor:innen und ECRs nicht gleichrangig diskutiert werden kann (Leising, et al., 2021; Kuhnt et al., 2022).

Ursachen / beitragende Faktoren

Das große Machtgefälle zwischen Professor:innen auf der einen Seite und Studierenden und ECRs auf der anderen Seite ist vor allem die Folge einer Häufung von Funktionen bei Ersteren: Hier bündeln sich insbesondere für ECRs die folgenden vier Funktionen (vgl. P6): (a) Kooperationspartner:in der wissenschaftlichen Arbeit, (b) Vorgesetzte:r mit Personal- und Budgetverantwortung, (c) Gutachter:in für das Ergebnis der wissenschaftlichen Zusammenarbeit, (d) Unterstützer:in in fachlichen und nicht-fachlichen Fragen (z.B. Karriereberatung).

Besonders problematisch erscheint im Hinblick auf ECRs die gängige Ausübung der Funktion als Personalverantwortliche:r durch Professor:innen, in Verbindung mit der derzeit an den Hochschulen gängigen Befristungspraxis auf Basis des WissZVG (Kuhnt et al., 2022; und P7). Hier besteht die konkrete Gefahr, dass die Frage der Vertragsverlängerung als Druckmittel genutzt wird, und es mangelt nicht an entsprechenden Berichten, dass genau dies auch häufig passiert.

Ein weiterer möglicher Ansatzpunkt für Machtmissbrauch ergibt sich aus der vermeintlichen oder tatsächlichen Verfügungsgewalt von Professor:innen über die Daten, die im Rahmen von Forschungsprojekten anfallen. Der Zugang zu diesen Daten ist für viele ECRs eine

unabdingbare Voraussetzung dafür, ihre wissenschaftlichen Fragestellungen überhaupt bearbeiten zu können. Die Rechtslage in diesem Bereich („wem gehören die Daten?“) ist den meisten Beteiligten aber bis heute offenbar weitgehend unklar.

Gegenüber *Studierenden* üben Professor:innen nicht nur die Funktion als Lehrpersonen mit hohem Ermessensspielraum in der Gestaltung der Lehre aus, sondern auch die Funktion als Prüfende. Sie lenken zudem im Rahmen der akademischen Selbstverwaltung die Geschicke ihrer jeweiligen Fakultäten maßgeblich, während andere Personengruppen (Studierende, ECRs, Dauerstelleninhaber:innen, Verwaltungspersonal) dort bislang nur relativ geringen Einfluss ausüben können.

Empfehlungen

(8.1.) Abgabe der (bisher alleinigen) Personalverantwortung der Professur für einzelne Mitarbeiter:innen an die Fakultät. Das bedeutet: Professuren *ohne* zugewiesene Haushaltsstellen sind der Regelfall. Um dabei eine unerwünschte Machtkonzentration beim jeweiligen Dekanat zu verhindern, sollten Personalentscheidungen möglichst gemeinsam von einem Gremium mit regelmäßig wechselnder Zusammensetzung getroffen werden.

(8.2.) Öffentliche Sichtbarmachung vorbildlicher Lösungen in dieser Hinsicht (z.B. durch Bericht auf der DGPs-Webseite), als Orientierung für andere Fakultäten.

(8.3.) Entscheidungen über die Allokation finanzieller und materieller Forschungsressourcen (z.B. Laborkapazitäten) sollten fakultätsintern möglichst durch demokratisch legitimierte Gremien unter Beteiligung von ECRs getroffen werden.

(8.4.) Abgabe (eines Teils) der Coachingfunktion der Professur an besonders dafür qualifizierte Personen, etwa im Rahmen von Graduiertenschulen / Career Centers (dies betrifft auch das Budget für Qualifizierungsmaßnahmen).

(8.5.) Verpflichtende *externe* Bewertung wissenschaftlicher Arbeiten als Regelfall: Personen, die an der Erstellung einer wissenschaftlichen Arbeit selbst beteiligt waren (und deren Kooperationspartner:innen), dürfen an der Bewertung dieser Arbeit nicht beteiligt sein.

(8.6.) *Ein* bei einem wissenschaftlichen Journal zur Begutachtung eingereichtes Manuskript sollte als Promotionsgrundlage ausreichen. Dies berücksichtigt die deutlich gestiegenen Anforderungen in Zeiten von Open Science sowie das Ziel, der Qualität der Forschung in Zukunft mehr Gewicht gegenüber der Quantität zu geben.

(8.7.) *Externe* Begutachtung dieses Manuskripts (8.6.), unabhängig vom Peer-Review-Verfahren des Journals - andernfalls könnten ECRs Leidtragende langer Reviewverfahren werden, für die sie selbst nicht verantwortlich sind. Auswahl der externen (gerne auch internationalen) Gutachter:innen durch eine unabhängige Stelle. Dieser Ansatz bedeutet auch, dass der Abschluss der Promotion von einem eventuellen Publikationserfolg explizit entkoppelt wird.

(8.8.) Abschaffung der Promotionsnoten.

(8.9.) Nachdenken über eine paritätische Besetzung von Selbstverwaltungsgremien mit Angehörigen der verschiedenen Statusgruppen (ECRs, Studierende, Verwaltung, Professor:innen). Hier muss die Rechtslage berücksichtigt und evtl. problematisiert werden.

(8.10.) Bei Einstellung: Feste Zusage eines individuellen Budgets für Kongressreisen, über das ECRs eigenständig - das heißt unabhängig von Professor:innen - verfügen können.

(8.11.) Einrichtung unabhängiger juristischer Beratung für ECRs und Finanzierung derselben über die Institution (z.B. Graduiertenschulen).

(8.12.) Erarbeitung und Veröffentlichung einer juristisch abgesicherten Leitlinie zum Thema "Rechte an Forschungsdaten" durch die DGPs (und/oder die DFG).

Verstärkte Nutzung von Betreuungsvereinbarungen (siehe **3.4.**).

Explizite, rechtsverbindliche Regelung des Verfahrens zum Betreuungswechsel in der Promotionsordnung (siehe **3.6.**).

Geänderter Umgang mit Befristungen (siehe **7.3.**).

Problembereich 9: Versagen des Peer-Review-Systems (P9)

Das Problem / die Betroffenen

Peer-Review (also die Begutachtung von Forschungsideen und Forschungsergebnissen durch Kolleg:innen) spielt eine zentrale Rolle im Wissenschaftsbetrieb. Es ist notwendig, da eine kompetente Beurteilung wissenschaftlicher Leistungen in der Regel hohe Expertise voraussetzt, über die nur andere im selben Feld arbeitende Forscher:innen in ausreichendem Maß verfügen. Die Kommission geht davon aus, dass Peer-Review auch in Zukunft die entscheidende Rolle bei solchen Beurteilungen spielen muss und wird. Peer-Review wird seiner Qualitätssicherungsfunktion bisher jedoch nicht ausreichend gerecht (Rennie, 2016; Tennent & Ross-Hellhauer, 2022; Vazire & Holcombe, 2021). Dies betrifft sowohl die Begutachtung von Manuskripten durch wissenschaftliche Zeitschriften als auch die Begutachtung von Förderanträgen durch Drittmittelgeber. Immer wieder werden unter anderem die folgenden Missstände beklagt:

- (1) Peer-Review ist kaum in der Lage, betrügerisches Handeln (P1) zu entdecken.
- (2) Anfälligkeit des Systems für Nepotismus und Reziprozitätseffekte ("eine Hand wäscht die andere"). Rollen von Autor:innen und Reviewer:innen wechseln kontinuierlich in beide Richtungen. Dabei ist Anonymität oft nicht zu gewährleisten oder wird sogar aktiv durchbrochen, ohne dass dies transparent oder geregelt wäre.
- (3) Starke Rolle politischer und strategischer Überlegungen (z.B. Beförderung oder Behinderung der Interessen einer bestimmten "Schule" des Denkens oder eines bestimmten Forschungsgegenstandes), die oft hinter nur scheinbar wissenschaftlichen Argumenten versteckt werden.
- (4) Ausnutzen der eigenen Machtposition als Reviewer:in, um Zitationen der eigenen Arbeiten zu erzwingen oder zumindest wahrscheinlicher zu machen, oder um Publikationen anderer zu verhindern/verzögern/verschlechtern, die dem Erfolg der eigenen Arbeiten der Reviewer:innen (vermeintlich) im Weg stehen könnten.
- (5) Nachlässigkeit in der Begutachtung, sowie manchmal unangemessen feindselige und verletzende Ausdrucksweise seitens der Reviewer.
- (6) Das System ist oft ineffizient, weil Artikel, die von einer Zeitschrift abgelehnt wurden, bis zum Erfolg bei weiteren Zeitschriften eingereicht und dort wieder begutachtet werden.

Leidtragende dysfunktionaler Reviews sind vor allem solche Forschende, die Wert auf eine hohe Qualität der eigenen Arbeit legen, da sie sich nicht sicher sein können, dass diese Qualität ausreichend gewürdigt werden wird. Sie müssen befürchten, dass andere (persönliche, politische) Gesichtspunkte im Peer-Review-Verfahren die entscheidende Rolle spielen werden und evtl. andere, schwächere Arbeiten aus diesen Gründen den Vorzug erhalten. Dies kann sich unmittelbar in unfairen Nachteilen bei Berufungsverfahren übersetzen. Betroffene sind auch die Fleißigen unter den Reviewer:innen, da sich ihre hochqualifizierte und anstrengende Arbeit für sie selbst bisher ebenfalls nicht genug auszahlt.

Ursachen / beitragende Faktoren

- 1) Die Anzahl der zu begutachtenden Artikel ist konstant sehr hoch, was zu deutlichen Ermüdungserscheinungen auf Seiten der (potenziellen) Reviewer führt ("Reviewer Fatigue"). Ein Hauptgrund für diese Artikelschwemme wiederum liegt in Fehlanreizen (P5): Forschende werden vor allem dafür belohnt, schnell möglichst viel zu publizieren und möglichst hohe Drittmittelsummen einzuwerben, was die Zahl der eingereichten Manuskripte und Anträge in die Höhe treiben dürfte.
- 2) Die dargestellte Überlastung der Gutachtenden mit Arbeitsaufgaben (P6) führt dazu, dass bei der Gründlichkeit von Reviews Abstriche gemacht werden (müssen).
- 3) Es bestehen für Gutachtende kaum extrinsische Anreize dafür, sich beim Reviewen Mühe zu geben.
- 4) Wissenschaftsverlage haben ein Interesse daran, dass publiziert wird. Besonders gründliche Reviews verlangsamten den Publikationsprozess oder reduzieren die Anzahl publizierbarer Manuskripte, was den kommerziellen Interessen der Verlage entgegensteht.
- 5) Die Möglichkeiten von Reviewern, ein vollständiges und ungeschöntes Bild von der Forschungsarbeit der Autor:innen eines Manuskripts zu bekommen, sind begrenzt, da in der Regel kein direkter (ungefilterter) Zugang zu Informationen (z.B. Rohdaten) besteht.
- 6) Nicht alle Gutachter:innen verfügen über alle Kompetenzen, die sie bräuchten (z.B. ausreichende Kenntnisse in den vielfältigen Programmiersprachen oder komplexen Auswertungsprozeduren), um eine Forschungsarbeit angemessen beurteilen zu können.
- 7) Traditionell findet Peer-Review in anonymisierter Form statt. Dies soll es den reviewenden Personen ermöglichen, auch sehr kritische Einschätzungen abzugeben, ohne z.B. Vergeltungsversuche seitens der Autor:innen befürchten zu müssen. Andererseits bietet eben diese Anonymität den Reviewenden auch Raum für nachlässiges oder politisch motiviertes Handeln, ohne sich dafür rechtfertigen zu müssen.

Empfehlungen

- (9.1.)** Klare öffentliche Positionierung von Fachgesellschaften und Forschungsinstitutionen zugunsten einer deutlichen Reduktion des rein quantitativen Publikations-Outputs.
- (9.2.)** Reduktion der Menge zu begutachtender Manuskripte durch eine deutliche Erhöhung der methodischen Mindestanforderungen (z.B. Prä-Registrierung und mandatorische Replikationsversuche bei empirischen Arbeiten) und "Desk Rejection" bei Nichterfüllung derselben. Hierbei sind Besonderheiten zu berücksichtigen, die sich aus besonders anspruchsvollen Designs oder schwer rekrutierbaren Stichproben ergeben.
- (9.3.)** Schnellstmöglicher Übergang zu Not-for-Profit-Publishing als Regelfall, da es in diesem Modell keine kommerziellen Interessen gibt, die mit wissenschaftlichen Zielen in Konflikt geraten können. Not-for-Profit-Journals können an zentraler Stelle (z.B. *Personality*

Science als PsychOpen GOLD Journal des ZPID) gehostet werden oder an einzelnen Hochschulen (z.B. *Meta-Psychology* an der Linnaeus University in Schweden).

(9.4.) Die Kommission empfiehlt, den (o.g.) Übergang zu Not-for-Profit-Publishing öffentlich zu einem zentralen Entwicklungsziel zu erklären.

(9.5.) Die Inhalte von Peer-Reviews sollten im Regelfall öffentlich gemacht werden. Dadurch würde vor allem die Transparenz des Prozesses im Hinblick auf die Stichhaltigkeit der vorgebrachten Argumente gesteigert.

(9.6.) Einsatz spezialisierter Expert:innen, die - zusätzlich zum inhaltlichen Review - alle eingereichten Arbeiten im Hinblick auf dieselben technischen Aspekte (z.B. Einhaltung von Prä-Registrierungen; Reproduzierbarkeit von Ergebnissen anhand offener Daten und R-Skripte) bewerten.

(9.7.) Peer-Reviews sollten, wenn möglich, von den reviewenden Personen namentlich unterzeichnet werden. So würde die Transparenz im Hinblick auf mögliche Seilschaften und Voreingenommenheiten erhöht. Auch sollte dies dazu beitragen, dass Reviewer:innen sich Mühe geben, einen konstruktiven Ton anzuschlagen. Allerdings kann es natürlich immer Situationen geben, in denen ein anonymes Peer-Review weiterhin angebracht erscheint.

(9.8.) Eine solche Entblindung birgt das Risiko der Beförderung von (positiven wie negativen) Reziprozitäten: Autor:innen könnten sich für Beurteilungen durch ihnen bekannte Personen bei späteren Begutachtungen von deren Arbeiten revanchieren wollen (Lynam, Hyatt, Hopwood, Wright & Miller, 2019). Ein weiteres Risiko besteht in einer generell reduzierten Bereitschaft von Reviewer:innen, angebrachte Kritik auch öffentlich zu äußern. Beide Risiken gilt es daher im Blick zu behalten.

(9.9.) Peer-Reviews sollten als eigenständiges Publikationsformat anerkannt, mit DOIs versehen und damit zitierbar werden. Dies könnte helfen, die Motivation der Reviewer:innen zu gründlichem Vorgehen zu steigern. Allerdings läge es damit im Ermessen von Herausgeber:innen, wer derartige Publikationen beisteuern darf. Um Missbräuchen dieser Macht vorzubeugen, sollte (a) transparent gemacht werden, wer wen als Reviewer:in für ein Manuskript eingeladen hat, (b) eine solche Entscheidung nicht von einer einzelnen Person getroffen werden, und (c) die Anzahl solcher Einladungen an dieselbe Person von vornherein (z.B. auf 3 pro Jahr und Zeitschrift) limitiert sein. Auch für anonyme Reviews ist eine DOI-Vergabe vorstellbar.

(9.10.) Gezielte Förderung von *Registered Reports* als Ideal, um einerseits den Publikationsbias zu bekämpfen und andererseits die Expertise von Gutachter:innen bereits in der Phase der Studienplanung nutzbar zu machen. Registered Reports könnten beispielsweise mithilfe eines Fast-Trackings durch Journals incentiviert werden.

Problembereich 10. Mangelnde Qualität von Berufungsverfahren (P10)

Das Problem / die Betroffenen

Schon seit Jahren wird ein dringender Verbesserungsbedarf im Bereich der Personalauswahlverfahren für wissenschaftliches Führungspersonal artikuliert (Abele-Brehm & Bühner, 2016a; 2016b). Beklagt wird u.a. mangelnde Sorgfalt, die Verwendung von Daten (z.B. Metriken) fragwürdiger Validität, die Nichterfassung hoch relevanter überfachlicher Merkmale (z.B. Integrität), und ein übermäßig starker Einfluss von lokaler Mikropolitik. Aus all dem können Fehlentscheidungen zu Lasten mehrerer Akteur:innen resultieren:

- 1) Studierende sind potenziell von schlechter Lehre betroffen (P4).
- 2) Wissenschaftliche Mitarbeiter:innen sind über lange Zeiträume vom (fachlichen und zwischenmenschlichen) Fehlverhalten problematischer Führungspersonen betroffen (P2). Fragwürdige wissenschaftliche und zwischenmenschliche Wertvorstellungen werden so an die jüngere Generation weitergegeben und in ihrer problematischen Wirkung vervielfacht.
- 3) Die Berufung wissenschaftlichen Führungspersonals mit fraglicher persönlicher Integrität kann auch zu einer mangelhaften Integrität und/oder Relevanz des wissenschaftlichen Outputs führen, was zu Lasten anderer Forschender und der Allgemeinheit geht (P1).
- 4) Die Besetzung akademischer Führungspositionen mit dafür ungeeigneten Personen schafft ein erhebliches Reputationsrisiko für betroffene Institutionen ("Skandale").

Ursachen / beitragende Faktoren

(1) Mangelhafte Sorgfalt

Eine valide Diagnostik der Eignung von Bewerber:innen erfordert (a) allgemeine diagnostische Kompetenz (z.B. Wissen um Konzepte wie Reliabilität, prädiktive und Konstruktvalidität sowie typische Urteilsfehler), (b) spezifische fachliche Kompetenz, die Qualität wissenschaftlicher Beiträge (Publikationen) richtig einschätzen zu können (anstatt diese z.B. nur zu zählen), und (c) die Möglichkeit, diese Kompetenzen gewissenhaft und gründlich einzusetzen.

Eine angemessen gründliche Beschäftigung mit den wissenschaftlichen Leistungen von Bewerber:innen kostet viel Zeit. Dazu ein Rechenbeispiel: Veranschlagt man für das *sorgfältige* Studium der Bewerbungsunterlagen und die *sorgfältige* Lektüre von (nur) drei Publikationen einen ganzen Arbeitstag, ergäben sich bei zehn Bewerbungen (nach einer weniger aufwändigen Vorauswahl) ein Gesamtaufwand von zehn ganzen Arbeitstagen. Dieser Aufwand ist aufgrund der weit verbreiteten Überlastung mit Arbeitsaufgaben (P6) und des Fehlens von Anreizen für gründliche Kommissionsarbeit (P5) kaum zu leisten. Zusätzlich ergibt sich das Problem, dass durch die zum Teil bestehende Pflicht zur geschlechterparitätischen Besetzung von Kommissionen auf Frauen eine Mehrbelastung zukommt, wenn

diese in der Gruppe des wissenschaftlichen Personals, wie oft noch der Fall, unterrepräsentiert sind. Eine rationale Kosten-Nutzen-Abwägung legt daher individuell eine Entscheidung im Sinne der Aufwandsminimierung nahe, was Einbußen bei der diagnostischen Qualität der Beurteilungen wahrscheinlich macht.

(2) Verwendung von Daten und Kriterien fragwürdiger Validität

Als Alternative zur gründlichen inhaltlichen Befassung mit den Forschungsleistungen einzelner Bewerber:innen wird häufig auf die Verwendung quantitativer Metriken zurückgegriffen (z.B. Anzahl von Publikationen und Zitationen, Drittmittelsummen, h-Indizes, Impact Faktoren), die einfach zu ermitteln sind und zudem vergleichsweise objektiv anmuten (siehe P5; Abele-Brehm & Bühner, 2016a). Dieses Vorgehen führt zu einer drastischen Arbeits- und Zeitersparnis und ist daher entsprechend attraktiv. Die zuvor genannten Metriken werden in ihrer Aussagekraft inzwischen jedoch zunehmend skeptisch gesehen, was u.a. an ihrer Empfänglichkeit für Manipulationen liegt (siehe P5), sowie am weitgehenden Fehlen ernstzunehmender Qualitätskontrollen im Wissenschaftssystem (siehe P9, P12) (Chapman et al., 2019; Dougherty & Horne, 2022; Fong & Wilhite, 2017; Leising et al., 2022a; Schönbrodt et al., 2022). Im schlimmsten Fall hat eine Orientierung an diesen Kennwerten eine Auswahl derjenigen Personen zur Folge, die den gegebenen Anreizen am konsequentesten folgen und die individuelle Metrikoptimierung über die inhaltliche Qualität ihrer Arbeit stellen.

Im Hinblick auf die Erfassung der Lehrkompetenz ist zu hinterfragen, inwiefern ein etwa viertelstündiger Lehrvortrag geeignet ist, die relevanten Charakteristika (z.B. Fähigkeit zur nachhaltigen Vermittlung anspruchsvoller Inhalte) abzubilden. Gleiches gilt bei der Heranziehung von studentischen Lehrevaluationen, da diese durch Einflüsse wie persönliche Sympathie für die Lehrperson und Schwierigkeit der Inhalte beeinflusst sein können (Kreitzer & Sweet-Cushman, 2021).

(3) Nichterfassung hoch relevanter überfachlicher Merkmale (z.B. Integrität)

Übereinstimmend mit den Ergebnissen psychologischer Entscheidungsforschung geht die Kommission davon aus, dass Machtmissbrauch (P2) und wissenschaftliches Fehlverhalten (P1) aus Interaktionen zwischen Merkmalen von Personen und Merkmalen ihrer jeweiligen Handlungskontexte resultieren. Eine valide Diagnostik relevanter Persönlichkeitsmerkmale und überfachlicher Kompetenzen wie etwa professionelle Führung und Konfliktmanagement (Peus et al., 2015) ist deshalb im Rahmen von Berufungsverfahren unabdingbar. Es besteht jedoch ein grobes Missverhältnis zwischen der hohen Relevanz solcher Merkmale und der marginalen Rolle, die sie im Rahmen der meisten Berufungsverfahren bislang spielen.

Fehlverhalten in der Wissenschaft kann als eine Form von Counterproductive Work Behavior (CWB) aufgefasst werden und tritt u.a. als sogenannte "Abusive Supervision" in Erscheinung (Fischer et al., 2021). Einzelstudien und Metaanalysen zufolge zählen zu den Persönlichkeitsprädiktoren des CWB: (geringe) Selbstkontrolle (Marcus & Schuler, 2004; Pusch & Holtfreter, 2020), (schwache) moralische Orientierung (Pusch & Holtfreter, 2020), Sensation-Seeking (O'Connor et al., 2017), Narzissmus (insbesondere die Facetten Entitlement und

Exploitativeness; Grijalva & Newman, 2014), Machiavellismus (O’Boyle et al., 2012), (geringe) Gewissenhaftigkeit, (geringe) Verträglichkeit und Neurotizismus als Domänen aus dem Fünf-Faktoren-Modell der Persönlichkeit (Grijalva & Newman, 2014; Pusch & Holtfreter, 2020; Runge et al., 2020), (geringe) Honesty-Humility, (geringe) Extraversion, (geringe) Verträglichkeit und (geringe) Gewissenhaftigkeit aus dem HEXACO-Modell der Persönlichkeit (Anglim et al., 2018; Pletzer et al., 2019; Pusch & Holtfreter, 2020; Zettler et al., 2020), sowie die so genannte dunkle Persönlichkeit (D) als Kern der so genannten dunklen Triade (Narzissmus, Machiavellismus, Psychopathie) (Moshagen et al., 2018; Palmer et al., 2020; Pusch & Holtfreter, 2020; Schwarzinger, 2020). Einige dieser Persönlichkeitseigenschaften von Führungskräften bewirken auch erhöhtes CWB ihrer Mitarbeiter:innen (Palmer et al., 2020). Die umfangreiche Literatur zu dieser Frage kann zwar einerseits als weiterer Beleg für die unter P1 beklagte Unübersichtlichkeit und wahrscheinliche Redundanz vieler psychologischer Konstrukte gelesen werden, zeigt aber auch, dass die Bedeutsamkeit von Persönlichkeitsfaktoren in diesem Bereich allgemein anerkannt wird.

(4) Übermäßig starker Einfluss lokaler Mikropolitik

Systematische Untersuchungen zum Einfluss lokaler Mikropolitik auf Berufungsverfahren sind der Kommission nicht bekannt. Eine große Menge anekdotischer Evidenz legt jedoch nahe, dass persönliche Interessen von Kommissionsmitgliedern nicht selten sogar eine *entscheidende* Rolle bei der Auswahl von Bewerber:innen spielen (z.B. Austragung persönlicher Feuden zwischen Kommissionsmitgliedern im Wege der Personalauswahl). Einige der im Folgenden gegebenen Empfehlungen verfolgen daher auch das Ziel, diese unerwünschten Einflüsse zu begrenzen.

Empfehlungen

Wie alle diagnostischen Entscheidungen beinhalten auch solche über die Besetzung akademischer Führungspositionen zwei Risiken: ungeeignete Bewerber:innen einzustellen (“falsch positive Entscheidungen”) und geeignete Bewerber:innen zu übersehen (“falsch negative Entscheidungen”). Ziel der hier vorgestellten Empfehlungen ist es, beide Fehlerarten zu minimieren.

(10.1.) Mitglieder von Berufungskommissionen brauchen diagnostische Expertise. Konzepte wie Reliabilität und Validität, typische Urteilsfehler und deren Ursachen, sowie Wege diese zu vermeiden, sind den meisten ausgebildeten Psycholog:innen präsent. Da Kommissionen jedoch in der Regel mit Personen ganz unterschiedlicher Expertise besetzt sind, und zudem die meisten Berufungsverfahren außerhalb der Psychologie stattfinden, scheint es ratsam, Kommissionsmitglieder gezielt für ihre Aufgabe zu schulen. Als Vorbild hierfür könnten die Schulungen nach DIN 33430 dienen (Diagnostik- und Testkuratorium, 2018).

(10.2.) Mitglieder von Berufungskommissionen brauchen Zeit und Energie für eine gründliche Auseinandersetzung mit Bewerbungen. Deshalb muss der von ihnen geleistete Einsatz ausgeglichen werden, z.B. durch Entlastungen in der Lehre oder durch die Bewilligung von

zusätzlichen Personal- und Sachmitteln. Dies betrifft explizit nicht nur Professor:innen, sondern auch Mittelbauvertreter:innen.

(10.3.) Verwendung von Kriterien Guter Wissenschaftlicher Praxis bei der Evaluation des wissenschaftlichen Outputs und des Potenzials von Bewerber:innen im Rahmen von Berufungsverfahren (Gärtner, Leising & Schönbrodt, 2022; Leising et al., 2022a, 2022b; Schönbrodt et al., 2022).

(10.4.) Entwicklung und Verwendung alternativer Indikatoren für gute Lehre. Lehrvorträge im Rahmen von Berufungsverfahren könnten etwa bzgl. inhaltlicher Richtigkeit, Anforderungen und Didaktik von Expert:innen beurteilt werden. Anschließende Verständnistests würden eine objektivere Erfassung des erzielten Lernfortschrittes ermöglichen als dies z.B. durch globale Ratings der Zufriedenheit seitens Studierender möglich ist. Eine weitere Möglichkeit besteht in der Aufzeichnung realer Unterrichtseinheiten auf Video und der Einsendung solcher Aufzeichnungen als Teil von Bewerbungen auf Professuren.

(10.5.) Relevante Persönlichkeitsmerkmale und überfachliche Kompetenzen (z.B. Führung, Mentoring und Konfliktmanagement) müssen explizit in ihrer Bedeutung anerkannt und in die Eignungsbeurteilung aufgenommen werden (Abele-Brehm & Bühner, 2016a). Dies erfordert die systematische und möglichst valide Messung derselben. Selbsteinschätzungen sind u.a. aufgrund der sozialen Erwünschtheit der Merkmale hierfür nur eingeschränkt geeignet. Allerdings gibt es vielversprechende Inventare zur Messung der "dunklen Triade" im Arbeitskontext (Schwarzinger & Schuler, 2016), die offenbar vergleichsweise wenig anfällig für Verfälschungstendenzen sind. Die Kommission empfiehlt die zusätzliche Verwendung weiterer Verfahren zur Messung dieser Merkmale, insbesondere:

- Assessment Center, wie sie z.B. an der Universität Bremen seit mehreren Jahren durchgeführt werden (Müller, 2015)
- Strukturierte Eignungsinterviews mit biographischen Elementen und kritischen Szenarien aus dem Wissenschafts- und Hochschulbetrieb (Strobel et al., 2018)
- Situational Judgment Tests (Corstjens et al., 2017; Kasten et al., 2020)
- Integritätstests (Marcus et al., 2007; Ones & Viswesvaran, 2001; Ones et al., 1993, 2012; Van Iddekinge et al., 2012)
- Selbstbeschreibungen im Forced-Choice-Format (Cao & Drasgow, 2019)
- Einschätzungen durch Kolleg:innen und Studierende, die Bewerber:innen aus früheren Interaktionskontexten kennen.

(10.6.) Perspektivisch empfiehlt die Kommission die Implementierung des in der Wirtschaft bewährten 360°- Feedbacks als Maßnahme vor bedeutenden Karriereschritten (z.B. als Bestandteil der Tenure-Track-Evaluation).

(10.7.) Deutliche Erhöhung des Anteils externer Mitglieder in Berufungskommissionen und Auswahl derselben durch möglichst unabhängige Instanzen wie z.B. Gremien von Fachgesellschaften. Auf diesem Weg kann der Einfluss der lokalen Mikropolitik wahrscheinlich zurückgedrängt und evtl. auch mehr diagnostische Expertise in die Kommissionsarbeit eingebracht werden.

(10.8.) Die Kommission empfiehlt, den wissenschaftlichen Forschungsstand zur Erfassung der persönlichen Integrität im Bewerbungskontext (siehe 10.5.) einmal gründlich zu recherchieren und die Ergebnisse dieser Recherche der Community als Handreichung für Bewerbungsverfahren zur Verfügung zu stellen. Dabei ergeben sich mit großer Wahrscheinlichkeit auch Hinweise auf weitere Forschungsbedarfe.

Problembereich 11. Mangelnde Klarheit und Verbindlichkeit ethischer Normen (P11)

Das Problem / die Betroffenen

Das Vorhandensein ethischer Normen ist eine wesentliche Voraussetzung dafür, das Handeln von Menschen als mehr oder weniger moralisch akzeptabel beurteilen zu können. Es ist daher auch zentral für *alle* hier adressierten Problemfelder. Die Etablierung einer ethisch ausgerichteten Organisationskultur erfordert aber nicht nur die *Existenz* von Normen, sondern auch, dass diese allen Handelnden *bekannt* sind und von diesen *als verbindlich gültig ernstgenommen* werden. Empirische Befunde zu wissenschaftlichem Fehlverhalten und Machtmissbrauch (z.B. Chapman et al., 2019; Elson et al., 2021; John et al., 2012) legen nahe, dass in der Wissenschaft erheblicher Nachholbedarf in dieser Hinsicht besteht.

Dabei unterscheiden sich die Problematiken zwischen den einzelnen Handlungsfeldern teilweise deutlich: Den meisten Wissenschaftler:innen sollte der Unterschied zwischen lauterem und fragwürdigen wissenschaftlichen Praktiken (QRPs) bekannt sein. Dennoch bestehen offenbar erhebliche Diskrepanzen zwischen dem, was als theoretisch wünschenswert akzeptiert wird, und dem, was im Forschungsalltag tatsächlich geschieht (Anderson, 2007; Elson et al., 2021; John et al., 2012).

Ethisch verwerfliches Forschungshandeln mag der eigenen Karriere und den Ranking-Ergebnissen akademischer Institutionen nutzen, schadet zugleich aber anderen Akteur:innen ganz erheblich (P1). Ein anhaltender Widerspruch zwischen nach außen dargestellten hohen ethischen Ansprüchen und ständigen Verstößen gegen dieselben im Inneren der Organisation kann zudem zu Zynismus und Entfremdung auf Seiten der Angestellten führen (z. B. Mete, 2013) und damit langfristig auch ganz konkret der jeweiligen Organisation selbst schaden.

Im Bereich des Machtmissbrauchs (P2) (Ausbeutung, Bullying) scheint sogar Unsicherheit über die Normen selbst zu herrschen. Gespräche der Kommission mit verschiedenen Akteur:innen im System ergaben, dass vielen unklar zu sein scheint, wer im akademischen System eigentlich welche Macht zu welchem Zweck verliehen bekommt, und wo die Grenze zwischen legitimem Machtgebrauch und illegitimem Machtmissbrauch verläuft. Geringe Kenntnisse scheinen insbesondere im Hinblick auf das Thema der *destruktiven Führung* zu bestehen, welche zum Beispiel von Führungskräften mit ausgeprägt narzisstischen bzw. psychopathischen Eigenschaften ausgehen kann (P10). Ebenso wissen viele Akteur:innen wenig über arbeitsrechtliche Belange. Mangelnde Klarheit darüber, was akzeptabel ist und was nicht, kann jedoch Fehlverhalten befördern. Hier erkennt die Kommission einen ausgeprägten Bedarf an Maßnahmen zur Schaffung von Problembewusstsein ("Awareness") bei allen Akteur:innen.

Ursachen / beitragende Faktoren

1) Einer konsequenten Umsetzung von GWP-Richtlinien in der Praxis steht einerseits das derzeit geltende wissenschaftliche Anreizsystem entgegen, das die Anwendung *niedriger* wissenschaftlicher Qualitätsmaßstäbe im Dienste einer Maximierung des rein quantitativen Outputs incentiviert (P5) sowie andererseits das Fehlen zuverlässiger Qualitätskontrollen (P9, P12). Die offiziell deklarierten Werte in diesem Bereich widersprechen oft diametral dem, was auf der Ebene realen Verhaltens belohnt bzw. sanktioniert wird.

2) Ein zweiter wichtiger Faktor besteht darin, dass Themen wie Macht und Machtmissbrauch in der Arbeitswelt bislang generell zu wenig ernst genommen und reflektiert werden. Dies mag auch an einer gewissen Tendenz zur Vermeidung liegen: Die meisten Menschen beschäftigen sich eher ungern mit der Frage, inwiefern sie (eigenen oder fremden) moralischen Ansprüchen eigentlich gerecht werden, oder mit so unangenehmen Themen wie Manipulation, Täuschung, Ausbeutung, Korruption und sexueller Grenzverletzung.

Empfehlungen

(11.1.) Angebot von und ggfs. Verpflichtung zur Teilnahme an Informationsveranstaltungen oder laufenden Fortbildungen zu Themen mit ethischem Bezug: Macht und Machtmissbrauch, psychische und physische Formen von Gewalt, GWP und Open Science, Sexuelle Selbstbestimmung, Mobbing, Bullying und Ostrazismus, Konformität und Widerspruch, Psychopathie und Narzissmus, Führung, Korruption inklusive rechtlicher Bestimmungen, Umgang mit Fehlern, Arbeitsrecht, Quantität versus Qualität in der Forschung, Whistleblowing und Beschwerdesystem.

(11.2.) Solche Veranstaltungen sollten dezidiert psychoedukativen Charakter haben. Die Bearbeitung konkreter Fallbeispiele erscheint dafür als Methode besonders geeignet. Zugleich aber ist die Existenz einer gewissen Grauzone in den meisten Handlungsfeldern anzuerkennen. Vorrangiges Ziel sollte sein, diese Themen überhaupt besprechbar und so die Teilnehmer:innen selbst urteilsfähiger und handlungsbereiter im Hinblick darauf zu machen.

(11.3.) Der Adressat:innenkreis solcher Maßnahmen umfasst Studierende, Professor:innen, Verwaltungspersonal, Hochschulleitungen, sowie alle wissenschaftlichen Mitarbeiter:innen - das heißt: *alle* Personen, die am Wissenschafts- und Lehrbetrieb beteiligt sind.

(11.4.) Der effizienteste Umgang mit dieser Aufgabe könnte darin bestehen, in Zusammenarbeit zwischen Kolleg:innen an verschiedenen Standorten gemeinsam ein entsprechendes Curriculum zu entwickeln. Ein mögliches Arbeitsergebnis wäre zum Beispiel die Bereitstellung entsprechender Informationsvideos im Netz.

(11.5.) Eine weitere mögliche Maßnahme besteht in der Ausbildung von Multiplikator:innen, die explizit als Ansprechpersonen für ethische Belange benannt werden. Dies könnte helfen, die Hemmschwellen, über schwierige Themen zu reden, zu reduzieren.

Problembereich 12. Schwache Kontroll- und Sanktionsinstanzen (P12)

Das Problem / die Betroffenen

In vielen Bereichen der Privatwirtschaft (z.B. Banken, Mobilität, Pharmaforschung, Nahrungsmittelindustrie) ist es seit Jahrzehnten selbstverständlich, dass das Handeln der Verantwortlichen im öffentlichen Interesse wenigstens stichprobenartig kontrolliert und gegebenenfalls auch sanktioniert wird. Dabei ist ebenso selbstverständlich, dass Kontrollen und Sanktionen auch *von extern* erfolgen, und nicht der jeweiligen Organisation selbst überlassen bleiben.

In der Wissenschaft liegt hingegen bis heute der Großteil der Verantwortung für Aufsicht, Kontrollen und eventuelle Sanktionen bei den Institutionen selbst. Fallberichte und systematische Untersuchungen belegen jedoch zunehmend, dass die gegenwärtigen Beschwerdesysteme wenig effektiv sind (Elson et al., 2021; Höbel et al., 2022). Beispielsweise zeigt die Umfrage des DGPs-Ombudsgremiums (Elson et al., 2021), dass nur ein sehr geringer Teil des vorkommenden Fehlverhaltens überhaupt jemals berichtet wird, und dass gegen die berichteten Missstände nur selten effektiv vorgegangen wird.

Wissenschaftliches Fehlverhalten schädigt die wissenschaftliche Community und die Allgemeinheit (P1). Nicht-wissenschaftliches Fehlverhalten (Ausbeutung, Bullying) schadet den Karrieremöglichkeiten und/oder der körperlichen und seelischen Gesundheit der jeweiligen Opfer (P4). Wird Fehlverhalten entdeckt, entsteht zudem oft ein immenser Reputationsschaden, sowohl für die unmittelbar Verantwortlichen, als auch für die betroffene Institution. Dieser Schaden ist umso größer, wenn auch noch deutlich wird, dass es zuvor einen zweifelhaften Willen zur Aufklärung oder aktive Vertuschungsversuche gegeben hat.

Ursachen / beitragende Faktoren

Eine ganze Reihe von Gründen tragen dazu bei, dass Kontroll- und Sanktionsmechanismen im Wissenschaftsbetrieb ihre Funktion zu häufig nur unzureichend erfüllen:

- 1) Potenzielle Hinweisgeber:innen oder Beschwerdeführer:innen kennen die zuständigen Stellen und Abläufe nicht.
- 2) Potenziellen Hinweisgeber:innen oder Beschwerdeführer:innen fehlt es an Vertrauen in die vorhandenen Beschwerdekanäle (Elson et al., 2021; Höbel et al., 2022). Dies kann in Zweifeln an der Vertraulichkeit oder Effektivität der Verfahren begründet sein, oder in Zweifeln an der Unparteilichkeit der zuständigen Personen. Alle drei Arten von Zweifeln scheinen durchaus berechtigt, wie persönliche Erfahrungsberichte gegenüber Mitgliedern der Kommission vermuten lassen.
- 3) Die Ermittlungs- und Sanktionsbefugnisse der Ombudspersonen sind eng begrenzt. So dürfen z.B. die Ombudspersonen verschiedener Hochschulen ihren jeweiligen Wissensstand nicht miteinander teilen, und die schwerwiegendste Sanktion, die die DFG bei Verstößen gegen die GWP vorsieht, ist eine fünfjährige Antragssperre.

- 4) Die zuständigen Stellen sind personell sehr schwach besetzt, sodass es schlicht an Personal mangelt, um die erforderliche Arbeit zu leisten.
- 5) Potenzielle Hinweisgeber:innen oder Beschwerdeführer:innen berichten von Angst vor Vergeltungsmaßnahmen seitens der Beschuldigten. Wenn es sich bei den Beschuldigten um Professor:innen handelt, verstärkt das Machtgefälle (P8) zwischen diesen und den potenziellen Beschwerdeführer:innen oft die entsprechenden Befürchtungen. So sind Erstere als Beamt:innen in der Regel quasi unkündbar, während Letztere oft auf zeitlich eng befristeten Stellen arbeiten (P7).
- 6) Möglichkeiten, Hinweisgeber:innen effektiv vor Vergeltungsversuchen seitens der Beschuldigten zu schützen, sind bislang eng begrenzt.
- 7) Die Effektivität eines internen Beschwerdeweges kann durch Interessenkonflikte geschmälert werden: Diese können sich aus dem Bestehen persönlicher und/oder beruflicher Beziehungen zwischen (internen) Ombudspersonen und denjenigen Personen ergeben, deren Verhalten Gegenstand einer Beschwerde ist. Auch stehen die Institutionen (und die *bei diesen angestellten* Ombudspersonen) regelmäßig vor dem Konflikt, dass ein Ernstnehmen angezeigter Probleme zu einem größeren Reputationsschaden führen könnte als der Versuch einer Verheimlichung derselben.
- 8) Die Tatsache, dass so wenige Fälle überhaupt angezeigt und dann von den vorhandenen Beschwerdesystemen bearbeitet werden, kann von Entscheidungstragenden als Hinweis auf eine niedrige Vorkommenshäufigkeit anzeigewürdiger Probleme und damit eine geringe Wichtigkeit des Themas insgesamt fehlinterpretiert werden.
- 9) Nicht selten besteht Unklarheit über Prozessabläufe und Zuständigkeiten auf Seiten der Institution (d.h. mangelhaft konkretisierte Compliance-Strukturen).
- 10) Auch mangelnde Bewusstheit über die geltenden ethischen Standards (P11) trägt wahrscheinlich dazu bei, dass potenzielle Hinweisgeber:innen sich oft nicht sicher sind, ob ihnen eine Meldung überhaupt "zusteht".

Angesichts dieser Hürden - gänzlich zum Nachteil potenzieller Whistleblower - verwundert es nicht, dass erfolgreiche Beschwerden im Wissenschaftssystem offenbar die Ausnahme darstellen (Elson et al., 2021; Höbel et al., 2022).

Empfehlungen

(12.1.) Spezifikation von hochschulinternen Meldesystemen und Bekanntmachung derselben auf leicht auffindbaren Webseiten. Hier muss insbesondere deutlich werden, welche Stelle für welche Art von Beschwerde (Wissenschaftliches Fehlverhalten, Bullying, Korruption, Betreuungsprobleme) zuständig ist, wie die entsprechenden Prozesse ablaufen, wie der vertrauliche Umgang mit Informationen (z.B. über die Identität hinweisgebender Personen) sichergestellt wird, und welche weiteren Schutzmechanismen solchen Personen zur Verfügung stehen.

(12.2.) Erhebliche Stärkung der zuständigen Stellen durch angemessene personelle und finanzielle Ausstattung sowie Verleihung ernstzunehmender Ermittlungs- und evtl. auch Sanktionsbefugnisse (z.B. Recht auf Akteneinsicht und ggfs. -sicherstellung bei allen an der Institution laufenden Forschungsprojekten). Die Kommission hält die Klärung der Rechtslage in diesem Bereich für ein wichtiges Nahziel, ebenso wie die Erarbeitung von mit möglichst geringem Zusatzaufwand implementierbaren, transparenten Prozessstandards.

(12.3.) Einrichtung unabhängiger Kontrollinstanzen (außerhalb der Forschungsinstitutionen) auf Länder- oder Bundesebene. Für diese gelten dieselben in Punkt 12.2. genannten Erfordernisse bzgl. Ressourcen und Befugnisse. Die Kommission schlägt vor, das Gremium *Ombudsman für die Wissenschaft* entsprechend zu stärken.

(12.4.) Standardmäßige Beteiligung einer unabhängigen *externen* Stelle zu Beginn jedes Beschwerdeverfahrens. Die Entscheidung darüber, ob einer initialen Beschwerde zunächst intern nachgegangen werden soll, darf nicht bei Personen liegen, die im Dienst der Institution stehen, an der die Beschwerde erfolgt.

(12.5.) Ablösung des bisher verbreiteten Prinzips, interne Ombudsstellen mit einzelnen Personen zu besetzen, durch *Ombudsgremien* als Regelfall (Mehraugenprinzip, besserer Umgang mit möglichen Befangenheiten).

(12.6.) Regelmäßige, *proaktive Nachfragen* (z.B. in Form von Umfragen oder persönlichen Turnusgesprächen) der Institutionen bei ihren Mitgliedern (auf allen Ebenen) hinsichtlich Entwicklungsbedarf und Verbesserungsmöglichkeiten in Strukturen und Prozessen, und als Maßnahme zur Senkung der Schwelle, Fehlverhalten anzuzeigen.

(12.7.) Erprobung der Möglichkeit *anonymer* Meldungen von Fehlverhalten.

(12.8.) Evaluationsforschung zur Effektivität bestehender Beschwerdeverfahren.

Die Kommission weist darauf hin, dass seit der Erstellung der ersten Version des vorliegenden Berichts das neue Hinweisgeberschutzgesetz (HinSchG) vom Bundestag beschlossen wurde, und zwischen diesem und einigen der hier gegebenen Empfehlungen (z.B. 12.1 und 12.3) eine hohe inhaltliche Überschneidung besteht.

Reduktion der mit einer Professur verbundenen Macht, u.a. durch Abgabe und Verteilung von Funktionen (siehe **8.1.**, **8.3.** bis **8.5.**, **8.9.** bis **8.12.**). Dies sollte die Hürden für eine Anzeige von Fehlverhalten erheblich senken.

Schaffung bzw. Stärkung einer kulturellen Sensibilität ("Awareness") für mögliche Ethik-Verstöße (siehe **11.1.** bis **11.5.**). Damit gemeint sind stärker geteilte, explizite Verhaltensstandards, die die *Einordnung* dessen, was beobachtet wird, erleichtern.

Literatur

- Abele-Brehm, A. E., & Bühner, M. (2016a). Wer soll die Professur bekommen? *Psychologische Rundschau*, 67(4), 250–261. <https://doi.org/10.1026/0033-3042/a000335>
- Abele-Brehm, A. E., & Bühner, M. (2016b). Überlegungen zur Optimierung von Berufungsverfahren in der Psychologie. *Psychologische Rundschau*, 67(4), 262–268. <https://doi.org/10.1026/0033-3042/a000333>
- Anderson, M. S., Martinson, B. C., & De Vries, R. (2007). Normative dissonance in science: Results from a National Survey of U.S. scientists. *Journal of Empirical Research on Human Research Ethics*, 2(4), 3–14. <https://doi.org/10.1525/jer.2007.2.4.3>
- Allen, L., Scott, J., Brand, A., Hlava, M., & Altman, M. (2014). Publishing: Credit where credit is due. *Nature*, 508(7496), 312–313. <https://doi.org/10.1038/508312a>
- Anglim, J., Lievens, F., Everton, L., Grant, S. L., & Marty, A. (2018). HEXACO personality predicts counterproductive work behavior and organizational citizenship behavior in low-stakes and job applicant contexts. *Journal of Research in Personality*, 77, 11–20. <https://doi.org/10.1016/j.jrp.2018.09.003>
- Arcudi, A., Cumurovic, A., Gotter, C., Graeber, D., Joly, P., Ott, V., Schanze, J.-L., Thater, S., Weltin, M., & Yenikent, S. (2019). *Doctoral Researchers in the Leibniz Association: Final Report of the 2017 Leibniz PhD Survey*. Leibniz PhD Network. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-61363-9>
- Beadle, B., Do, S., El Youssoufi, D., Felder, D., Gorenflos López, J., Jahn, A., Pérez-Bosch Quesada, E., Rottlieb, T., Rüter, F., Schanze, J.-L., Stroppe, A.-K., Thater, S., Verrière, A., & Weltin, M. (2020). Being a Doctoral Researcher in the Leibniz Association: 2019 Leibniz PhD Network Survey Report. Leibniz PhD Network. <https://www.ssoar.info/ssoar/handle/document/69403>
- Block, J. (1995). A contrarian view of the five-factor approach to personality description. *Psychological Bulletin*, 117(2), 187–215. <https://doi.org/10.1037/0033-2909.117.2.187>
- Brachem, J., Frank, M., Kvetnaya, T., Schramm, L. F., & Volz, L. (2022). Replikationskrise, p -hacking und Open Science. Eine Umfrage zu fragwürdigen Forschungspraktiken in studentischen Projekten und Impulse für die Lehre. *Psychologische Rundschau*, 73(1), 1–17. <https://doi.org/10.1026/0033-3042/a000562>
- Bramlage, J. K., Jackenkroll, B., Julmi, C., & Pereira, J. M. (2021). Konsequenzen unzumutbarer Aufgaben bei wissenschaftlichen Mitarbeiter*innen: Ein Pfadmodell [Konferenzbeitrag]. AKempor – Arbeitskreis Empirische Personal- und Organisationsforschung. Hamburg, Deutschland.
- Bundesministerium für Bildung und Forschung, InterVal GmbH & HIS-Institut für Hochschulentwicklung e.V. (2022). *Evaluation des Wissenschaftszeitvertragsgesetzes*. https://his-he.de/wp-content/uploads/2022/05/Bericht-WissZeitVG_220517.pdf

- Cao, M., & Drasgow, F. (2019). Does forcing reduce faking? A meta-analytic review of forced-choice personality measures in high-stakes situations. *Journal of Applied Psychology, 104*(11), 1347–1368. <https://doi.org/10.1037/apl0000414>
- Chapman, C. A., Bicca-Marques, J. C., Calvignac-Spencer, S., Fan, P., Fashing, P. J., Gogarten, J., Guo, S., Hemingway, C. A., Leendertz, F., Li, B., Matsuda, I., Hou, R., Serio-Silva, J. C., & Chr Stenseth, N. (2019). Games academics play and their consequences: how authorship, *h*-index and journal impact factors are shaping the future of academia. *Proceedings. Biological Sciences, 286*(1916), 20192047. <https://doi.org/10.1098/rspb.2019.2047>
- Christian, K., Johnstone, C., Larkins, J.-ann, Wright, W., & Doran, M. R. (2021). A survey of early-career researchers in Australia. *ELife, 10*. <https://doi.org/10.7554/elife.60613>
- Cohen, A., & Baruch, Y. (2022). Abuse and exploitation of doctoral students: A conceptual model for traversing a long and winding road to academia. *Journal of Business Ethics, 180*, 505-522. <https://doi.org/10.1007/s10551-021-04905-1>.
- Corstjens, J., Lievens, F., & Krumm, S. (2017). Situational judgement tests for selection. In H. W. Goldstein, E. D. Pulakos, J. Passmore, & C. Semedo (Eds.), *The Wiley Blackwell Handbook of the Psychology of Recruitment, Selection, and Employee Retention* (pp. 226–246). Hoboken, NJ: Wiley-Blackwell.
- DGPs Vorstand. (2022, 17. Januar). *#IchBinHanna: Tenure Track statt Befristung* [Presse-meldung]. <https://www.dgps.de/aktuelles/details/ichbinhanna-tenure-track-anstatt-befristung/>
- DHV (2022, Juni) Zu wenig Fortschritte bei Befristungsregelungen? *DHV-Newsletter Archiv*. <https://s3-eu-west-1.amazonaws.com/files.crsend.com/230000/230245/rss/media/13667635.htm>
- Diagnostik- und Testkuratorium (Hrsg.) (2018). *Personalauswahl kompetent gestalten: Grundlagen und Praxis der Eignungsdiagnostik nach DIN 33430*. Berlin: Springer.
- Dougherty, M. R., & Horne, Z. (2022). Citation counts and journal impact factors do not capture some indicators of research quality in the behavioural and brain sciences. *Royal Society Open Science, 9*, 220334. <http://doi.org/10.1098/rsos.220334>
- Elson, M., Fiedler, S., Kirsch, P., & Stahl, J. (2021, July 28). Wissenschaftliche Integrität in der deutschen akademischen Psychologie. <https://doi.org/10.17605/OSF.IO/MK9ST>
- Falagas, M.E., & Alexiou, V.G. (2008). The top-ten in journal impact factor manipulation. *Archivum Immunologiae et Therapiae Experimentalis, 56*, 223-226. <https://doi.org/10.1007/s00005-008-0024-5>
- Fischer, T., Tian, A. W., Lee, A., & Hughes, D. J. (2021). Abusive supervision: A systematic review and fundamental rethink. *The Leadership Quarterly, 32*(6), 101540. <https://doi.org/10.1016/j.leaqua.2021.101540>

- Flake, J. K., & Fried, E. I. (2020). Measurement schmeasurement: Questionable measurement practices and how to avoid them. *Advances in Methods and Practices in Psychological Science*, 3(4), 456–465. <https://doi.org/10.1177/2515245920952393>
- Fong, E. A., & Wilhite, A. W. (2017). Authorship and citation manipulation in academic research. *PloS one*, 12(12), e0187394. <https://doi.org/10.1371/journal.pone.0187394>
- Frith, U. (2020). Fast lane to slow science. *Trends in Cognitive Sciences*, 24(1), 1-2. <https://doi.org/10.1016/j.tics.2019.10.007>
- Gasparian, A.Y., Ayvazyan, L. & Kitas, G.D. (2013). Authorship problems in scholarly journals: considerations for authors, peer reviewers and editors. *Rheumatology International*, 33, 277–284. <https://doi.org/10.1007/s00296-012-2582-2>
- Gärtner, A., Leising, D., & Schönbrodt, F. D. (2022, November 25). Responsible research assessment II: A specific proposal for hiring and promotion in psychology. <https://doi.org/10.31234/osf.io/5yexm>
- Grijalva, E., & Newman, D. A. (2014). Narcissism and counterproductive work behavior (CWB): Meta-analysis and consideration of collectivist culture, big five personality, and narcissism's facet structure. *Applied Psychology*, 64(1), 93–126. <https://doi.org/10.1111/apps.12025>
- Höbel, M., Durglishvili, A., Reinold, J., & Leising, D. (2022). Sexual harassment and coercion in German academia: a large-scale survey study. *Sexual Offending: Theory, Research, and Prevention*, 17, 1-34. <https://doi.org/10.5964/sotrap.9349>
- Horbach, S. P., Breit, E., Halffman, W., & Mamelund, S.-E. (2020). On the willingness to report and the consequences of reporting research misconduct: The role of power relations. *Science and Engineering Ethics*, 26(3), 1595–1623. <https://doi.org/10.1007/s11948-020-00202-8>
- Jeder Fehler zählt. (n.d.). *Informationen zum Fehlerberichtssystem*. Retrieved August 31, 2022, from <https://jeder-fehler-zaehlt.de/>
- John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, 23(5), 524–532. <https://doi.org/10.1177/0956797611430953>
- Kasten, N., Freund, P. A., & Staufienbiel, T. (2020). “Sweet little lies”: An in-depth analysis of faking behavior on Situational Judgment Tests compared to personality questionnaires. *European Journal of Psychological Assessment*, 36(1), 136–148. <https://doi.org/10.1027/1015-5759/a000479>
- Kis, A., Tur, E.M., Lakens, D., Vaesen, K., & Houkes, W. (2022) Leaving academia: PhD attrition and unhealthy research environments. *PLoS ONE* 17(10), e0274976. <https://doi.org/10.1371/journal.pone.0274976>
- Konferenzrat der Psychologie-Fachschaften-Konferenzen. (2018). Positionspapier der Psychologie-Fachschaften-Konferenz (PsyFaKo) zum Thema “Replikationskrise und

- Open Science in der Psychologie“. <https://psyfako.org/wp-content/uploads/27-PsyFaKo-PP-Open-Science-Replikationskrise-und-Open-Science-in-der-Psychologie.pdf>
- Kuhnt, M., Reitz, T. & Wöhrle, P. (2022). Arbeiten unter dem Wissenschaftszeitvertrags-gesetz: Eine Evaluation von Befristungsrecht und -realität an deutschen Universitäten. <https://mittelbau.net/evaluation-des-wisszeitvg/>
- Kreitzer, R. J., & Sweet-Cushman, J. (2021). Evaluating student evaluations of teaching: A review of measurement and equity bias in sets and recommendations for ethical reform. *Journal of Academic Ethics*, 20(1), 73–84. <https://doi.org/10.1007/s10805-021-09400-w>
- Lasser, J., Bultema, L., Jahn, A., Löffler, M., Minneker, V., & van Scherpenberg, C. (2021). Power abuse and anonymous accusations in academia—Perspectives from early career researchers and recommendations for improvement. *Beiträge zur Hochschulforschung*, 43(1-2), 48-61.
- Leising, D., Thielmann, I., Glöckner, A., Gärtner, A., & Schönbrodt, F. (2022a). Ten steps toward a better personality science – how quality may be rewarded more in research evaluation. *Personality Science*, 3, 1-44. <https://doi.org/10.5964/ps.6029>
- Leising, D., Thielmann, I., Glöckner, A., Gärtner, A., & Schönbrodt, F. (2022b). Ten steps toward a better personality science – a rejoinder to the comments. *Personality Science*, 3, 1-15. <https://doi.org/10.5964/ps.7961>
- Lynam, D. R., Hyatt, C. S., Hopwood, C. J., Wright, A. G. C., & Miller, J. D. (2019). Should psychologists sign their reviews? Some thoughts and some data. *Journal of Abnormal psychology*, 128(6), 541–546. <https://doi.org/10.1037/abn0000426>
- Mackie, S. A., & Bates, G. W. (2018). Contribution of the doctoral education environment to Phd candidates' Mental Health Problems: A scoping review. *Higher Education Research & Development*, 38(3), 565–578. <https://doi.org/10.1080/07294360.2018.1556620>
- Marcus, B., & Schuler, H. (2004). Antecedents of counterproductive behavior at work: A general perspective. *Journal of Applied Psychology*, 89(4), 647–660. <https://doi.org/10.1037/0021-9010.89.4.647>
- Marcus, B., Lee, K., & Ashton, M. C. (2007). Personality dimensions explaining relationships between integrity tests and counterproductive behavior: Big five, or one in addition? *Personnel Psychology*, 60(1), 1–34. <https://doi.org/10.1111/j.1744-6570.2007.00063.x>
- Mete, Y. A. (2013). Relationship between organizational cynicism and ethical leadership behaviour: A study at higher education. *Procedia-Social and Behavioral Sciences*, 89, 476-483. <https://doi.org/10.1016/j.sbspro.2013.08.880>
- Moshagen, M., Hilbig, B. E., & Zettler, I. (2018). The dark core of personality. *Psychological Review*, 125(5), 656–688. <https://doi.org/10.1037/rev0000111>
- Müller, W. (2015). Assessment Center in Berufungsverfahren an der Universität Bremen. In C. Preus, S. Braun, T. Hentschel & D. Frey (Hrsg.). *Personalauswahl in der Wissenschaft* (S. 261-267). Berlin: Springer.

- Niemann, Y. F., Muhs, G. G. y, & Gonzalez, C. G. (Eds.). (2020). *Presumed incompetent II: Race, class, power, and resistance of women in Academia*. Utah State University Press.
- O'Boyle, E. H., Forsyth, D. R., Banks, G. C., & McDaniel, M. A. (2012). A meta-analysis of the Dark Triad and work behavior: A Social Exchange perspective. *Journal of Applied Psychology, 97*(3), 557–579. <https://doi.org/10.1037/a0025679>
- O'Connor, P. J., Stone, S., Walker, B. R., & Jackson, C. J. (2017). Deviant behavior in constrained environments: Sensation-seeking predicts workplace deviance in shallow learners. *Personality and Individual Differences, 108*, 20–25. <https://doi.org/10.1016/j.paid.2016.11.062>
- Olsthoorn, L. H. M., Heckmann, L. A., Filippi, A., Vieira, R. M., Varanasi, R. S., Lasser, J., Bäuerle, F., Zeis, & P., Schulte-Sasse, R. (2020). *PhDnet Report 2019*. Max Planck PhDnet. https://pure.mpg.de/pubman/faces/ViewItemOverviewPage.jsp?itemId=item_3243876_4
- Ones, D. S., & Viswesvaran, C. (2001). Integrity tests and other criterion-focused occupational personality scales (cops) used in personnel selection. *International Journal of Selection and Assessment, 9*(1-2), 31–39. <https://doi.org/10.1111/1468-2389.00161>
- Ones, D. S., Viswesvaran, C. & Schmidt, F. L. (1993). Comprehensive meta-analysis of integrity tests validities: Findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology, 78*(4), 679-703. <https://doi.org/10.1037/0021-9010.78.4.679>
- Ones, D. S., Viswesvaran, C. & Schmidt, F. L. (2012). Integrity tests predict counterproductive work behavior and job performance well: Comment on Van Iddekinge, Roth, Raymond, and Odle-Dusseau (2012). *Journal of Applied Psychology, 97*(3), 537-542. <https://doi.org/10.1037/a0024825>
- Open Science Collaboration (2015). PSYCHOLOGY. Estimating the reproducibility of psychological science. *Science, 349*(6251), aac4716. <https://doi.org/10.1126/science.aac4716>
- Palmer, J. C., Holmes, R. M., & Perrewé, P. L. (2020). The cascading effects of CEO Dark Triad personality on subordinate behavior and firm performance: A multilevel theoretical model. *Group & Organization Management, 45*(2), 143–180. <https://doi.org/10.1177/1059601120905728>
- Paulhus, D. L., & Williams, K. M. (2002). The Dark Triad of personality: Narcissism, machiavellianism, and psychopathy. *Journal of Research in Personality, 36*(6), 556–563. [https://doi.org/10.1016/s0092-6566\(02\)00505-6](https://doi.org/10.1016/s0092-6566(02)00505-6)
- Peukert, K., Jacobi, L., Geuer, J., Paredes Cisneros, I., Löffler, M., Lienig, T., Taylor, S., Gusic, M., Novakovic, N., Kuhl, T., Ordini, E., Runge, A., Samoylov, O., Härtel, M., Amend, A.-L., & Nagel, M. (2020). *Survey Report 2019*. Helmholtz Juniors. https://www.helmholtz.de/fileadmin/user_upload/06_jobs_talente/Helmholtz-Juniors/Survey_Report2019.pdf

- Peus, C., Braun, S., Hentschel, T. & Frey, D. (Hrsg.) (2015). *Personalauswahl in der Wissenschaft*. Berlin: Springer.
- Pletzer, J. L., Bentvelzen, M., Oostrom, J. K., & de Vries, R. E. (2019). A meta-analysis of the relations between personality and workplace deviance: Big five versus hexaco. *Journal of Vocational Behavior*, 112, 369–383. <https://doi.org/10.1016/j.jvb.2019.04.004>
- Prinz, F., Schlange, T., & Asadullah, K. (2011). Believe it or not: How much can we rely on published data on potential drug targets? *Nature Reviews Drug Discovery*, 10(9), 712–712. <https://doi.org/10.1038/nrd3439-c1>
- Pruschak, G., & Hopp, C. (2022). And the credit goes to ... - Ghost and honorary authorship among social scientists. *PloS one*, 17(5), e0267312. <https://doi.org/10.1371/journal.pone.0267312>
- Pusch, N., & Holtfreter, K. (2020). Individual and organizational predictors of white-collar crime: A meta-analysis. *Journal of White Collar and Corporate Crime*, 2(1), 5–23. <https://doi.org/10.1177/2631309x19901317>
- Rennie, D. (2016). Let's make peer review scientific. *Nature*, 535, 31–33. <https://doi.org/10.1038/535031a>
- Regler, B., Einhorn, L., Lasser, J., Vögele, M., Elizarova, S., Bäuerle, F., Wu, C., Förste, S., Shenolikar, J., & Group 2018, P. S. (2019). *PhDnet Report 2018*. Max Planck PhDNet. <https://doi.org/10.17617/2.3052826>
- Runge, J. M., Lang, J. W. B., Zettler, I., & Lievens, F. (2020). Predicting counterproductive work behavior: Do implicit motives have incremental validity beyond explicit traits? *Journal of Research in Personality*, 89, 104019. <https://doi.org/10.1016/j.jrp.2020.104019>
- Scherpenberg, C., Bultema, L., Jahn, A., Löffler, M., Minneker, V., & Lasser, J. (2021). *Manifestations of power abuse in academia and how to prevent them*. Elephant in the Lab. <https://doi.org/10.5281/zenodo.4608392>
- Schönbrodt, F. D., Gärtner, A., Frank, M., Gollwitzer, M., Ihle, M., Mischkowski, D., ... Leising, D. (2022, November 25). Responsible Research Assessment I: Implementing DORA for hiring and promotion in psychology. <https://doi.org/10.31234/osf.io/rgh5b>
- Schraudner, M., Striebing, C., & Hochfeld, K. (2019). *Work culture and work atmosphere in the Max Planck Society*. Fraunhofer Center for Responsible Research and Innovation. <https://www.mpg.de/14284109/mpg-arbeitskultur-ergebnisbericht-englisch.pdf>
- Schwarzinger, D., & Schuler, H. (2016). *Dark Triad of Personality at Work (TOP)*. Göttingen: Hogrefe.
- Schwarzinger, D. (2020). *Die Dunkle Triade der Persönlichkeit in der Personalauswahl*. Göttingen: Hogrefe.
- Strobel, A., Franke-Bartholdt, L. & Kersting, M. (2018). Eignungsinterviews / direkte mündliche Befragungen. In: Diagnostik- und Testkuratorium (Hrsg.), *Personalauswahl*

- kompetent gestalten: Grundlagen und Praxis der Eignungsdiagnostik nach DIN 33430* (S. 66-83). Berlin: Springer.
- Tennant, J.P., Ross-Hellauer, T. (2020). The limitations to our understanding of peer review. *Research Integrity and Peer Review*, 5, 6. <https://doi.org/10.1186/s41073-020-00092-1>
- Van Iddekinge, C. H., Roth, P. L., Raymark, P. H., & Odle-Dusseau, H. N. (2012). The criterion-related validity of integrity tests: An updated meta-analysis. *Journal of Applied Psychology*, 97(3), 499–530. <https://doi.org/10.1037/a0021196>
- Vazire, S., & Holcombe, A. O. (2021). Where are the self-correcting mechanisms in science? *Review of General Psychology*, 26(2), 212–223. <https://doi.org/10.1177/10892680211033912>
- Zettler, I., Thielmann, I., Hilbig, B. E., & Moshagen, M. (2020). The nomological net of the HEXACO MODEL OF PERSONALITY: A large-scale meta-analytic investigation. *Perspectives on Psychological Science*, 15(3), 723–760. <https://doi.org/10.1177/1745691619895036>