

# Forschungstransparenz als hohes wissenschaftliches Gut stärken

## Konkrete Ansatzmöglichkeiten für Psychologische Institute

Felix D. Schönbrodt, Markus Maier, Moritz Heene und Markus Bühner

**Zusammenfassung:** Groß angelegte Replikationsprojekte der letzten Jahre legen ein aus unserer Sicht beunruhigendes Ausmaß an nicht-replizierbaren Befunden in der wissenschaftlichen Literatur nahe, sowohl in der Psychologie als auch in anderen Disziplinen. Basierend auf einer Analyse einiger Ursachen dieser Situation argumentieren wir, dass der Wandel hin zu einer offenen Wissenschaft („Open Science“) eine Konsequenz aus der Glaubwürdigkeitskrise sein muss. Wir plädieren für konkrete und machbare Änderungen in den Arbeitseinheiten und Instituten vor Ort, und zeigen exemplarisch, welche Schritte am Department Psychologie der Ludwig-Maximilians-Universität München umgesetzt wurden. Diese Schritte betreffen Anreizstrukturen, die Forschungskultur, die Lehre und die Verzahnung mit der Ethikkommission. Sie haben das Ziel, eine reproduzierbarere und glaubwürdigere Forschung zu unterstützen, ohne unnötige bürokratische Belastungen zu erzeugen.

**Schlüsselwörter:** Replikationskrise, Glaubwürdigkeitskrise, Anreize, Qualitätssicherung

### Fostering Research Transparency as a Key Property of Science: Concrete Actions for Psychological Departments

**Abstract:** Recent large-scale replication projects suggest an amount of nonreplicable results in the scientific literature, in psychology but also in other sciences, which is concerning from our point of view. We analyze some causes for this situation, and argue that the change toward more research transparency (“open science“) must be one consequence that should be drawn from the credibility crisis. We call for feasible changes in the local research units and departments and show as an example the steps that have been taken at the Department of Psychology of the Ludwig-Maximilians-Universität München. These changes concern incentive structures, research culture, teaching, and a close integration with the local ethics committee. The goal is to foster a more credible and more reproducible research output without generating unnecessary bureaucratic overhead.

**Keywords:** replication crisis, credibility crisis, incentives, research quality

Groß angelegte Replikationsprojekte der letzten Jahre haben gezeigt, dass eine überraschend hohe Anzahl an nicht-replizierbaren Befunden in der wissenschaftlichen psychologischen Literatur existiert. Je nach Zusammenstellung der zu replizierenden Studien befand sich die Quote der erfolgreichen Replikationen nur zwischen 12% und 39% (z. B. Klein et al., 2014; Nosek & Lakens, 2014; Open Science Collaboration, 2015). Auch in anderen Feldern wird dies als Problem erkannt und diskutiert, zum Beispiel in der Medizin (Begley & Ellis, 2012: 11% Replikationsquote; Prinz, Schlange & Asadullah, 2011: 20–25% Replikationsquote). Wir glauben, dass sich unser Fach, nicht zuletzt durch die kritische Diskussion in der nationalen und internationalen Presse, in einer Glaubwürdigkeitskrise befindet.

Aufgrund der kumulativen Evidenz der letzten Jahre ist aus unserer Sicht folglich nicht mehr die Frage, ob wir in vielen Bereichen der Psychologie (und anderen Disziplinen) eine Glaubwürdigkeitskrise haben, sondern welche Konsequenzen sinnvollerweise aus der Krise gezogen werden können, um verlorenes Vertrauen wiederherzustellen und die Glaubwürdigkeit unserer Forschung deutlich zu erhöhen. Im Folgenden möchten wir (1) einige Ursachen für die Replikationskrise andiskutieren, (2) aufzeigen wie eine offene Forschung („Open Science“) eine Antwort auf die Krise sein kann, und (3) exemplarisch vorstellen, welche Schritte konkret am Department Psychologie der LMU München als Reaktion auf die Krise umgesetzt wurden.

Wir danken Anne Scheel und Caroline Zygarr für hilfreiche Anmerkungen zum Manuskript.

## Einige Ursachen für die Replikationskrise

Warum kann ein großer Anteil unserer wissenschaftlichen Literatur nicht erfolgreich repliziert werden? Aus einer evolutionären Perspektive lässt sich diese „Warum“-Frage auf zwei Ebenen beantworten (Tinbergen, 1963): Man kann in der proximativen Betrachtungsweise nach den Wirkursachen fragen. Diese stellen die konkreten Mechanismen dar, die ein Phänomen verursachen. Die ultimative Betrachtungsweise hingegen fragt nach der Zweckursache: Was ist der Nutzen eines bestimmten Verhaltens oder Merkmals?

Als Wirkmechanismen möchten wir drei Prozesse diskutieren, die zu einer Inflation von falsch-positiven (und somit nicht replizierbaren) Ergebnissen führen: (a) unabsichtliche Fehler, (b) Flexibilität in der Datenanalyse und (c) *questionable research practices* (QRPs). Als Zweckursache lässt sich (d) das aktuelle Anreizsystem in der Wissenschaft identifizieren (vgl. auch hierzu Abele-Brehm & Bühner, 2016), welches einen Forschungsstil belohnt, der nicht unbedingt zu belastbaren Erkenntnissen führt.

(a) *Unabsichtliche Fehler*. Während mit „Replikation“ der Versuch gemeint ist, ein Forschungsergebnis in einer neuen Stichprobe noch einmal zu zeigen, zielt die „Reproduktion“ eines berichteten Befundes lediglich darauf ab, anhand desselben Datensatzes noch einmal zu den berichteten Ergebnissen zu kommen (Asendorpf et al., 2013). Ist ein Befund aufgrund eines Rechenfehlers falsch-positiv und nicht reproduzierbar, so ist dieser auch nicht replizierbar. Wie steht es um die Reproduzierbarkeit der Forschung? Die Zeitschrift *Quarterly Journal of Political Science* verlangt seit 2005, dass bei Einreichungen von Manuskripten die Primärdaten und die Analyseskripte mitgeliefert werden. Während des Reviewprozesses wird die Reproduzierbarkeit überprüft, also nachgerechnet, ob die Zahlenangaben im Manuskript mit den Ergebnissen der Analyseskripte von den Autoren selbst übereinstimmen. Diese scheinbar triviale Anforderung an wissenschaftliche Sorgfalt wurde regelmäßig verletzt (Eubank, 2016): 58 % aller Einreichungen hatten Diskrepanzen bis hin zu falschen Vorzeichen, und das, obwohl die Autoren vorher wussten, dass ihre Analysen verifiziert werden. Ähnliche Ergebnisse zeigte eine Untersuchung aus der Ökonomie, welche die Ergebnisse von nur 49 % der 59 untersuchten Artikel reproduzieren konnte (Chang & Li, 2015).

(b) *Flexibilität in der Datenanalyse*. Gelman und Loken (2014) haben das Bild des „garden of forking paths“ geprägt. Diese Metapher beschreibt, dass Forscher während des Prozesses der Datenanalyse viele Ad-hoc-Entscheidungen treffen, welche kontingent zu den Daten sind

(z.B.: Wie gehe ich genau mit Ausreißern um?). Diese Entscheidungen führen jeweils zu unterschiedlichen Analysepfaden, wobei einige dieser Pfade zu signifikanten Ergebnissen führen, andere nicht. Wenn nun im Prozess der Datenanalyse ein optionales Stoppen stattfindet, in dem Sinne, dass man solange Analysevarianten ausprobiert bis ein signifikantes Ergebnis erreicht wurde, erhöht dies die Wahrscheinlichkeit von falsch-positiven und somit nicht replizierbaren Ergebnissen.

(c) *Questionable research practices* (QRPs). Unter QRPs werden Praktiken der Datenerhebung und -analyse verstanden, die in der Wahrnehmung der Forscher noch nicht als offensichtlicher Betrug gelten, sondern eher eine Grauzone von „halb-legalen“ Praktiken darstellen – oft dadurch legitimiert, „dass es ja alle so machen“. Dazu gehört zum Beispiel, dass man eine Vielzahl von Variablen erhebt, aber nur die berichtet, bei denen eine statistische Analyse ein signifikantes Ergebnis erbrachte, oder dass man bei fast-signifikanten Ergebnissen so lange nacherhebt, bis der  $p$ -Wert unter die kritische Grenze rutscht. Wenn diese und andere Praktiken in Kombination angewendet werden, kann die Wahrscheinlichkeit für ein falsch-positives Ergebnis vom nominalen 5 %-Niveau leicht auf über 50 % gehoben werden (Simmons, Nelson & Simonson, 2011). In großen Umfragen versuchten John, Loewenstein und Prelec (2012) und Fiedler und Schwarz (2015) die Prävalenz solcher Praktiken zu erheben. Die Ergebnisse beider Umfragen zeigen, dass diese Techniken regelmäßig und in nicht zu vernachlässigendem Ausmaß angewandt werden.

(d) *Anreizstrukturen*. Solange die Anreizstrukturen Quantität (möglichst lange Publikationslisten,  $h$ -index) und unrealistisch perfekte Ergebnisse belohnen (Gervais, Jewell, Najle & Ng, 2015; Giner-Sorolla, 2012), wird die Inflation falsch-positiver Ergebnisse zunehmen, insbesondere da häufig nur signifikante Ergebnisse publiziert werden, die zum Teil aus Studien mit geringer Power stammen. Damit wird ein Forschungsstil gefördert, der die Ausbeutung des Stichprobenfehlers für Forscher attraktiv erscheinen lässt (Bakker, van Dijk & Wicherts, 2012; Engel, 2015; Higginson & Munafò, 2016; Smaldino & McElreath, 2016).

## Forschungstransparenz als (Teil-)Antwort

Zusammenfassend muss man davon ausgehen, dass ein großer Anteil der publizierten Ergebnisse entweder auf einem Rechenfehler beruht, durch QRPs „schöngerechnet“ wurde oder durch unintentionale Biases während der

Datenanalyse stärker dargestellt wurde, als es angemessen wäre. Viele dieser proximalen Mechanismen funktionieren auch unbewusst, so dass nicht unbedingt Absicht unterstellt werden muss. In Kombination tragen diese neben anderen Faktoren jedoch dazu bei, dass ein Großteil der Literatur wie oben berichtet nicht replizierbar ist. Welche konkreten Maßnahmen könnten zu einer Verbesserung der Situation führen?

Im Folgenden möchten wir argumentieren, dass Forschungstransparenz (aktuell auch oft unter dem etwas breiteren Begriff der „Open Science“ zusammengefasst) ein wichtiger Baustein zur Lösung dieser Probleme ist.

*Offene Daten.* Die bisher erfolgten Korrekturen von Rechenfehlern konnten nur stattfinden, wenn die Primärdaten verfügbar waren<sup>1</sup>. Gleichzeitig zeigten Publikationen mit offenen Daten weniger Rechenfehler und eine starke Evidenz für die berichteten Effekte im Vergleich zu Publikationen ohne bereitgestellte Daten (Wicherts, Bakker & Molenaar, 2011). Offene Daten sind also (momentan zumindest) ein valides Qualitätsmerkmal. Offene Primärdaten ermöglichen weiterhin, im Sinne des „garden of forking paths“ alternative plausible Analysevarianten zu explorieren. So kann die Robustheit der Ergebnisse gegenüber Variationen in den Analysepfaden festgestellt werden, welche wiederum einen Rückschluss auf die vermutliche Replizierbarkeit erlauben (Simonsohn, Simmons & Nelson, 2015; Steegen, Tuerlinckx, Gelman & Vanpaemel, 2016).

Die Bereitstellung von Daten zum Zwecke der Reproduktion publizierter Ergebnisse wird häufig als eine Selbstverständlichkeit guter wissenschaftlicher Praxis angesehen. Zum Beispiel muss bei jeder Publikation in einer APA-Zeitschrift unterschrieben werden, dass die Daten zum Zwecke der Nachvollziehbarkeit von Analysen jedem bereitgestellt werden. Dies ist jedoch praktisch nicht gegeben, denn Untersuchungen zeigten, dass die Primärdaten nur in 27 % (Wicherts, Borsboom, Kats & Molenaar, 2006), bzw. 38 % (Vanpaemel, Vermorgen, Deriemaecker & Storms, 2015) tatsächlich verfügbar gemacht wurde, und das oft auch nur nach mehrmaligem Nachhaken und monatelangem Warten. Teilweise fordern Datenberesteller unrealistisch hohe Geldbeträge (Ferguson, 2016) oder stellen so restriktive Auflagen an Datennachnutzer, dass eine Reproduktion oder sinnvolle Reanalyse von Datensätzen unmöglich ist (Coyne, 2016). Auch aus diesem Grund fordern DFG und DGPs auch einen prinzipiell unbeschränkten Zugang zu Primärdaten.

*Reproduzierbare Analyseskripte.* Wenn zusammen mit den Primärdaten auch die Datenanalysen als Skripte ab-

gespeichert werden (z. B. SPSS-Syntax oder R-Skripte), ist jede Analyse mit allen Optionen exakt reproduzierbar. Werden vor der Publikation alle Skripte unter den Koautoren zirkuliert, kann von denen im Sinne eines „code reviews“ auch die Qualität, Dokumentation und Fehlerfreiheit der Analysen überprüft werden.

*Präregistrierungen.* Legt man sich vor der Datenerhebung auf die Hypothesen, die fokalen Variablen und die Vorgehensweise in der statistischen Auswertung fest, dann sind die Freiheitsgrade so eingeschränkt, dass die meisten Varianten von *p*-hacking verhindert werden. Gleichzeitig zieht eine Präregistrierung eine klare Trennlinie zwischen konfirmatorischer und explorativer Forschung (Wagenmakers, Wetzels, Borsboom, van der Maas & Kievit, 2012). Gerade durch diese Vorabfestlegung gewinnt eine präregistrierte Studie an Glaubwürdigkeit (Chambers, Feredoes, Muthukumaraswamy & Etchells, 2014). Auch wenn empirische Daten dazu bisher fehlen, kann man davon ausgehen, dass Effekte, die in präregistrierten Studien gezeigt wurden, eine höhere Wahrscheinlichkeit für eine erfolgreiche Replikation haben als diese aus nicht präregistrierten Studien.

*Anreizstrukturen.* Gerade unter Nachwuchswissenschaftlern besteht die Sorge, dass ein Engagement in Forschungstransparenz strategisch zu Nachteilen in der individuellen Karriere führt. Wer auf *p*-hacking verzichtet, mit hoher Teststärke forscht (was im Regelfall zu größerem Aufwand bei der Datenerhebung führt), und seine imperfekten Primärdaten offen zur kritischen Begutachtung bereitstellt, hat möglicherweise eine geringere Menge an Publikationen als Mitbewerber, die diese Strategien nicht nutzen (John et al., 2012). Wenn die Wissenschaft als System nicht genau diejenigen Forscher herausselektieren will, die aufwändige Maßnahmen ergreifen, um belastbare Erkenntnisse zu produzieren, müssen die Auswahlkriterien bei der Vergabe von Stellen, Professuren und Forschungsmitteln so angepasst werden, dass gute Forschung belohnt, aber zumindest nicht bestraft wird (Nelson, Simmons & Simonsohn, 2012; Nosek, Spies & Motyl, 2012).

## Maßnahmen auf Ebene der Zeitschriften, Fachgesellschaften und Drittmittelgeber

Einige Veränderungen haben bereits auf der Ebene von Zeitschriften stattgefunden. So bieten viele Zeitschriften

<sup>1</sup> Der GRIM-Test von Brown und Heathers (2016) ist eine Methode, mit der Rechenfehler auch anhand von Summenstatistiken entdeckt (wenn auch nicht korrigiert) werden können.

nun „Badges“ für Forschungstransparenz an (Open Data, Open Material, und Preregistered Study; <https://osf.io/tvyxz>), zum Beispiel Psychological Science, das European Journal of Personality, oder Social Psychology. Im Jahr 2015 wurden auch die Transparency and Openness Promotion (TOP) Guidelines für Zeitschriften eingeführt (Nosek et al., 2015), welche acht Transparenzstandards mit jeweils vier Abstufungen definieren. So reicht zum Beispiel der Standard „Datentransparenz“ von „Zeitschrift ermuntert zu offenen Daten oder sagt nichts“ (Level 0) bis zu „Die Analysen werden basierend auf den öffentlichen unverfügbaren Daten von unabhängiger Stelle nachgerechnet und verifiziert“ (Level 3). Im November 2017 hatten bereits über 5000 Zeitschriften und 78 Organisationen die Richtlinien unterzeichnet<sup>2</sup>.

Auch auf der Ebene der Wissenschaftsorganisationen und der Fachgesellschaften ist vieles in Bewegung. So bekennt sich die DFG in ihren aktualisierten Leitlinien zum Umgang mit Forschungsdaten vom September 2015<sup>3</sup> klar zu offenen Daten. Sie betont dort, dass Daten, die aus öffentlich-rechtlichen Mitteln finanziert wurden, im Grundsatz der Öffentlichkeit zum Zwecke der Nachprüfbarkeit und Nachnutzung frei zur Verfügung stehen. Die EU hat Open Science ebenfalls zu einem Kernthema gemacht: „The European Commission is now moving decisively from ‚Open access‘ into the broader picture of ‚Open science‘“<sup>4</sup>. Die Deutsche Gesellschaft für Psychologie (DGPs) greift das Thema ebenfalls vielfältig auf, zum Beispiel durch die im September 2016 verabschiedeten Empfehlungen zum „Umgang mit Forschungsdaten“ (Schönbrodt, Gollwitzer & Abele-Brehm, 2016) oder den Empfehlungen der DGPs-Kommission „Qualität der psychologischen Forschung“ (Fiedler et al., 2015).

## Maßnahmen an den psychologischen Instituten vor Ort

Diese Reformen auf institutioneller Ebene sind sehr zu begrüßen. Welche konkreten Möglichkeiten gibt es jedoch für einzelne Wissenschaftler und an den psychologischen Instituten, um die Forschungstransparenz zu erhöhen?

Im Folgenden möchten wir einige konkrete Schritte vorstellen, die das Department Psychologie an der Ludwig-Maximilians-Universität München mit dem Ziel unternommen hat, die Forschungstransparenz zu erhöhen. Wir wollen exemplarisch aufzeigen, welche konkreten Schritte möglich sind und aus der Praxis berichten, was funktioniert hat. Dabei möchten wir betonen, dass das hier skizzierte „Münchner Modell“ nur einer von vielen möglichen Wegen ist, diese Ziele zu befördern. Es haben sich an mehreren deutschsprachigen Universitäten vergleichbare Initiativen zusammengeschlossen<sup>5</sup>, welche dieselben Ziele vertreten, aber durchaus unterschiedliche Schwerpunkte und Wege der Umsetzung wählen. Eine Pluralität in den Ansätzen ist begrüßenswert, da sich nur so gute Strategien der Zielerreichung von schlechten trennen lassen. Die verschiedenen Initiativen haben sich im „Netzwerk der Open-Science-Initiativen“ (NOSI) zusammengeschlossen (<https://osf.io/tbkzh>) und stehen im regelmäßigen Austausch. Neue Initiativen oder Interessenten sind herzlich willkommen in dem offenen Netzwerk.

Bei allen Schritten an unserem Department geht es uns darum, positive Anreize zu setzen, und einen inklusiven Dialog mit möglichst vielen Beteiligten zu erreichen. Es geht nicht darum, eine „Replikationspolizei“ zu installieren, die Forschung am Department in „gut“ und „schlecht“ einzuteilen oder gar das Publizieren durch „neue Transparenzhürden“ zu erschweren. Es geht uns auch nicht darum, zurückzuschauen und frühere Forschung unseres oder anderer Departments mithilfe von Indizes zu be- oder verurteilen. Es geht darum die Forschungstransparenz als ein hohes wissenschaftliches Gut zu stärken.

All das geschieht in dem Bewusstsein, dass wir in vielen Punkten Neuland betreten, dass sich die verschiedenen Teilbereiche der Psychologie unterscheiden, und dass Maßnahmen nach einiger Zeit daraufhin evaluiert werden sollten, ob sie den gewünschten Effekt erzielen und ob unerwünschte Nebenwirkungen auftreten (Ioannidis, 2014). Maßnahmen sollten möglichst wenig bürokratischen Aufwand erzeugen: Sie sollen Forschung besser machen, aber nicht von der eigentlichen Forschung abhalten. Im Folgenden möchten wir skizzieren, was im Laufe der letzten zwei Jahre entwickelt wurde.

<sup>2</sup> Siehe <http://centerforopenscience.org/top/>

<sup>3</sup> [http://www.dfg.de/foerderung/antragstellung\\_begutachtung\\_entscheidung/antragstellende/antragstellung/nachnutzung\\_forschungsdaten/](http://www.dfg.de/foerderung/antragstellung_begutachtung_entscheidung/antragstellende/antragstellung/nachnutzung_forschungsdaten/); siehe auch [http://www.dfg.de/download/pdf/foerderung/programme/lis/ua\\_inf\\_empfehlungen\\_200901.pdf](http://www.dfg.de/download/pdf/foerderung/programme/lis/ua_inf_empfehlungen_200901.pdf)

<sup>4</sup> <https://ec.europa.eu/programmes/horizon2020/en/h2020-section/open-science-open-access>

<sup>5</sup> Im November 2017 gab es acht offizielle Open-Science-Initiativen an deutschsprachigen Instituten: LMU München, Universität Koblenz-Landau, FernUniversität Hagen, Universität Münster, Universität Göttingen, Universität Wien, Humboldt-Universität Berlin und Universität Leipzig (sortiert nach Gründungsdatum).

## Selbstverpflichtung zu Forschungstransparenz

Die effizienteste Maßnahme gegen „questionable research practices“ ist es, sie nicht mehr anzuwenden. Daher wurde eine Selbstverpflichtung zu Forschungstransparenz entwickelt, die in 12 Punkten darlegt, wie die Unterzeichner in Zukunft agieren werden (Schönbrodt, Maier, Heene & Zehetleitner, 2015; <http://www.researchtransparency.org>). Dabei geht es erstens um Aspekte der eigenen Forschung (z. B. bei jeder Erstautor-Publikation die zugrundeliegenden Primärdaten offenzulegen). Zweitens werden Unterzeichner in ihrer Rolle als Reviewer nach offenen Daten und reproduzierbaren Skripten fragen, oder nach einer expliziten Begründung, wenn dies nicht möglich ist (vgl. auch die Peer Reviewers' Openness Initiative; Morey et al., 2016). Drittens erwarten Unterzeichner bei der Betreuung von Dissertationen dieselben Kriterien der Transparenz, die sie selbst anwenden. Bei einer Serie von Experimenten in einer Dissertation wird erwartet, dass mindestens eines davon präregistriert wird. Diese Erwartungen werden zu Beginn der Promotion kommuniziert, und es werden post-graduale Fortbildungen angeboten, um die notwendigen Fertigkeiten zu erlernen. Viertens werden Unterzeichner aktiv für die Werte von Forschungstransparenz eintreten, etwa in Komitees (zum Beispiel in Berufungs- oder Tenure-Track-Verfahren), als Editoren und in den Fachgesellschaften. Zu jedem Aspekt der Selbstverpflichtung kann es natürlich begründete Ausnahmen geben, zum Beispiel, wenn offene Daten aus Datenschutzgründen nicht möglich sind.

## Das Open-Science-Komitee

Im Juli 2015 hat das Leitungskollegium des Departments einstimmig beschlossen, ein Open-Science-Komitee (OSK) einzurichten. Das Komitee hat mehr als 40 Mitglieder, wobei alle Lehrstühle und alle Statusgruppen von Studierenden bis zu Professorinnen und Professoren vertreten sind. Das Komitee hat vier Aufgaben:

*Aktuelle internationale Entwicklungen zum Thema „Open Science“ beobachten und in das Department kommunizieren.* In der Psychologie (und der Wissenschaft im Allgemeinen) ist gerade viel Bewegung was die Etablierung neuer Standards und neuer Technologien betrifft. Das OSK informiert regelmäßig in seinen Treffen und in Newslettern über diese Entwicklungen, damit keine strategischen Nachteile durch Unkenntnis dieser Veränderungen entstehen.

*Konkrete Vorschläge bezüglich Tenure-Track-Kriterien, Berufungsverfahren, Doktorandenbetreuung, etc. (weiter)entwickeln.* Viele strukturelle Bedingungen können einen Nährboden für QRPs bieten. Hängt zum Beispiel die Beurteilung einer Dissertation oder die Entfristung in einem Tenure-Track-Verfahren davon ab, ob diese letzte Studie mit  $p = .08$  noch publiziert werden kann, stellt das einen enormen Anreiz für eine „kreative“ Datenanalyse dar. Das OSK entwickelt Ideen, wie solche Anreizstrukturen verändert werden können.

*Workshops und Fortbildungen organisieren.* Transparenz erfordert Wissen und Fertigkeiten, die in der bisherigen wissenschaftlichen Ausbildung kaum vermittelt wurden. In regelmäßigen Workshops vermitteln wir den Forschern unseres Departments Fertigkeiten in den neuen Techniken (z. B. „Wie mache ich eine gute Präregistrierung?“, „Einführung in das Open Science Framework“ oder „Do's and Don't's der Datenanalyse“).

*Die departmentsinterne Diskussion über mögliche Konsequenzen aus der Replikationskrise zusammenführen; ausloten, ob gemeinsame Standards entwickelt werden können.* Die Mitglieder des OSK dienen als Botschafter der einzelnen Arbeitseinheiten: Sie sollen aktuelle Informationen aus den Komitee-Sitzungen in die Lehrstühle kommunizieren, und gleichzeitig offene Fragen und Diskussionspunkte aus der Praxis der Lehrstühle mitbringen, damit diese im OSK diskutiert werden können. Dort wo ein Konsens möglich ist, werden departmentsweite Schritte angestrebt.

Viele Präsentationen, Workshopmaterialien und Dokumente unseres OSK sind offen verfügbar (<https://osf.io/mgwk8>) und können von anderen genutzt werden.

## Lehre

Die Folien der Vorlesung zu Statistik I und II wurden von Mitgliedern des OSK kritisch geprüft und neue Schwerpunkte gesetzt. So lernen nun alle Studierenden im Rahmen der Statistikausbildung im zweiten Semester wie eine unangemessene (aber nicht untypische; John et al., 2012) Datenauswertung mit dem Ziel, signifikante Ergebnisse zu erzielen, zu einer Inflation falsch-positiver Ergebnisse führt. Dazu werden auch interaktive Tools eingesetzt, mit denen aufgezeigt wird, wie schnell man durch  $p$ -hacking ein signifikantes Ergebnis erzielen kann (Schönbrodt, 2016). Weiterhin gibt es einen längeren Exkurs über Meta-Analysen und welchen Einfluss ein Publikationsbias auf metaanalytische Effektschätzungen hat.

Für empirische Praktika wurde ein verpflichtendes Kern-Curriculum mit vier Elementen eingeführt<sup>6</sup>: (a) eine explizite Stichprobenplanung (im Regelfall: eine A-priori-Poweranalyse), (b) die Präregistrierung der Hypothesen, Operationalisierungen und geplanten konfirmatorischen Analysen, (c) offene Daten mit Codebuch, sowie (d) reproduzierbare Analyseskripte. Somit wird mit allen Studierenden mindestens einmal ganz praktisch ein reproduzierbarer und transparenter Forschungsprozess von Anfang bis zum Ende durchdekliniert.

## Zusammenarbeit mit Ethikkommission

Auch die Ethikkommission an der Fakultät 11 der LMU unterstützt in ihrer Arbeit die Forschungstransparenz. Dabei hat sie unnötig restriktive Bestimmungen gestrichen und Empfehlungen zu Probandeneinverständniserklärungen so umformuliert, dass eine Bereitstellung offener Daten rechtlich möglich ist. Sie greift die DFG-Leitlinien zum Umgang mit Forschungsdaten auf, indem sie in dem Antragsformular<sup>7</sup> explizit zur Veröffentlichung der anonymisierten Primärdaten in öffentlich zugänglichen wissenschaftliche Repositorien auffordert. Bei pseudonymisierten Daten empfiehlt die Ethikkommission nach 10 Jahren die Personen-Zuordnungslisten, aber nicht die eigentlichen Primärdaten zu löschen. Hierdurch wird aus unserer Sicht eine Balance zwischen Datenschutz und Forschungstransparenz gewährleistet, welche jedoch regelmäßig, zum Beispiel im Licht der neuen Datenschutz-Grundverordnung, neu bewertet werden muss.

## Anreize neu setzen: Qualität statt Quantität

Wenn Glaubwürdigkeit und Qualität als Kriterien guter Forschung dienen, haben transparente und reproduzierbare Forscher gute Argumente auf ihrer Seite. So schreibt zum Beispiel die DFG in Bezug auf offene Forschungsdaten: „Das Engagement und die Bemühungen von Wissenschaftlern und Wissenschaftlerinnen um die Verfügbarmachung von Forschungsdaten [...] sollten bei der Würdigung von wissenschaftlichen Qualifikationen und Leistungen zukünftig stärker berücksichtigt werden.“<sup>8</sup> Jean-Claude Burgelman, Leiter von „Research and Innovation“ der Europäischen Kommission, betont ebenfalls: „The career system has to gratify open science“<sup>9</sup>. In die-

sem Sinne versucht das Department, die Anreizstrukturen zu justieren indem positive Anreize bei Mittelverteilung und Stellenausschreibungen gesetzt werden.

*Verteilung der Mittel.* Die Verteilung der Mittel am Department bemisst sich unter anderem nach Kriterien der Forschungstransparenz, indem bei der internen Berechnung Bonuspunkte für offene Daten, offenes Material, und Präregistrierungen vergeben werden.

*Schwerpunkte bei Stellenausschreibungen.* Im Dezember 2015 wurde an der LMU die W3-Professur für Sozialpsychologie neu ausgeschrieben. Dabei wurde folgender Absatz in die Ausschreibung aufgenommen: „Das Department Psychologie legt Wert auf transparente und replizierbare Forschung und unterstützt diese Ziele durch Open Data, Open Material und Präregistrierungen. Bewerber/innen werden daher gebeten, in ihrem Anschreiben darzulegen, auf welche Art und Weise sie diese Ziele bereits verfolgt haben und in Zukunft verfolgen möchten.“ Derselbe Absatz wurde in 2016 bei drei weiteren Ausschreibungen für Professuren unseres Departments verwendet. Vergleichbare Transparenzstatements wurden auch bei Professurausschreibungen der Universitäten Köln, Göttingen und Münster hinzugefügt.

## Koordinierte Schritte hin zu einer reproduzierbaren und transparenten Forschung

Robert Merton (1942) hat vier Normen definiert, die im Kern jeder Wissenschaft stehen. Zwei davon haben direkt mit offener Wissenschaft zu tun: „Kommunitarismus“ bezeichnet die Norm, dass wissenschaftliche Erkenntnisse, Methoden und Daten ein Gemeingut sind. Es gibt auch eine Gegennorm zum Kommunitarismus: „Secrecy is the anti-thesis of this norm; full and open communication its enactment“ (S. 274).

Eine weitere Norm der Wissenschaft ist laut Merton der „organisierte Skeptizismus“. Kern wissenschaftlichen Arbeitens ist demnach, dass Behauptungen auf Basis empirischer Fakten und logischer Konsistenz überprüft werden und nachvollziehbar sind. Dies ist nur dann ganz möglich, wenn Zugang zu den methodischen Details und den Primärdaten gegeben ist. Ansonsten müsste man auf Autoritätsbeweise und guten Willen vertrauen. Letzteres

<sup>6</sup> siehe [http://www.fak11.lmu.de/dep\\_psychologie/studium/lehrelounge/kerncurriculum\\_empra](http://www.fak11.lmu.de/dep_psychologie/studium/lehrelounge/kerncurriculum_empra)

<sup>7</sup> einsehbar auf <http://www.fak11.lmu.de/forschung/ethikkommission/>

<sup>8</sup> [http://www.dfg.de/foerderung/antrag\\_gutachter\\_gremien/antragstellende/antragstellung/nachnutzung\\_forschungsdaten/index.html](http://www.dfg.de/foerderung/antrag_gutachter_gremien/antragstellende/antragstellung/nachnutzung_forschungsdaten/index.html)

<sup>9</sup> <http://www.euroscientist.com/jean-claude-burgelman-the-new-open-science-paradigm-requires-fine-tuning/>

ist zwischenmenschlich angenehm, aber wissenschaftlich nicht angemessen.

Was verlieren wir, wenn wir diese Kernelemente wissenschaftlichen Arbeitens vernachlässigen? Akademische Psychologen sind nicht die Einzigen, die Erklärungen für menschliches Verhalten und Erleben anbieten. Es gibt eine Vielzahl von selbsternannten Selbsthilfe-Gurus, Talkshow-Kommentatoren oder politischen Interessengruppen, die ebenfalls Erklärungen zur *conditio humana* im Angebot haben. Auf wessen Aussagen sollten sich politische Entscheidungen, Therapien, Personalentwicklungsmaßnahmen oder Gutachten in Gerichtsprozessen stützen? Die Normen und Methoden der Wissenschaft heben wissenschaftlich-psychologische Erklärungen über diese alternativen Alltagserklärungen hinaus, und stellen letztendlich auch eine notwendige Legitimation für eine öffentliche Finanzierung von sozialwissenschaftlicher Forschung dar (Lupia, 2014).

Die aktuellen Entwicklungen, sowohl auf technologischer Ebene, als auch auf Ebene der Publikationsstrukturen, ermöglichen nun den Übergang zu einer wesentlich höheren Stufe einer transparenten Wissenschaft. Dieser Übergang muss auch ganz konkret vor Ort in den Instituten und Arbeitsgruppen implementiert werden. Forschungstransparenz alleine ist sicher kein hinreichender Faktor für gelungene Wissenschaft, denn man kann mit maximaler Transparenz schlechte Untersuchungsdesigns anwenden, fehlerhaft argumentieren, oder äußerst triviale oder abwegige Forschungsfragen untersuchen. Erfolgreiche Wissenschaft bedarf wesentlich mehr, und die aktuelle Replikationskrise kann auch als Krise der psychologischen Theoriebildung gesehen werden (siehe auch Glöckner, Fiedler & Renkewitz, in dieser Ausgabe). Wir sind jedoch der festen Überzeugung, dass Forschungstransparenz ein notwendiger Baustein einer glaubwürdigen und replizierbaren Forschung ist, der uns näher an unseren gesellschaftlichen Auftrag und unsere eigentlichen Werte als Wissenschaftler heranbringt (Anderson, Martinson & De Vries, 2007). Viele Worte sind über die Glaubwürdigkeitskrise geschrieben worden. Nun ist es an der Zeit, zu handeln und verschiedene Stellschrauben unserer Forschungspraxis und unseres wissenschaftlichen Systems zu justieren<sup>10</sup>.

## Literatur

Abele-Brehm, A. E. & Bühner, M. (2016). Überlegungen zur Optimierung von Berufungsverfahren in der Psychologie. *Psychologische Rundschau*, 67, 262 – 268.

- 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, 3 – 14. <https://doi.org/10.1525/jer.2007.2.4.3>
- Asendorpf, J. B., Conner, M., De Fruyt, F., De Houwer, J., Denissen, J. J. A. & Fiedler, K. (2013). Recommendations for increasing replicability in psychology. *European Journal of Personality*, 27, 108 – 119. <https://doi.org/10.1002/per.1919>
- Bakker, M., Dijk, A. van & Wicherts, J. M. (2012). The rules of the game called psychological science. *Perspectives on Psychological Science*, 7, 543 – 554. <https://doi.org/10.1177/1745691612459060>
- Begley, C. G. & Ellis, L. M. (2012). Drug development: Raise standards for preclinical cancer research. *Nature*, 483, 531 – 533. <https://doi.org/10.1038/483531a>
- Brown, N. J. L. & Heathers, J. A. J. (2016). The GRIM test: A simple technique detects numerous anomalies in the reporting of results in psychology. *Social Psychological and Personality Science*, 8 (4), 363 – 369. <https://doi.org/10.1177/1948550616673876>
- Chambers, C. D., Feredoes, E., Muthukumaraswamy, S. D. & Etchells, P. (2014). Instead of „playing the game“ it is time to change the rules: Registered Reports at AIMS Neuroscience and beyond. *AIMS Neuroscience*, 1, 4 – 17.
- Chang, A. C. & Li, P. (2015). Is economics research replicable? Sixty published papers from thirteen journals say „usually not“. *Finance and Economics Discussion Series 2015 – 083*. <https://doi.org/10.17016/FEDS.2015.083>
- Coyne, J. C. (December 4, 2016). *Spokesperson clarifies that American Psychological Association does not endorse routine data sharing*. Retrieved from <https://jcoynester.wordpress.com/2016/12/04/spokesperson-clarifies-that-american-psychological-association-does-not-endorse-routine-data-sharing/>
- Engel, C. (2015). Scientific disintegrity as a public bad. *Perspectives on Psychological Science*, 10, 361 – 379. <https://doi.org/10.1177/1745691615577865>
- Eubank, N. (2016). Lessons from a decade of replications at the Quarterly Journal of Political Science. *Political Science & Politics*, 49, 273 – 276. <https://doi.org/10.1017/S1049096516000196>
- Ferguson, C. J. (November 29, 2016). *A quixotic quest to obtain a dataset on media violence with an unexpected price tag*. Retrieved from <https://jcoynester.wordpress.com/2016/11/29/a-quixotic-quest-to-obtain-a-dataset-on-media-violence-with-an-unexpected-price-tag/>
- Fiedler, K. & Schwarz, N. (2015). Questionable research practices revisited. *Social Psychological and Personality Science*, 7 (1), 45 – 52. <https://doi.org/10.1177/1948550615612150>
- Fiedler, K., Elbert, T., Erdfelder, E., Freund, A. M., Kliegl, R. & Stahl, C. (2015). *Empfehlungen der DGPs-Kommission „Qualität der psychologischen Forschung“*. Verfügbar unter [https://www.dgps.de/index.php?id=143&tx\\_ttnews\[tt\\_news\]=1669&cHash=ff3ae479fbbfa2ca45aebf7b4db8e504](https://www.dgps.de/index.php?id=143&tx_ttnews[tt_news]=1669&cHash=ff3ae479fbbfa2ca45aebf7b4db8e504)
- Gelman, A. & Loken, E. (2014). The statistical crisis in science. *American Scientist*, 102, 460. <https://doi.org/10.1511/2014.111.460>
- Gervais, W. M., Jewell, J. A., Najle, M. B. & Ng, B. K. L. (2015). A powerful nudge? Presenting calculable consequences of underpowered research shifts incentives toward adequately powered designs. *Social Psychological and Personality Science*, 6 (7), 847 – 854. <https://doi.org/10.1177/1948550615584199>
- Giner-Sorolla, R. (2012). Science or art? How aesthetic standards grease the way through the publication bottleneck but undermine science. *Perspectives on Psychological Science*, 7, 562 – 571.

<sup>10</sup> Beziehungsweise lange bekannte und angemahnte Qualitätsstandards tatsächlich umzusetzen.

- Glöckner, A., Fiedler, S. & Renkewitz, F. (2016). *Belastbare und effiziente Wissenschaft: Strategische Ausrichtung von Forschungsprozessen als Weg aus der Replikationskrise*. Hagen: FernUniversität Hagen.
- Higginson, A. D. & Munafò, M. R. (2016). Current incentives for scientists lead to underpowered studies with erroneous conclusions. *PLOS Biology*, 14, e2000995. <https://doi.org/10.1371/journal.pbio.2000995>
- Ioannidis, J. P. A. (2014). Handling the fragile vase of scientific practices. *Addiction*, 110, 9–10. <http://doi.org/10.1111/add.12720>
- John, L. K., Loewenstein, G. & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, 23, 524–532. <https://doi.org/10.1177/0956797611430953>
- Klein, R. A., Ratliff, K. A., Vianello, M., Adams Jr., R. B., Bahník, Š. & Bernstein, M. J. (2014). Investigating variation in replicability: A „many labs“ replication project. *Social Psychology*, 45, 142–152. <https://doi.org/10.1027/1864-9335/a000178>
- Lupia, A. (2014). What is the value of social science? Challenges for researchers and government funders. *Political Science & Politics*, 47, 1–7. <https://doi.org/10.1017/S1049096513001613>
- Merton, R. K. (1942). A note on science and democracy. *Journal of Legal and Political Sociology*, 1, 115–126.
- Morey, R. D., Chambers, C. D., Etchells, P. J., Harris, C. R., Hoeks-tra, R. & Lakens, D. (2016). The Peer Reviewers' Openness Initiative: Incentivizing open research practices through peer review. *Royal Society Open Science*, Published online January 13, 2016. <https://doi.org/10.1098/rsos.150547>
- Nelson, L. D., Simmons, J. P. & Simonsohn, U. (2012). Let's publish fewer papers. *Psychological Inquiry*, 23, 291–293. <https://doi.org/10.1080/1047840X.2012.705245>
- Nosek, B. A. & Lakens, D. (2014). Registered reports: A method to increase the credibility of published results. *Social Psychology*, 45, 137–141. <https://doi.org/10.1027/1864-9335/a000192>
- Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D. & Breckler, S. J. (2015). Promoting an open research culture. *Science*, 348, 1422–1425. <https://doi.org/10.1126/science.aab2374>
- Nosek, B. A., Spies, J. R. & Motyl, M. (2012). Scientific utopia II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science*, 7, 615–631. <https://doi.org/10.1177/1745691612459058>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349, aac4716. <https://doi.org/10.1126/science.aac4716>
- 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, 712–712. <https://doi.org/10.1038/nrd3439-c1>
- Schönbrodt, F. (June 21, 2016). *Introducing the p-hacker app: Train your expert p-hacking skills*. Retrieved from <http://www.nicebread.de/introducing-p-hacker/>
- Schönbrodt, F. D., Gollwitzer, M. & Abele-Brehm, A. (2016). *Entwurf für DGPs-Richtlinien zum Umgang mit Forschungsdaten*. Verfügbar unter [https://www.dgps.de/uploads/media/DGPs-Richtlinien\\_Forschungsdaten\\_22\\_01\\_16.pdf](https://www.dgps.de/uploads/media/DGPs-Richtlinien_Forschungsdaten_22_01_16.pdf)
- Schönbrodt, F. D., Maier, M., Heene, M. & Zehetleitner, M. (2015). *Commitment to research transparency*. Retrieved from <http://www.researchtransparency.org>
- Simmons, J. P., Nelson, L. D. & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22, 1359–1366.
- Simonsohn, U., Simmons, J. P. & Nelson, L. D. (2015). *Specification curve: Descriptive and inferential statistics on all reasonable specifications*. Retrieved from <https://papers.ssrn.com/abstract=2694998>
- Smaldino, P. E. & McElreath, R. (2016). *The natural selection of bad science*. Retrieved from <http://arxiv.org/abs/1605.09511>
- Steen, S., Tuerlinckx, F., Gelman, A. & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, 11, 702–712. <https://doi.org/10.1177/1745691616658637>
- Tinbergen, N. (1963). On aims and methods of ethology. *Zeitschrift für Tierpsychologie*, 20, 410–433.
- Vanpaemel, W., Vermorgen, M., Deriemaeker, L. & Storms, G. (2015). Are we wasting a good crisis? The availability of psychological research data after the storm. *Collabra*, 1, 1–5. <https://doi.org/10.1525/collabra.13>
- Wagenmakers, E.-J., Wetzels, R., Borsboom, D., Maas, H. L. J. van der & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science*, 7, 632–638. <https://doi.org/10.1177/1745691612463078>
- Wicherts, J. M., Bakker, M. & Molenaar, D. (2011). Willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results. *PLoS ONE*, 6, e26828. <https://doi.org/10.1371/journal.pone.0026828>
- Wicherts, J. M., Borsboom, D., Kats, J. & Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. *American Psychologist*, 61, 726–728. <https://doi.org/10.1037/0003-066X.61.7.726>

**PD Dr. Felix Schönbrodt**

Ludwig-Maximilians-Universität  
 Department Psychologie  
 Leopoldstr. 13  
 80802 München  
 felix@nicebread.de