

Akzeptierte Manuskriptfassung (nach peer review) des folgenden Artikels:

[Geukes, K., Schönbrodt, F. D., Utesch, T., Geukes, S. & Back, M. D. \(2016\). Wege aus der Vertrauenskrise. Individuelle Schritte hin zu verlässlicher und offener Forschung. Zeitschrift für Sportpsychologie, 23 \(3\). Doi:10.1026/1612-5010/a000167](#)

© Hogrefe Verlag, Göttingen 2016

Diese Artikelfassung entspricht nicht vollständig dem in der Zeitschrift veröffentlichten Artikel. Dies ist nicht die Originalversion des Artikels und kann daher nicht zur Zitierung herangezogen werden.

Die akzeptierte Manuskriptfassung unterliegt der Creative Commons License CC-BY-NC.

Zeitschrift für Sportpsychologie

Wege aus der Vertrauenskrise - Individuelle Schritte hin zu verllässlicher und offener Forschung Ways out of the crisis of confidence - Individual steps towards a reliable and open science --Manuskript-Entwurf--

| | |
|--|--|
| Manuskriptnummer: | SPO-D-16-00006R1 |
| Vollständiger Titel: | Wege aus der Vertrauenskrise - Individuelle Schritte hin zu verllässlicher und offener Forschung Ways out of the crisis of confidence - Individual steps towards a reliable and open science |
| Artikeltyp: | Originalarbeit |
| Schlüsselwörter: | Vertrauenskrise; Replikation; Präregistrierung; Open Science; Forschungspraktiken; crisis of confidence; replication; preregistration; research practices |
| Korrespond. Autor: | Katharina Geukes, Dr. Institut für Psychologie, Universität Münster Münster, GERMANY |
| Korrespondierender Autor, Zweitinformationen: | |
| Korrespondierender Autor, Institution: | Institut für Psychologie, Universität Münster |
| Korrespondierender Autor, zweite Institution: | |
| Erstautor: | Katharina Geukes, Dr. |
| Erstautor, Zweitinformationen: | |
| Reihenfolge der Autoren: | Katharina Geukes, Dr. Felix D. Schönbrodt Till Utesch Sebastian Geukes Mitja D. Back |
| Reihenfolge 'Zweite Informationen' von Autoren: | |
| Zusammenfassung: | <p>Die Psychologie, und mit ihr auch die Sportpsychologie, befindet sich in einer sogenannten Vertrauenskrise (siehe Titel des Themenhefts). Während die Diskussion ihrer Ursachen und Konsequenzen lebendig geführt wird, fehlt es an der Herausarbeitung individueller Möglichkeiten, Vertrauen wieder kollektiv herzustellen. In diesem Beitrag zeigen wir Wege aus dieser Vertrauenskrise auf und beschreiben insbesondere individuelle Schritte hin zu verllässlichem und offenem Forschen: Verllässliches Forschen bezieht sich auf das Publizieren von robusten Ergebnissen, sowie von direkten und konzeptuellen Replikationsstudien und offenes Forschen auf die Transparenz hinsichtlich der Planung (Präregistrierung), der Durchführung (Open Material) und der Analyse (Open Data, Reproducible Code) wissenschaftlicher Untersuchungen. Das Bekenntnis zu verllässlicher und offener Forschung wird unser Handeln in den verschiedenen Rollen im Wissenschaftssystem (als Forscher, Reviewer, Herausgeber, Betreuer, Kommissionsmitglied etc.) verändern. In diesem Sinne begreifen wir in diesem Beitrag die momentane Diskussion als Chance, die Zuverlässigkeit unserer Befunde nachhaltig zu steigern und langfristig gerechtfertigtes Vertrauen zu schaffen.</p> <p>Psychology faces a so-called crisis of confidence and so does sport psychology (see title of this special issue). While the debate on its causes and consequences is lively, the deduction of individual opportunities to collectively increase trust is missing. We propose ways out of this crisis and above all describe individual steps towards a reliable and open science. Reliable science refers to the publication of robust effects, as well as to direct and conceptual replications and open science to the transparency</p> |

regarding the design (Preregistration), the conduction (Open Material), and analysis (Open Data, Reproducible Code) of scientific studies. The commitment to reliable and open science will change our behavior in the diverse roles within the scientific system (e.g., as researcher, reviewers, supervisors, editors, members of commissions). In this sense, we consider the current discussion as a chance to enhance the trustworthiness of our findings and to ultimately create justified confidence.

Running Head: WEGE AUS DER VERTRAUENKRISE

**Wege aus der Vertrauenskrise –
Individuelle Schritte hin zu verlässlicher und offener Forschung**

**Ways out of the crisis of confidence –
Individual steps towards a reliable and open science**

Deutsches Abstract

Die Psychologie, und mit ihr auch die Sportpsychologie, befindet sich in einer sogenannten Vertrauenskrise (siehe Titel des Themenhefts). Während die Diskussion ihrer Ursachen und Konsequenzen lebendig geführt wird, fehlt es an der Herausarbeitung individueller Möglichkeiten, Vertrauen wieder kollektiv herzustellen. In diesem Beitrag zeigen wir Wege aus dieser Vertrauenskrise auf und beschreiben insbesondere *individuelle* Schritte hin zu verlässlichem und offenem Forschen: *Verlässliches* Forschen bezieht sich auf das Publizieren von robusten Ergebnissen, sowie von direkten und konzeptuellen Replikationsstudien und *offenes* Forschen auf die Transparenz hinsichtlich der Planung (Präregistrierung), der Durchführung (Open Material) und der Analyse (Open Data, Reproducible Code) wissenschaftlicher Untersuchungen. Das Bekenntnis zu verlässlicher und offener Forschung wird unser Handeln in den verschiedenen Rollen im Wissenschaftssystem (als Forscher, Reviewer, Herausgeber, Betreuer, Kommissionsmitglied etc.) verändern. In diesem Sinne begreifen wir in diesem Beitrag die momentane Diskussion als Chance, die Zuverlässigkeit unserer Befunde nachhaltig zu steigern und langfristig gerechtfertigtes Vertrauen zu schaffen.

Anzahl Wörter: 149

Schlüsselwörter: Vertrauenskrise; Replikation; Präregistrierung; Open Science; Forschungspraktiken

Englisches Abstract

Psychology faces a so-called crisis of confidence and so does sport psychology (see title of this special issue). While the debate on its causes and consequences is lively, the deduction of individual opportunities to collectively increase trust is missing. We propose ways out of this crisis and above all describe *individual* steps towards a reliable and open science. *Reliable* science refers to the publication of robust effects, as well as to direct and conceptual replications and *open* science to the transparency regarding the design (Preregistration), the conduction (Open Material), and analysis (Open Data, Reproducible Code) of scientific studies. The commitment to reliable and open science will change our behavior in the diverse roles within the scientific system (e.g., as researcher, reviewers, supervisors, editors, members of commissions). In this sense, we consider the current discussion as a chance to enhance the trustworthiness of our findings and to ultimately create justified confidence.

Word count: 150

Key words: crisis of confidence; replication; preregistration; open science; research practices

Wege aus der Vertrauenskrise –

Individuelle Schritte hin zu verlässlicher und offener Forschung

Die Vertrauenswürdigkeit einer Wissenschaft steht und fällt mit der Replizierbarkeit ihrer Befunde. Gemäß Asendorpf et al. (2013) differenzieren wir zwischen der Reproduzierbarkeit, der Generalisierbarkeit, und der Replizierbarkeit von Forschungsbefunden. Reproduzierbarkeit bedeutet, dass verschiedene Forscher bei der Analyse des gleichen Datensatzes zu denselben Ergebnissen kommen. Generalisierbarkeit bedeutet in unterschiedlichen Kontexten zu den gleichen Ergebnissen zu kommen. Unter Replizierbarkeit verstehen wir schließlich, dass die Befunde einer Originalstudie A und die einer Replikationsstudie B nicht substantiell verschieden sind und, falls es Unterschiede zwischen diesen Studien gibt, diese auf unsystematischen Fehlern beruhen. Das heißt die Replizierbarkeit beschreibt das Ausmaß, zu dem konsistente Ergebnisse beobachtet werden, wenn wissenschaftliche Untersuchungen wiederholt werden. Da die Replizierbarkeit eine der defintorischen Eigenschaften von Wissenschaft ist (Bacon, 1267/1859; Jasny, Chin, Chong, & Vignieri, 2011; Kuhn, 1962; Popper, 1934/1992; Rosenthal, 1991) und sogar als Demarkationslinie zwischen Wissenschaft und Nicht-Wissenschaft bezeichnet wird, (Braude, 1979), wurde schon seit längerem in Frage gestellt, ob die gängige Wissenschaftspraxis tatsächlich zu in der Regel replizierbaren Befunden führt¹ (Cohen, 1994, Ionannidis, 2005; Makel, Plucker, & Hegarty, 2012; Maxwell, 2004; Meehl, 1990). In jüngerer Zeit gab es insbesondere im Fach Psychologie (Ebersole, Atherton, Belanger, Skulborstad, Adams, Allen et al., 2014; Klein et al., 2014; Klein et al., 2016; Open Science Collaboration, 2015), aber

¹ Selbstverständlich sind dennoch Studien denkbar, die sich nicht (mehr) replizieren lassen, obwohl die Originalstudie einen „wahren“ Befund berichtete, beispielsweise weil es keine weitere Beobachtung Gelegenheit gibt oder weil der wahre Effekt nach gesellschaftlichen Veränderungen nicht mehr existiert (für eine ausführlichere Darstellung von Replizierbarkeit und ihrer Quantifizierung siehe Asendorpf et al., 2013 oder auch Valentine et al., 2011 oder aber Cumming und Finch, 2005).

mittlerweile auch in anderen Disziplinen wie der Krebsforschung (Errington et al., 2014) oder den Wirtschaftswissenschaften (Camerer, et al., 2016, Chang & Li, 2015), Bestrebungen, die Frage der Replizierbarkeit veröffentlichter Befunde in systematischer Weise zu untersuchen. Insbesondere das sogenannte Reproducibility Project: Psychology² (Open Science Collaboration, 2015) hat hier wichtige Aufschlüsse geliefert und zu einem großen Echo sowohl in der wissenschaftlichen als auch in der breiten Öffentlichkeit geführt. Ziel dieses groß angelegten Vorhabens war es, die tatsächliche Replizierbarkeit psychologischer Studien abzuschätzen. Dazu wurden 100 experimentelle und korrelative Studien aus drei wichtigen psychologischen Zeitschriften ausgewählt und möglichst exakt wiederholt. Während 97 % der Originalstudien signifikante Effekte berichteten, war dies nur noch in 36 % der Replikationsstudien der Fall— neben der einfachen Gegenüberstellung der Signifikanztests wurden in der Publikation auch viele weitere Indizes der Replizierbarkeit untersucht. Diese systematisch ermittelten Zahlen legen nahe, dass die veröffentlichten Befunde keinesfalls in der Regel verlässlich und replizierbar sind (s. auch Schimmack & Heene, 2016). Diese empirischen Befunde gesellen sich zu vielen weiteren Indizien, die den Eindruck einer Vertrauenskrise unterstreichen. Dazu zählen beispielsweise Betrugs- und/oder Manipulationsfälle (vgl. Stroebe, Postmes, & Spears, 2012), die hochrangige Veröffentlichung von Befunden zu angeblich zeitlich rückwirkenden Effekten (Bem, 2011) und ihre öffentliche Diskussion (Galak, LeBoeuf, Nelson, & Simmons, 2012; Wagenmakers, Wetzels, Borsboom, & van der Maas, 2011), Befunde, die zeigen, dass Daten trotz schriftlicher Zusicherung in der überwiegenden Mehrheit nicht für Reanalysen zur Verfügung gestellt werden (Vanpaemel, Vermorgen, Deriemaecker, & Storms, 2015; Wicherts, Bakker, & Molenaar, 2011; Wicherts, Borsboom, Kats, & Molenaar, 2006), und Analysen, die

² Nach der oben genannten Definition hätte das Projekt eher „Replicability Project“ genannt werden sollen.

zeigen, wie unter den aktuellen Veröffentlichungsbedingungen und Anreizsystemen sehr leicht falsch positive Befunde ermittelt werden können (Ioannidis, 2005; Simmons, Nelson, & Simonsohn, 2011).

Es ist anzunehmen, dass die mangelnde Replizierbarkeit und damit der Kern der Vertrauenskrise in der Psychologie in ähnlichem Maße auch auf die Sportpsychologie zutrifft. Schweizer und Furley (2015) beschreiben hierzu bereits sehr anschaulich die Problematik der Stichprobengröße und der angemessenen Teststärke (Power). Es gab zuvor auch schon vereinzelte spezifische Ansätze zur Verbesserung der Methoden (z. B. zu Faktorenanalysen, zu p -Werten und zur Reliabilität; Myers, Ahn, & Jin, 2011; Zhu, 2012, 2013). Die globale Diskussion der durch die Vertrauenskrise nun vermehrt in den Fokus gerückten Kritikpunkte an der psychologischen Forschung findet in der Sportpsychologie explizit bislang kaum statt. Da die Sportpsychologie—ob experimentell oder korrelativ forschend—aber (1) konzeptionell auf mehr oder weniger vertrauenswürdige Befunde der psychologischen Grundlagenfächer rekurriert und sich (2) paradigmatisch wie methodisch an den in der Psychologie gängigen Forschungspraktiken orientiert, ist die Vertrauenskrise gleichermaßen präsent und diskussionswürdig. Dies deutet auf die einmalige Chance der Sportpsychologie hin, die lebhafteste Diskussion der Vertrauens- und Replikationskrise in der Psychologie nun in eine eigene konstruktive Systemkritik zu überführen und ihr gemeinsam zu einem vergleichsweise frühen Zeitpunkt zu begegnen.

Ursachen der Vertrauenskrise

Wie kommt es überhaupt zu einer Vertrauenskrise, zur Veröffentlichung von Befunden, die häufig nicht replizierbar sind? Es ist anzunehmen, dass das Bestreben von Wissenschaftlern grundsätzlich gegensätzlicher Natur ist. Es geht Ihnen darum, sicheres neues Wissen zu generieren (Dreiskämper, dieses Heft). Jedoch wurde bereits von vielen Autoren festgestellt, dass es insbesondere falsche Anreizstrukturen und Belohnungssysteme sind, die Wissenschaftler zur

Veröffentlichung wenig robuster Forschungsergebnisse verleiten (Engel, 2015; Nelson, Simmons & Simonsohn, 2012; Nosek, & Lakens, 2014; Nosek, Spies & Motyl, 2012; Pashler & Wagenmakers, 2012; Smaldino, & McElreath, 2016).

Wissenschaftliche Karrieren sind heutzutage davon abhängig, dass Forscher in kurzer Zeit eine große Anzahl an Veröffentlichungen ansammeln, möglichst in hochrangigen Zeitschriften. Diese Anforderungen laden dazu ein, eine Vielzahl von Studien mit kleinen Stichproben durchzuführen (Bakker, van Dijk & Wicherts, 2012) und—ob bewusst oder unbewusst—fragwürdige Entscheidungen bei der Datenerhebung und -analyse zu treffen (Questionable Research Practices; QRPs; De Vries, Anderson, & Martinson, 2006; Gelman & Loken, 2013; John, Loewenstein, & Prelec, 2012; siehe aber auch: Fiedler & Schwarz, 2015). Dazu zählen beispielsweise: (1) Das optionale Stoppen der Datenerhebung, sobald signifikante Ergebnisse gefunden werden (optional stopping, Szucs, 2016). (2) Das selektive Berichten nur derjenigen Probanden, Variablen oder Experimente, die signifikante Effekte zeigen bzw. das selektive Nicht-Berichten von Probanden, Variablen oder Experimenten, die nichtsignifikante Effekte zeigen (file drawer; Spellman, 2012). (3) Die erst post-hoc getroffenen Entscheidungen für die Hinzunahme von Kovariaten oder Subgruppen in die Analyse, ohne die ein Effekt nicht bestehen würde (Simmons, Nelson & Simonsohn, 2011). (4) Das nachträgliche Anpassen der Hypothesen an die Daten bzw. Ergebnisse („Hypothesizing after the results are known“, HARKing, Kerr, 1998) mit der Folge, dass explorative Forschung als konfirmatorische Forschung deklariert wird. Damit geht häufig die unzulängliche Beachtung dessen einher, dass die Wahrscheinlichkeit der Daten gegeben die Hypothese nicht dieselbe Wahrscheinlichkeit besitzt wie die Hypothese gegeben die Daten (Wagenmakers, Wetzels, Borsboom & van der Maas, 2011).

Es ergibt sich die kuriose Situation, dass viele der resultierenden falsch positiven Befunde oft kontraintuitiv sind und sich somit umso besser eignen, hochrangig veröffentlicht zu werden,

da sie eben scheinbar einen zuvor unerwarteten Zusammenhang aufzeigen. In jedem Fall aber führt diese Jagd nach statistischer Signifikanz (vgl. Büsch & Strauß, 2015) in der Mehrheit offenbar nicht zu vertrauenswürdigen und replizierbaren Ergebnissen. Grundlage der Vertrauenskrise ist also das Dilemma, dass im aktuellen Publikations- und Belohnungssystem Quantität und überraschende Befunde gegenüber Qualität und replizierbaren Befunden bevorzugt werden.

Dass diese Vertrauenskrise Veränderungen nach sich ziehen wird, ist aus unserer Sicht so nötig wie erwünscht. *Wie* diese Veränderungen allerdings aussehen können, ist Zentrum einer aktuellen Debatte, die bislang noch keinen breiten Konsens erzielt hat. Neben vielen Fachartikeln und Präsentationen auf Konferenzen, die sich mit vertrauenswürdiger und replizierbarer Forschung beschäftigen, findet dieser Diskurs zu einem guten Teil informell statt, beispielsweise in Blogs, auf Twitter, oder in Facebook-Diskussionsgruppen³. Im Folgenden stellen wir zwei sich ergänzende Wege aus dieser Vertrauenskrise zur Diskussion vor—verlässliches und offenes Forschen—und beschreiben im Anschluss, wie jeder von uns diese Wege in Form von *individuellen* Schritten gehen kann, d.h. wie wir verlässliches und offenes Forschen in unser Handeln integrieren können.

Verlässliches Forschen

A science of progress is reliable science (Feeg, 2011).

Wenn der Fortschritt einer wissenschaftlichen Disziplin davon abhängt, wie verlässlich in ihr geforscht wird, ist die Schlussfolgerung naheliegend: Wir müssen verlässlich forschen! Aber wie? Unter verlässlichem Forschen verstehen wir ein Forschen, das dem übergeordneten Ziel,

³ siehe z.B. die Psychological Methods Discussion Group auf Facebook (<https://www.facebook.com/groups/853552931365745/>) oder diese Liste von Diskutanten auf Twitter (<https://twitter.com/nicebread303/lists/openscience>)

replizierbare Befunde zu generieren, d.h. eine realistische Schätzung von wahren Effekten zu erzielen, gerecht wird.

Für jede einzelne Studie bedeutet dies, dass sie jeweils so geplant sein sollte, dass die Wahrscheinlichkeit für falsch-positive oder falsch-negative Befunde minimiert wird. Insbesondere sollte die minimal notwendige Stichprobengröße auf einer realistischen Teststärken/Power-Berechnung basieren, was in der (Sport-)Psychologie in der Regel eine deutliche Erhöhung der bisherigen Stichprobengrößen bedeutet (vgl. Schweizer & Furley, 2015; siehe auch: Bakker, Hartgerink, Wicherts, & van der Maas, 2016)—und selbst wenn dies mit einem größeren Rekrutierungs- und Durchführungsaufwand für jede einzelne Untersuchung einhergeht.

Gleichzeitig bedeutet das übergeordnete Ziel der Replizierbarkeit auch, dass wir Replikationsversuchen anderer Forscher wohlwollend gegenüberstehen sollten. Hier ergänzen sich direkte Replikationen (Durchführung mit möglichst identischen Materialien und Prozeduren) und konzeptuelle Replikationen (Variation im Material, in Operationalisierungen, in der Stichprobe, etc.), um die Replizierbarkeit von Effekten sowie ihre Generalisierbarkeit über Labor-, Stichproben-, Länder- und Kulturgrenzen hinweg abzuschätzen. Aufgrund der vielen Negativbefunde wurden manche Replikationsstudien im Grundsatz kritisiert. Hier sollte jedoch bedacht werden, dass, im Gegensatz zu axiomatischen Beweisen (z. B. in der Mathematik), ein einzelner negativer Befund keineswegs sofort die Existenz eines Effektes in Frage stellt. Eine Replikation bietet zunächst einmal lediglich einen neuerlichen empirischen Test eines angenommenen Effekts und vergrößert somit die Datenlage für seine Bewertung.

Schließlich umfasst verlässliches Forschen auch die metaanalytische Integration über Einzelstudien (also Replikationen) hinweg, da sie die Möglichkeit bietet, eine realistische Schätzung des wahren Effekts zu erzielen. Dies wird in der Psychologie bereits regelmäßig

praktiziert, jedoch kommt die wichtige Rolle verlässlicher Forschung hier erneut zum Tragen: Der metaanalytisch ermittelte Effekt kann nur diejenigen Beobachtungen abbilden, die in Einzelstudien berichtet wurden. Enthält die Literatur aber viele falsch positive Befunde bzw. fehlen tatsächliche Negativbefunde (publication bias), so kommt es zu einer Überschätzung des wahren Effektes. Trotz möglicher Verfahren zur Prüfung des „publication bias“ (z. B. Eggers test: Egger, Smith, Schneider & Minder, 1997; *p*-curve: Simonsohn, Nelson & Simmons, 2014; *p*-uniform: van Assen, Van Aert, & Wicherts, 2015; PET-PEESE: Stanley & Doucouliagos, 2013), können Metaanalysen daher stets nur so vertrauenswürdig sein, wie ihre zugrundeliegende Literatur ist ("garbage in, garbage out"). Aktuelle Untersuchungen zur Performanz dieser Verfahren zur Bias-Korrektur zeigen jedoch, dass sie nur unter sehr eingeschränkten Bedingungen gut funktionieren (van Aert, Wicherts & van Assen, in press). Gerade deshalb ist es wichtig, dass jeder einzelne Effekt bereits eine verlässliche Grundlage hat.

Unter verlässlichem Forschen verstehen wir letztlich eine Rückbesinnung auf hinlänglich bekannte wissenschaftliche und methodische Standards und damit eine Selbstverpflichtung der Wissenschaftler. Ein ergänzender Weg aus der Vertrauenskrise ist das offene Forschen, das ganz entscheidend zur Erhöhung der Vertrauenswürdigkeit wissenschaftlicher Forschung beiträgt.

Offenes Forschen

Openness is not needed because we are untrustworthy; it is needed because we are human (Nosek, Spies, & Motyl, 2012, p. 626).

Unter offenem Forschen verstehen wir ein Forschen, bei dem vollständige Transparenz bezüglich der Studienplanung und Durchführung (Open Material) sowie der Datenanalyse (Open Data und Reproducible Code) besteht. Eine hervorragende Plattform hierfür bietet beispielsweise das Open Science Framework (www.osf.io). Der Anspruch nach Offenlegung des gesamten Forschungs- und Analyseprozesses sollte hierbei nicht als ein grundsätzliches Misstrauen

gegenüber Forschern verstanden werden, sondern stellt zunächst einmal grundsätzlich die Überprüfbarkeit veröffentlichter Befunde her (vgl. die wissenschaftliche Norm des "organisierten Skeptizismus", Merton, 1942).

Es ist zutiefst menschlich, dass wir Fehler machen. Dass solche Fehler (z. B. Fehler in der Datenaufbereitung oder -analyse) aufgespürt werden können und somit entsprechende Korrekturen möglich sind, ist sicherlich eine der wichtigsten positiven Konsequenzen des offenen Forschens. Unter dem klassischen (Blackbox-)Modell ist eine solche Korrektur kaum möglich, da in der Regel nur die Endergebnisse der Analysen im Artikel präsentiert werden. Deshalb können viele Fehler gar nicht auffallen oder zumindest nicht vollends nachvollzogen werden und bestehen unkorrigiert in der wissenschaftlichen Literatur fort. Allein das Wissen um die Veröffentlichung der Materialien und Daten regt uns dazu an, Studien gewissenhafter und verständlicher zu dokumentieren. So zeigt eine Analyse von Wicherts et al. (2011), dass Publikationen, bei denen Daten bereitgestellt wurden, weniger Fehler enthielten. Das bedeutet, dass sie dadurch nicht nur das Potential für mehr Glaubwürdigkeit haben, da die berichteten Ergebnisse prinzipiell nachprüfbar sind, sondern tatsächlich eine höhere Qualität besitzen. Auch Re-Analysen (z. B. mit neueren statistischen Verfahren oder in Metaanalysen) und Replikationen werden durch die Veröffentlichung aller Untersuchungsdetails und der Rohdaten wesentlich vereinfacht. Offenes Forschen bietet schließlich eine ideale Basis für forschungsgruppenübergreifende Kollaborationen, da die wichtigsten Materialien, Daten und Analysemethoden direkt zugänglich und in eigene Abläufe integrierbar sind. Ebenso wird auch das Zusammenfügen (Poolen) von Datensätzen über Arbeitsgruppen hinweg erleichtert, um die hohen Anforderungen an Stichprobengrößen insbesondere auch in sehr aufwändigen (z.B. zeit- und kostenintensiven, längsschnittlichen und verhaltensbasierten) Studien zu erfüllen.

Ein weiteres Element offener Forschung stellt die Präregistrierung von Studien dar. Die Präregistrierung meint hier die schriftlich dokumentierte Festlegung der Fragestellungen und Hypothesen und der Erhebungs- und Analysewege bereits vor Untersuchungsbeginn. Typischerweise werden Präregistrierungen, insbesondere bei experimentellen Designs, vor dem Start der Datenerhebung vorgenommen. Aber auch bei der Analyse von existierenden Datensätzen ist eine Präregistrierung möglich. In diesen Fällen wird die Fragestellung inkl. spezifischer Hypothesen, der zu untersuchenden Variablen und dem analytischen Vorgehen präregistriert und in diesem Zug auf die öffentliche Dokumentation der bereits erhobenen Daten verwiesen.

Zunehmend sind Präregistrierungen in Form der sogenannten *registered reports* auch direkt bei den Fachzeitschriften möglich. Dort findet dann der peer-review Prozess bereits vor der Datenerhebung auf Basis des vorgeschlagenen Untersuchungsdesigns statt, ähnlich wie es bereits bei vielen klinischen Studien üblich ist (vgl. Chambers, 2015; Chambers, Dienes, McIntosh, Rotshtein, & Willmes, 2015). Nach Diskussion, Überarbeitung und Annahme des Untersuchungsplanes gibt der Herausgeber dann eine "in-principle acceptance" für die Veröffentlichung der späteren Ergebnisse, unabhängig von den resultierenden Ergebnissen und deren statistischer Signifikanz. Innerhalb der Präregistrierung wird offenes Forschen bereits vor der Durchführung einer Untersuchung realisiert und kann als Königsweg, insbesondere von konfirmatorischer Forschung, betrachtet werden. In diesem Zuge ist es wichtig zu betonen, dass die Präregistrierung in keiner Weise eine weitere explorative Analyse der Daten in einem gesonderten Analyseabschnitt ausschließt. Es wird lediglich die (z.B. für das Signifikanztesten entscheidende) Anforderung betont, zwischen konfirmatorischen und explorativen Analysen klar zu unterscheiden. Auch ohne eine Präregistrierung stellen aber die Offenlegung von

Untersuchungsmaterialien, Daten und Analysecode mit der Veröffentlichung bereits wesentliche Verbesserungen im Sinne des offenen Forschens dar.

Individuelle Schritte zur Förderung von verlässlichem und offenem Forschen

Mit der Überzeugung, dass verlässliches und offenes Forschen zu mehr Vertrauenswürdigkeit und Replizierbarkeit führen werden, stellt sich nun die Frage, wie wir diese Aspekte in unser forschendes Handeln integrieren können. Wir präsentieren nun eine Zusammenfassung von bereits vorgeschlagenen *individuellen* Schritten (s. Tabelle 1), die für jeden Einzelnen von uns umsetzbar sind. Eine wichtige Grundlage dieser Zusammenfassung ist die von Schönbrodt, Maier, Heene und Zehetleitner (2015) formulierte Selbstverpflichtung zu Offenheit und Transparenz in der Wissenschaft („Commitment to Research Transparency and Open Science“, Schönbrodt, Maier, Heene & Zehetleitner, 2015; <http://www.researchtransparency.org>). Wie auch die Selbstverpflichtung gliedert sich die Präsentation der Handlungsoptionen anhand der Rollen, die Wissenschaftler im Wissenschaftssystem innehaben—beginnend bei der Rolle als Forscher, über die des Gutachters von Manuskripten, des Betreuers von Abschluss- und Doktorarbeiten, bis hin zur Rolle des Herausgebers und die des Kommissionsmitglieds.

Als Forscher

Als individueller Forscher können wir uns zunächst grundsätzlich zu den Werten verlässlichen und offenen Forschens und dem übergeordneten Ziel der Replizierbarkeit (öffentlich) bekennen. Dies gelänge beispielsweise durch das Unterzeichnen einer Selbstverpflichtung oder durch Mitgliedschaften in entsprechend ausgerichteten Interessengruppen (beispielsweise in der Society for Improving Psychological Science, SIPS). Neben dem Bekenntnis zu offener Forschung ist aber natürlich die Umsetzung derselben entscheidend. Die Planung, Durchführung und Auswertung von Einzelstudien beispielsweise

beinhaltet schon eine Reihe wichtiger Entscheidungen, die direkt beeinflussen, inwiefern Studien grundsätzlich replizierbare Befunde hervorbringen können. Das bewusste Treffen dieser Entscheidungen kann zum Beispiel über die Integration der „21-word solution“ von Simmons, Nelson und Simonsohn (2012) in unseren Veröffentlichungen dokumentiert werden. Sie lautet „We report how we determined our sample size⁴, all data exclusions (if any), all manipulations, and all measures in the study.“ und kann standardmäßig als Teil der *Author Note* kommuniziert werden.

Weiterhin können Materialien, Daten und Analyseskripte standardmäßig zusammen mit einer Publikation der Öffentlichkeit zugänglich gemacht werden. Dies bedeutet: (1) die Vorbereitung einer nachvollziehbaren Beschreibung des Studiendesigns und -ablaufs, inklusive aller Instruktionen, Gerätschaften, Manipulationen und Items/Stimuli, die eine direkte Replikation der Studie ermöglicht, (2) die Vorbereitung und Erklärung (z. B. via Codebook) eines finalen Datensatzes, der Basis der Analysen für eine Veröffentlichung war, und (3) die Vorbereitung einer bestenfalls verständlich kommentierten Analysesyntax, die gemeinsam mit dem Datensatz genau diejenigen Ergebnisse produziert, die in der Veröffentlichung zu finden sind.

Idealerweise aber beginnt der Dokumentationsprozess bereits in der Planungsphase mit der Ausarbeitung einer Präregistrierung (die in der Regel recht kurz sein kann und einem oft ohnehin erforderlichen Ethikantrag nicht unähnlich ist). Hier wird also die Studienidee samt

⁴ Insbesondere für die Bestimmung der Stichprobengröße haben die Befunde zur Replikationsüberprüfung und Effektgrößenschätzung wichtige Implikationen (s. Open Science Collaboration, 2015; oder auch Franco, Malhotra, & Simonovits, 2015). Demnach findet sich in Replikationsstudien typischerweise ein ca. halb so großer Effekt wie in der Originalstudie und ein reines Verlassen auf die veröffentlichten Befunde würde die Effektgröße überschätzen. Als Daumenregel bietet es sich also an, einen halb so großen Effekt wie in der Originalstudie anzunehmen bzw. sich an den Effektgrößen zu orientieren, die in bereits durchgeführten Replikationsstudien gefunden wurden.

Hypothesen offengelegt sowie ihre Durchführung und die geplanten Analysen festgehalten.

Geeignete Plattformen hierfür sind z.B. *osf.io* oder *aspredicted.org*.

Wie oben erwähnt, ginge das Bekenntnis zu verlässlicher und offener Forschung grundsätzlich auch mit einer wohlwollenden Haltung gegenüber Replizierbarkeitsüberprüfungen einher—sei es in der passiven Form, dass wir Replikationen unserer eigenen Studien wohlwollend und kooperativ gegenüber stehen (z. B. durch das Bereitstellen von Materialien und Ablaufplänen) oder in der aktiven Form, dass wir selbst Replikationsstudien durchführen oder uns konzertierten Replikationsinitiativen (z. B. Registered Replication Reports, RRRs) anschließen.

Als Gutachter

Als Gutachter können wir die Werte verlässlicher und offener Wissenschaft ebenso etablieren, indem wir eben solche Maßstäbe an die zu begutachtenden Manuskripte anlegen. Beispielsweise können wir die Aussagen der 21-word solution von den Autoren eines Manuskripts einfordern (z.B. per „standard reviewer disclosure request“⁵, Nosek, Simonsohn, Moore, Nelson, Simmons, Sallans, & LeBel, 2014; *osf.io/hadz3*) und/oder nach Open Material, Open Data und reproduzierbarem Code fragen. So wird Autoren(-gruppen) die Gelegenheit gegeben, sich während des Schreibens des Antwortbriefs aktiv mit den Werten verlässlichen und offenen Forschens auseinander zu setzen und sich explizit zu ihnen positionieren.

Die Peer Reviewers' Openness (PRO) Initiative (Morey et al., 2016; *opennessinitiative.org*) geht noch einen Schritt weiter. Ihre Unterzeichner werden ab dem 01.

⁵ Das standard reviewer disclosure request (Nosek et al., 2014) lautet: "I request that the authors add a statement to the paper confirming whether, for all experiments, they have reported all measures, conditions, data exclusions, and how they determined their sample sizes. The authors should, of course, add any additional text to ensure the statement is accurate. This is the standard reviewer disclosure request endorsed by the Center for Open Science [see <http://osf.io/hadz3>]. I include it in every review."

Januar 2017 nur dann ein umfassendes Review eines Manuskripts zur Verfügung stellen, wenn die Minimalkriterien von Open Data und Open Material erfüllt sind, oder wenn im Manuskript eine Begründung genannt wird, warum eine Veröffentlichung der Materialien/Daten nicht möglich ist.

Als Betreuer

Als Betreuer von Abschlussarbeiten können wir verlässliches und offenes Forschen wahrscheinlich am nachhaltigsten fördern, denn wir sozialisieren im Betreuungsprozess die nächste Generation von Wissenschaftlern, die bestenfalls schon mit Verlässlichkeit und Offenheit als Selbstverständlichkeit heranwächst. Wir können in der Entstehung von Abschlussarbeiten bereits Methoden propagieren, die den Informationsgehalt und die Replizierbarkeit einer Studie erhöhen und diese von Beginn an mit den Abschlussarbeitskandidaten explizit diskutieren. Wir können ebenso die Erwartung formulieren, dass die Kandidaten mindestens uns Betreuern gegenüber Materialien, Daten und Analyseskripte in klar dokumentierter Form zur Verfügung stellen. Im Falle einer Studienreihe innerhalb von Doktorarbeiten könnten wir z.B. mindestens eine Studie einfordern, der eine Präregistrierung vorausgeht.

Als Herausgeber

Als Herausgeber von Zeitschriften können wir verlässliches und offenes Forschen fördern, indem wir auf die Umsetzung von Richtlinien hinwirken, die Autoren zur Einhaltung von Standards offener und transparenter Forschung verpflichten. Ein weit verbreitetes Set von Richtlinien sind die Transparency and Openness (TOP) Guidelines (Nosek et al., 2015; www.cos.io/top/), bei denen sich Zeitschriften zu einem bestimmten Level an verpflichtender Transparenz bekennen können. Eine weitere Maßnahme auf Zeitschriftenebene, die von Herausgebern mit unterstützt werden kann, ist das Kenntlichmachen transparenter Forschung mithilfe sogenannter „Badges“ (www.osf.io/tvyxz/), also deutlich sichtbaren Emblemen, die auf

der ersten Artikelseite anzeigen, ob die Materialien und/oder Daten verfügbar sind, und ob der Studie Präregistrierung vorausging. Schließlich können wir als Herausgeber von Zeitschriften deutlich machen, dass wir die Veröffentlichung von präregistrierten Studien und Replikationsstudien explizit begrüßen, indem wir uns für die Einführung von Sektionen für Replikationen und Registered Reports einsetzen.

Als Kommissionsmitglied

Als Kommissionsmitglied können wir dafür Sorge tragen, dass sich verlässliches und offenes Forschen auch in verbesserten Karrierechancen niederschlagen. Dazu gehört unter anderem eine über quantitative Indikatoren hinausgehende Evaluation von wissenschaftlichen Lebensläufen. Diese kostet unbestritten Zeit und ist aufwändig, aber schnell gezählte Anzahlen von Publikationen oder die Übersetzung der individuellen Publikationsleistung in vergleichsweise eindimensionale Indizes werden sowohl den Bewerbern als auch der Bedeutsamkeit und Seltenheit der zu besetzenden Stellen nicht gerecht. Ein solches Beispiel findet sich in einer Ausschreibung der LMU München für die W3 Professur Sozialpsychologie (https://www.academics.de/jobs/professur_w3_fuer_sozialpsychologie_121431.html). In dieser Ausschreibung werden die Bewerber gebeten, in ihrem Anschreiben darzulegen, auf welche Art und Weise sie die Ziele von transparenter und replizierbarer Forschung (inkl. Open Materials, Open Data, reproduzierbarem Code und Präregistrierungen) bereits verfolgt haben und in Zukunft verfolgen möchten. In diesem Sinne empfiehlt auch 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."⁶ . Sowohl die Bewerber als auch die Kommission sind also gefordert, sich zumindest explizit mit dem Thema zu befassen. Natürlich gilt hier zu berücksichtigen, dass viele Entwicklungen zur Forschungstransparenz erst in den letzten Jahren an Dynamik gewonnen haben und sie sich daher noch nicht bei jedem publizierten Artikel von Bewerbern niederschlagen können. Aber immerhin bietet diese Betonung einen weiteren Anreiz, aktuelle und zukünftige Forschung verlässlich und offen zu gestalten.

In allen wichtigen typischen Rollen eines Wissenschaftlers bieten sich uns also Möglichkeiten, offenes und verlässliches Forschen mit Nachdruck zu fördern.

Diskussion

In diesem Beitrag haben wir einen Überblick über die Vertrauenskrise in der (Sport-)Psychologie und ihre Ursachen gegeben und zwei sich ergänzende Wege aus dieser Krise aufgezeigt—verlässliches und offenes Forschen. Um diese Wege zu gehen und das angestrebte Ziel einer vertrauenswürdigen Wissenschaft, die auf replizierbaren Befunden beruht, zu erreichen, haben wir für die verschiedenen Rollen im Wissenschaftssystem jeweils individuelle Schritte vorgestellt, verlässliches und offenes Forschen zu fördern.

Wir haben betont, dass jeder einzelnen Untersuchung bereits eine verlässliche Grundlage unterliegen muss, um in Originalstudien wie Replikationsstudien oder in ihrer metaanalytischen Integration das Vorhandensein und die Größe eines Effekts abschätzen zu können. So schaffen wir eine verlässliche und replizierbare Basis für unsere theoretischen Schlussfolgerungen und steigern die Vertrauenswürdigkeit in die empirische Wissenschaft. Offenes Forschen hingegen stellt die Transparenz des Forschungsprozesses in den Mittelpunkt. Sie ermöglicht eine

6

http://www.dfg.de/foerderung/antragstellung_begutachtung_entscheidung/antragstellende/antragstellung/nachnutzung_forschungsdaten/

grundsätzliche Nachprüfbarkeit von Forschungsbefunden und bietet eine sehr gute Grundlage für ein verlässliches und kooperatives Forschen.

Dass die Bewegung hin zu verlässlichem und offenem Forschen bereits begonnen hat und unter den Wissenschaftlern mehr und mehr Beachtung findet, zeigen die steigende Anzahl an Selbstverpflichtungen (researchtransparency.org: 90 Unterzeichner; opennessinitiative.org: 313 Unterzeichner; beides Stand 30.09.2016), die Veränderung der Zeitschriftenrichtlinien (vgl. TOP-Guidelines) oder die Gründung von Organisationen und Initiativen wie das Center for Open Science oder die SIPS. Aber auch die steigende Zahl von Studien, die Materialien, Daten und reproduzierbaren Analysecode mitveröffentlichen bzw. Studien, in deren Veröffentlichung die 21-word-solution integriert ist, zeigen den Trend deutlich an (s. <https://replicationindex.wordpress.com/2014/12/17/the-r-index-of-simmons-et-al-s-21-word-solution/>). Darüber hinaus untermauert die steigende Anzahl einzelner Replikationsstudien, sowie von Replikationssektionen bei Zeitschriften und von konzertierten Replikationsbewegungen (z. B. Registered Replication Reports; RRRs; s. <http://www.psychologicalscience.org/index.php/replication>; z.B. Ego-Depletion, Professor Priming, Intuition and Cooperative Decision Making u.v.m.) das Ziel, die Vertrauenswürdigkeit unserer Forschung sicherzustellen. Auch auf organisationaler Ebene werden diese Entwicklungen deutlich. So hat z.B. die NWO (die niederländische Organisation für Wissenschaft) aktuell Drittmittel in Höhe von 3 Mio. Euro für die Förderung von Replikationsforschung bereitgestellt (<http://www.nwo.nl/en/news-and-events/news/2016/nwo-makes-3-million-available-for-replication-studies-pilot.html>).

Bei diesen vielversprechenden Entwicklungen möchten wir nicht unerwähnt lassen, dass der Wunsch nach verlässlicher und offener Forschung individuelle Forscher durchaus in ein Dilemma bringen kann. Dieses Dilemma umfasst insbesondere den Konflikt zwischen dem

Streben nach hoher Studienqualität, Transparenz, und ergebnisunabhängiger Veröffentlichung einerseits, sowie den zur Zeit noch fortbestehenden Anreizen nach hoher Quantität, positiver Befunde und überraschender Ergebnisse andererseits.

Offenes und verlässliches Forschen ist aufwändig, weil empirische Studien mit höherem Durchführungs- und Dokumentationsaufwand einhergehen. Der Durchführungsaufwand bezieht sich hier vor allem auf die Rekrutierung und Testung von mehr Probanden, um angemessen große Stichproben zu erzielen. Der Dokumentationsaufwand bezieht sich auf die Erstellung, Vorbereitung und Veröffentlichung von Materialien, Daten und reproduzierbarem Analysecode. Sie erfordern eine gewissenhafte Zusammenstellung, die durchaus Ressourcen binden kann. Die Präregistrierung erfordert diese Art gewissenhafter Zusammenstellung schon zu einem weit früheren Zeitpunkt im Zyklus einer empirischen Forschungsarbeit und setzt schon sehr detaillierte Planungen der Erhebung und Auswertung voraus. Aus unserer Sicht sind diese Anforderungen jedoch durch die hinzugewonnene Qualität der Studie schnell aufgewogen und führen auch unabhängig von den allgemeinen Zielen bereits laborintern zu verlässlicheren Abläufen. Auch die Präregistrierung im Speziellen sollte bei sorgfältiger Studienplanung keinen allzu großen Mehraufwand bedeuten (eher eine zeitliche Verschiebung mancherlei Entscheidungen und Aufwände).

Über den Aufwand hinaus wird die Entwicklung in Richtung des verlässlichen und offenen Forschens oft (noch) mit Skepsis und Unsicherheit aufgenommen. Drei Beispiele zeigen, welche Arten von neuen Fragen sich ergeben können und wie sich aus unserer Sicht mit Ihnen umgehen ließe (auf diese und viele weitere Fragen liefern auch die neuen Richtlinien der Deutschen Gesellschaft für Psychologie zum Thema "Umgang mit Forschungsdaten" Antworten; Schönbrodt, Gollwitzer & Abele-Brehm, 2016): (1) Wie kann bei der Veröffentlichung der Daten die Anonymität der Probanden sichergestellt werden? (2) Wie kann sichergestellt werden, dass

die Gruppe, die den Aufwand der Datenerhebung geleistet hat, Anerkennung für Ihren Aufwand erfährt, wenn der Datensatz zeitnah veröffentlicht und somit für jedermann zur Verfügung steht?

(3) Wie sollten wir vorgehen, wenn wir einen Fehler in einem veröffentlichten Datensatz oder in einem Analyseskript finden?

Eine erste Antwort auf die Frage nach der Anonymität von Forschungsdaten ist, dass grundsätzlich die meisten (sport-) psychologischen Datensätze leicht und effektiv anonymisierbar sind, sodass keinerlei Rückschluss auf Individuen mehr möglich ist. Wenn es aber natürlich um sehr ausgewählte Populationen geht, in der Sportpsychologie beispielsweise die Untersuchung von nationalen Spitzenathleten, so mag eine Anonymisierung erschwert oder unmöglich sein. Auch die Nutzung von vorhandenen Datensätzen (z. B. Sozio-oekonomisches Panel, SOEP) kann gegen eine korrespondierende Veröffentlichung der Rohdaten sprechen. In solchen Fällen müssten dann entweder weiterführende Vorkehrungen in der Datenaufbereitung getroffen werden (z. B. durch die Einführung einer Random-ID, durch das Löschen von demographischen Variablen, etc.), um eine eindeutige Zuordnung von Probanden zu Identitäten zu verhindern, oder das Manuskript würde eine entsprechende Erklärung in der *Author Note* enthalten, aus der hervorgeht, warum eine Veröffentlichung der Daten nicht möglich ist (s. auch FAQs auf researchtransparency.org; vgl. dazu auch der aktuelle Richtlinienentwurf der Deutschen Gesellschaft für Psychologie zum Thema "Umgang mit Forschungsdaten"; Schönbrodt, Gollwitzer & Abele-Brehm, 2016).

Als Antwort auf die zweite Frage zur Nutzung von veröffentlichten Daten durch fremde Wissenschaftler ist zu sagen, dass die Nutzung oder Reanalyse von Daten natürlich mindestens in einer zusätzlichen Zitation resultiert. Typischerweise, und das spiegeln unsere ersten Erfahrungen wider, resultiert solch ein Reanalysen- oder Datennutzungsvorhaben aber in einer Kooperationsanfrage und entsprechend ggf. in einer Koautorenschaft. Somit ginge die

Veröffentlichung von Daten tatsächlich weniger mit „Trittbrettfahrerei“ oder gar „Diebstahl“ einher, sondern im Gegenteil mit vielversprechenden Möglichkeiten zur Kollaboration. Im Übrigen gilt es hier natürlich ebenfalls, die Entwicklung dahingehend zu fördern, dass Zitationen und Autorenschaften nicht den einzigen Weg zur Anerkennung wissenschaftlicher Leistungen darstellen, sondern dass die Veröffentlichung von Daten und die Bereitstellung von Software u.ä. als genauso wertvoll angesehen werden wie ein empirischer Bericht oder ein Überblicksartikel.

Die Antwort auf die dritte Frage zum Umgang mit Fehlern in einer Veröffentlichung ist sehr situationsabhängig und hängt von den Auswirkungen dieses Fehlers auf die zentrale Aussage des Artikels ab. Eine mögliche Variante, mit entdeckten Fehlern umzugehen, wäre es, den korrespondierenden Autor in Kenntnis zu setzen und so z.B. ein Erratum oder, falls nötig, eine Zurückziehung des Artikels zu erreichen, oder, im Falle eines Analysefehlers, eine gemeinsame Reanalyse zu veröffentlichen (siehe z.B. Silberzahn, Simonsohn & Uhlmann, 2014). Bei Uneinsichtigkeit der Originalautoren kann aber auch ein unabhängiger Kommentar bzw. eine unabhängige Reanalyse nötig sein, um einen entdeckten Fehler richtig zu stellen. In jedem Fall zahlt es sich hier wieder aus, wenn Daten und Analyseskripte zur Überprüfung frei verfügbar sind. Nur so kann unser Ziel vertrauenswürdiger und somit bestenfalls fehlerfreier Veröffentlichungen erreicht werden.

Da sich die Entwicklung zu verlässlichem und offenem Forschen noch in einer frühen Phase befindet, sind diese Unsicherheiten und deren Diskussion besonders wichtig und ein zielführender Prozess. Insofern bietet diese Vertrauenskrise auch einen Anlass über unsere Forschungspraxis und unseren kollegialen Umgang miteinander zu reflektieren und neue, zeitgemäße Lösungen zu entwickeln. Zum jetzigen Zeitpunkt allerdings verlangt diese Entwicklung durchaus noch eine klare Haltung von uns Wissenschaftlern, um deren Wert überzeugt und bestenfalls auch überzeugend zu vertreten—sei es gegenüber Koautoren,

gegenüber Autoren, deren Manuskripte wir begutachten, gegenüber unseren Mitarbeitern und Abschlussarbeitskandidaten, gegenüber Zeitschriften und Verlagen, mit denen und für die wir arbeiten, gegenüber anderen Kommissionsmitgliedern in universitären und auch außeruniversitären Arbeitsgruppen. Diese Überzeugungsarbeit, aber auch die Etablierung neuer Routinen und Standards in der Arbeit unserer Forschungsgruppen und -kooperationen kann mitunter einen kontroversen Prozess darstellen, der sowohl auf fachlicher als auch auf sozialer Ebene herausfordernd sein kann.

Auch wenn die Umsetzung von verlässlichem und offenem Forschen individuell aufwändig und herausfordernd sein kann, stellt sie eine wichtige bottom-up Veränderung des wissenschaftlichen Systems dar. Sicherlich sollte diese über zusätzliche top-down Veränderungen⁷ des Anreizsystems und der Belohnungsstruktur sowie strukturellen Maßnahmen, beispielsweise umgesetzt durch die Universitäten und Verlage, aber auch durch die ASP, das BISP, den DAAD, die DFG, die DGPs, die DVS, die FEPSAC, die NASPSA, SIPS etc., ergänzt werden. Diese Veränderungen könnten weiterhin auch dafür Sorge tragen sollten, dass möglichen parallelen, aber nicht-intendierten Entwicklungen vorgebeugt wird. So könnte die Forderung nach größeren Stichproben und mehr Power auch dazu führen, dass die zusätzlich erforderten Ressourcen beispielsweise die Methodenvielfalt einschränken (z.B. Selbstbericht statt Verhaltensbeobachtung). Wege aus der Vertrauenskrise werden somit bestenfalls sowohl bottom-up als auch top-down bestritten—auf der Basis des Idealismus und der Überzeugung der einzelnen Wissenschaftler, dass verlässliches und offenes Forschen lohnenswert sind, und auf der Basis, dass das System, also wir gemeinsam als Forscher, Reviewer, Betreuer, Herausgeber und

⁷ Da top-down-Veränderungen oder strukturellen Maßnahmen in diesem Beitrag nicht adressiert wurden, aber ebenso wichtig für eine nachhaltige Veränderung der wissenschaftlichen Praxis sind, verweisen wir für deren Diskussion auf Beiträge von z. B. Asendorpf et al. (2013), Fraley und Vazire (2014) und Nosek et al. (2012).

Kommissionsmitglieder, veränderte Anreizstrukturen schaffen, in denen sich verlässliches und offenes Forschen zuverlässig in entscheidende Vorteile für die wissenschaftliche Laufbahn niederschlagen—sei in der Psychologie im Allgemeinen oder der Sportpsychologie im Speziellen. Denn das Ziel ist dasselbe: Die Vertrauenswürdigkeit unserer Wissenschaft, die auf Replizierbarkeit beruht.

Fazit

Die Vertrauenswürdigkeit einer Wissenschaft steht und fällt mit der Replizierbarkeit ihrer Befunde. Dieser Beitrag gab einen Überblick darüber, dass es in der (Sport-)Psychologie aktuell um die Replizierbarkeit nicht allzu gut bestellt ist, beschrieb Ursachen dieser Vertrauenskrise und zeigte zwei sich ergänzende Wege aus dieser Krise auf—verlässliches und offenes Forschen. In jeder unserer Rollen im Wissenschaftssystem haben wir vielfältige Möglichkeiten verlässliches und offenes Forschen zu fördern. Auch wenn diese Maßnahmen individuell zunächst aufwändig und herausfordernd zu sein scheinen, sind sie aus unserer Sicht nötig und lohnenswert, um das kollektive Ziel einer replizierbaren und vertrauenswürdigen Forschung zu erreichen.

Literaturverzeichnis

- Asendorpf, J. B., Conner, M., De Fruyt, F., De Houwer, J., Denissen, J. J. A., Fiedler, K., et al. (2013). Recommendations for increasing replicability in psychology. *European Journal of Personality, 27*, 108–119.
- Bacon, R. (1859). *Fr. Rogeri Bacon Opera quaedam hactenus inedita. Vol. I. containing I.—Opus tertium. II.—Opus minus III.—Compendium philosophiæ*. Longman, Green, Longman and Roberts. Retrieved from <http://books.google.com/books?id=wMUKAAAAYAAJ> (Original work published 1267).
- Bakker, M., van Dijk, A., & Wicherts, J. M. (2012). The rules of the game called psychological science. *Perspectives on Psychological Science, 7*, 543–554.
- Bakker, M., Hartgerink, C. H., Wicherts, J. M., & van der Maas, H. L. (2016). Researchers' intuitions about power in psychological research. *Psychological Science, 0956797616647519*.
- Braude, S. E. (1979). *ESP and psychokinesis. A philosophical examination*. Philadelphia, PA: Temple University Press.
- Büsch, D. & Strauss, B. (in Druck). *Wider die Sternchenkunde. Diskussionsbeitrag zur empirischen Sportwissenschaft*. Sportwissenschaft.
- Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., . . . Chan, T. (2016). Evaluating replicability of laboratory experiments in economics. *Science, 351*(6277).
- Chambers, C. D. (2015). Ten reasons why journals must review manuscripts before results are known. *Addiction, 110*, 10-11.
- Chambers, C. D., Dienes, Z., McIntosh, R. D., Rotshtein, P., & Willmes, K. (2015). Registered reports: realigning incentives in scientific publishing. *Cortex, 66*, A1-A2.

- 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.
- Cumming, G., & Finch, S. (2005). Inference by eye: Confidence intervals, and how to read pictures of data. *The American Psychologist*, 60, 170–180.
- De Vries, R., Anderson, M. S., & Martinson, B. C. (2006). Normal misbehavior: Scientists talk about the ethics of research. *Journal of Empirical Research on Human Research Ethics*, 1, 43–50.
- Dreiskämper, D. (2016). Die „Vertrauenskrise“ der empirischen Sozialwissenschaften und deren Bedeutung für die Sportpsychologie - Ein Kommentar aus der Perspektive der Vertrauensforschung. *Zeitschrift für Sportpsychologie*.
- Ebersole, C. R., Atherton, O. E., Belanger, A. L., Skulborstad, H. M., Adams, R. B., Allen, J., ... Nosek, B. A. (2015, September 27). *Many Labs 3: Evaluating participant pool quality across the academic semester via replication*. Retrieved from osf.io/ct89g.
- Engel, C. (2015). Scientific disintegrity as a public bad. *Perspectives on Psychological Science*, 10, 361-379.
- Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, 315, 629–634.
- Errington, T. M., Iorns, E., Gunn, W., Tan, F. E., Lomax, J., Nosek, B. A. (2014). An open investigation of the reproducibility of cancer biology research. *eLife*. 3, 04333.
- Feeg, M. (2011). A science of progress is a reliable science: A reply to limits of scientific methods. *The Gauntlet*, 2011-10-13, retrieved from <http://www.archive.thegauntlet.ca/story/science-progress-reliable-science>.
- Fiedler, K., & Schwarz, N. (2015). Questionable research practices revisited. *Social Psychological and Personality Science*, 1948550615612150.

- Fraley, R. C., & Vazire, S. (2014). The N-pact factor: Evaluating the quality of empirical journals with respect to sample size and statistical power. *PLoS ONE*, 9, e109019.
- Franco, A., Malhotra, N., & Simonovits, G. (2015). Underreporting in Psychology Experiments: Evidence from a study registry. *Social Psychological and Personality Science*, 7, 8–12.
<http://doi.org/10.1177/1948550615598377>
- Galak, J., LeBoeuf, R. A., Nelson, L. D., & Simmons, J. P. (2012). Correcting the past: Failures to replicate psi. *Journal of Personality and Social Psychology*, 103, 933-948.
- Gelman, A., & Loken, E. (2013). *The garden of forking paths: Why multiple comparisons can be a problem, even when there is no fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time*. Unpublished Manuscript.
- Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Med* 2, e124.
- Jasny, B. R., Chin, G., Chong, L., & Vignieri, S. (2011). Again, and again, and again. *Science*, 334, 1225.
- John, L., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth-telling. *Psychological Science*, 23, 524–532.
- Klein, R. A. et al. (2014). Investigating variation in replicability: A „Many Labs“ replication project. *Social Psychology*, 45, 142-152.
- Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Alper, S., ... Friedman, M. (2016, March 5). Many Labs 2: Investigating Variation in Replicability Across Sample and Setting. Retrieved from osf.io/8cd4r.
- Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press.
- Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in Psychology Research: How Often Do They Really Occur? *Perspectives on Psychological Science*, 7, 537-542.

- Maxwell, S. E. (2004). The persistence of underpowered studies in psychological research: causes, consequences, and remedies. *Psychological Methods, 9*, 147.
- Meehl, P. E. (1990). Why summaries of research on psychological theories are often uninterpretable. *Psychological Reports, 66*, 195-244.
- Merton, R. K. (1942). A note on science and democracy. *Journal of Legal and Political Sociology, 1*, 115– 126.
- Morey, R. D. et al. (2016). The Peer Reviewers' Openness Initiative: Incentivizing open research practices through peer review. *The Royal Society*.
- Myers, N. D., Ahn, S., & Jin, Y. (2011). Sample size and power estimates for a confirmatory factor analytic model in exercise and sport: a Monte Carlo approach. *Research Quarterly for Exercise and Sport, 82*, 412-423.
- Nelson, L. D., Simmons, J. P., & Simonsohn, U. (2012). Let's publish fewer papers. *Psychological Inquiry, 23*, 291–293.
- Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., et al. (2015). Promoting an open research culture. *Science, 348*, 1422–1425.
- Nosek, B. A., & Lakens, D. (2014). Registered reports: A method to increase the credibility of published results. *Social Psychology, 45*, 137-141.
- Nosek, B. A., Simonsohn, U., Moore, D. A., Nelson, L. D., Simmons, J. P., Sallans, A., & LeBel, E. P. (2014, February 4). *Standard Reviewer Statement for Disclosure of Sample, Conditions, Measures, and Exclusions*. Retrieved from osf.io/hadz3
- 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.

Open Science Collaboration (in press). Maximizing the reproducibility of your research. In S. O.

Lilienfeld & I. D. Waldman (Eds.), *Psychological Science Under Scrutiny: Recent Challenges and Proposed Solutions*. New York, NY: Wiley.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349, aac4716.

Open Science Collaboration. (2012). An open, large-scale, collaborative effort to estimate the reproducibility of psychological science. *Perspectives on Psychological Science*, 7, 657–660.

Pashler, H., & Wagenmakers, E.-J. (Eds.). (2012). Special section on replicability in psychological science: A crisis of confidence? [Special section]. *Perspectives on Psychological Science*, 7, 528-654.

Popper, K. (1992). *The logic of scientific discovery*. New York, NY: Routledge. (Original work published 1934)

Schimmack, U. & Heene, M. (2016). Die Verdrängung des selektiven Publizierens: 7 Fallstudien von prominenten Sozialpsychologen. Gefunden unter:
<https://replicationindex.wordpress.com/2016/04/20/die-verdrangung-des-selektiven-publizierens-7-fallstudien-von-prominenten-sozialpsychologen/>

Schönbrodt, F. D., Gollwitzer, M., & Abele-Brehm, A. (2016). *DGPs-Richtlinien zum Umgang mit Forschungsdaten*. Verfügbar unter https://www.dgps.de/uploads/media/DGPs-Richtlinien_Forschungsdaten_22_01_16.pdf

Schönbrodt, F. D., Maier, M., Heene, M., & Zehetleitner, M. (2015). *Voluntary commitment to research transparency*. Retrieved from <http://www.researchtransparency.org>.

Schweizer, G., & Furley, P. (2015). Reproducible research in sport and exercise psychology: The role of sample sizes. *Psychology of Sport & Exercise*. Online First: doi: 10.1016/j.psychsport.2015.11.005

Scuzs, D. (2016). A Tutorial on Hunting Statistical Significance by Chasing N. *Frontiers in Psychology: Quantitative Psychology and Measurement*, 7, 1444. doi:10.3389/fpsyg.2016.01444

Silberzahn, R., Simonsohn, U., & Uhlmann, E. L. (2014). Matched-Names Analysis Reveals No Evidence of Name-Meaning Effects A Collaborative Commentary on Silberzahn and Uhlmann (2013). *Psychological Science*, 25, 1504–1505. doi:10.1177/0956797614533802

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.

Smaldino, P. E., & McElreath, R. (2016). The natural selection of bad science. Retrieved from <https://arxiv.org/abs/1605.09511>

Spellman, B. A. (2012). Introduction to the special section: Data, data, everywhere... especially in my file drawer. *Perspectives on Psychological Science*, 7, 58–59.

Stanley, T. D., & Doucouliagos, H. (2013). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods*, 5, 60–78.

Valentine, J. C., Biglan, A., Boruch, R. F., Castro, F. G., Collins, L. M., Flay, B. R., ... Schinke, S. P. (2011). Replication in prevention science. *Prevention Science*, 12, 103–117.

van Aert, R. C. M., Wicherts, J. M., & M. A. L. M. van Assen (in press). Conducting meta-analyses based on p-values: Reservations and recommendations for applying p- uniform and p-curve. *Perspectives on Psychological Science*.

- van Assen, M. A. L. M., Van Aert, R. C. M., & Wicherts, J. M. (2015). Meta-analysis using effect size distributions of only statistically significant studies. *Psychological Methods*, *20*, 293-309. doi: <http://dx.doi.org/10.1037/met0000025>
- Vanpaemel, W., Vermorgen, M., Deriemaecker, L., & Storms, G. (2015). Are We Wasting a Good Crisis? The Availability of Psychological Research Data after the Storm. *Collabra*, *1*, 1-5.
- Wagenmakers, E. J., Wetzels, R., Borsboom, D., & van der Maas, H. L. (2011). Why psychologists must change the way they analyze their data: the case of psi: comment on Bem (2011). *Journal of Personality and Social Psychology*, *100*, 426-432.
- 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.
- Wicherts, J. M., Borsboom, D., Kats, J., & Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. *American Psychologist*, *61*, 726-728.
- Zhu, W. (2012). Sadly, the earth is still round ($p < 0.05$). *Journal of Sport and Health Science*, *1*, 9-11.
- Zhu, W. (2013). Reliability: What type, please! *Journal of Sport and Health Science*, *2*, 62-64.

Tabellen

Tabelle 1. *Individuelle Schritte zu verlässlicher und offener Forschung*

| Rolle | Individuelle Schritte zu verlässlicher und offener Forschung |
|---------------------|---|
| Forscher | <ul style="list-style-type: none"> ▪ Bereitstellen von Open Materials, Open Data und reproduzierbaren Codes zu Studien (z. B. als Online Supplement oder als Projekt unter www.osf.io) ▪ Berechnen von Power a priori (Daumenregel: anzunehmen ist ein ca. halb so großer Effekt wie in der Originalstudie/orientieren an Effektgrößen aus Replikationsstudien) ▪ Präregistrieren von Studien (unter OSF.io, aspredicted.org oder direkt bei Journals als Registered Reports) ▪ Integrieren der „21-word solution“ in Veröffentlichungen (Simmons, Nelson, & Simonsohn, 2012) ▪ Unterzeichnen des „Commitment to Research Transparency and Open Science“ (http://www.researchtransparency.org) oder Beitreten von Organisationen wie SIPS |
| Gutachter | <ul style="list-style-type: none"> ▪ Einfordern von Aussagen der „21-word solution“ (Simmons, Nelson, & Simonsohn, 2012) oder Nutzen des „standard reviewer disclosure request“ (Nosek, Simonsohn, Moore, Nelson, Simmons, Sallans, & LeBel, 2014; osf.io/hadz3) ▪ Unterzeichnen der Peer Reviewers‘ Openness (PRO) Initiative (Morey et al., 2016; http://www.opennessinitiative.org) |
| Betreuer | <ul style="list-style-type: none"> ▪ Fördern (Vorleben, Diskutieren) von offener und verlässlicher Forschung ▪ Einfordern (zumindest gegenüber den Betreuern) von Open Materials, Open Data und reproduzierbaren Codes ▪ Einfordern von mindestens einer Studie innerhalb einer Doktorarbeit der eine Präregistrierung (inkl. a priori Poweranalyse, Beschreibung der Hypothesen, des Designs und der geplanten Analysen) voraus geht |
| Herausgeber | <ul style="list-style-type: none"> ▪ Hinwirken auf die Umsetzung von Guidelines, die Autoren zur Einhaltung von Standards offener und transparenter Forschung verpflichten ▪ Hinwirken auf die Unterzeichnung und Umsetzung der Transparency and Openness (TOP) Guidelines (Nosek et al., 2015; www.cos.io/top/) ▪ Hinwirken auf die Prämierung transparenter Forschung mithilfe sogenannter „Badges“ (www.osf.io/tvyxz/) ▪ Ermöglichen der Publikation von Replikationen (z.B. auch als registered reports) |
| Kommissionsmitglied | <ul style="list-style-type: none"> ▪ Hinwirken darauf, dass sich verlässlichere und offenere Forschung in verbesserten Karrierechancen niederschlägt ▪ Hinwirken auf eine faire Evaluation von individueller Publikationsleistung (auch) an den Werten verlässlicher und offener Forschung |

-
- Hinwirken darauf, dass die Haltung eines Kandidaten zu verlässlicher und offener Forschung explizite Berücksichtigung im Auswahlprozess findet
-