



Workstattberichte – subjektive und disziplinäre Zugänge zur Wissenschaft

Müllhalde Wissenschaft?

Zur Replikationskrise in der Psychologie

Anja S. Göritz

1 Ich danke Helmut Fink für hilfreiche Anmerkungen zu diesem Beitrag.

Seit gut zehn Jahren wütet die sogenannte Replikationskrise in der Psychologie und in anderen Wissenschaften wie der Medizin.¹ Zum persönlichen Weckruf wurde für mich der Artikel von John Ioannidis *Why most published research findings are false*. Seit der Veröffentlichung dieses Artikels haben vielfache Replikationsstudien gezeigt, dass möglicherweise sechzig und mehr Prozent – jedenfalls erschreckend viele – der in angesehenen Fachzeitschriften veröffentlichten Ergebnisse nicht replizierbar sind.

Die Replikationskrise darf als ein Produkt einer mußelosen Wissenschaft gelten, die sich vielfach aus Konkurrenz und Selbstbehauptung heraus begreift, die also nicht das aufrichtige Verstehen-Wollen und die gesicherte Erkenntnis als Hauptziel der eigenen Bemühungen fasst, sondern nach Aufmerksamkeit, Reputation und Forschungsmitteln heischt und sich damit entfernt hat vom selbstzweckhaften Anspruch der Erkenntnis um der Erkenntnis willen. Es scheint, als habe der Erfolg die Erkenntnis als Ziel abgelöst. Allerdings glaube ich nicht, dass diese systemischen Gründe die Replikationskrise bereits hinreichend erklären.

Wie konnte es passieren, dass

- meist vom Steuerzahler aufgebrauchte Mittel
- weltweit
- jahrzehntelang
- in astronomischer Höhe

für ‚Ausschuss‘ verwendet wurden? Denn mit dem Anspruch auf Verallgemeinerbarkeit veröffentlichte Ergebnisse, die in der Folge nicht repliziert werden können, stellen in empirisch-nomothetischen Fächern Ausschuss dar. Bei nicht-replizierbaren Ergebnissen handelt es sich gerade nicht um Gesetzmäßigkeiten, um deren Erforschung es geht und welche in der Tat publikationswürdig wären.

Welche therapeutischen Irrtümer, leeren Versprechen, nicht als solche gekennzeichneten Placebo-Interventionen, falschen Ratschläge und fehlgeleiteten Re-

gelungen, die ihrerseits enorme Ressourcen verschwendeten und millionenfach Schaden oder Tod über Betroffene brachten, wurden von Müll-Befunden der Wissenschaft ‚informiert‘?

Für mein Empfinden ist die Replikationskrise nicht irgendein nun glücklicherweise aufgedeckter Missstand, hinter dem gleich wieder die Sonne aufgeht, sondern sie offenbart ein Systemversagen der (Natur-)Wissenschaft. Diese Ent-Täuschung konnte mir passieren aufgrund meiner Sozialisation als (Natur-)Wissenschaftlerin; eine Wissenschaftlerin, die nicht so naiv war zu glauben, dass alles stimmt, was veröffentlicht wird – aber eben doch das meiste.

Auch ich bin seit Jahren Teil des Systems, sodass ich mich fragen muss, wie viel Ausschuss ich in meiner wissenschaftlichen Laufbahn produziert habe. Wie viel Geld habe ich verschwendet? Welchen Schaden habe ich verursacht? Bin ich ein Scharlatan?

Obwohl absichtliche Fälschungen vereinzelt nachgewiesen wurden, erklärt sich dadurch wohl nur ein kleiner Teil der nicht replizierbaren Befunde. Der weit größere Teil dürfte zurückgehen auf

- in die wissenschaftliche Methodik eingebaute Dilemmata (z.B. die Wahl des Signifikanzniveaus beim statistischen Testen, wobei sich eine Forscherin zwischen mehr falsch-positiven oder mehr falsch-negativen Urteilen entscheiden muss),
- Gesetzmäßigkeiten der Verbreitung wissenschaftlicher Befunde (z.B. bessere Publikationschancen und Rezeption von Studien, die neue, verblüffende oder ‚sexy‘ Ergebnisse berichten),
- als fragwürdig bekannte, aber kaum vermeidbare Forschungspraktiken (z.B. Ausschluss von Versuchspersonen, deren Daten sperrig sind),
- ganz „normale“, nicht fragwürdige Forschungspraktiken (z.B. Verwendung eines bestimmten Auswertungsverfahrens, für dessen Anwendung es gute Gründe gibt, bei gleichzeitigem Vorhandensein anderer Auswertungsverfahren, für deren Anwendung es ebenfalls gute Gründe gegeben hätte), somit aber eine mangelnde Offenlegung der Kontingenz von Forschungspraktiken

Vor den eingebauten Schwächen von Forschungspraktiken und den von ihnen bewirkten Methodenartefakten ist keine Wissenschaftlerin und kein Wissenschaftler gefeit. Wie sollte das auch möglich sein? Man kann nicht *nicht* handeln, wenn man forscht. Und man kann dieses Handeln auch nicht völlig transparent machen, unter Kontrolle bringen und von subjektiven und situativen Einflüssen bereinigen. Die ergebnisbeeinflussende Wirkung von Forschungspraktiken halte ich im Einzelnen für

- nicht vermeidbar, weil hier Dilemmata lauern oder weil schlicht eine Forschungspraktik ausgewählt werden muss,
- nicht ausrottbar, da Wissenschaft von Menschen betrieben wird. Die Wissenschaftler-Menschen weisen erkenntnisbeeinflussende kognitiv-emotionale Universalien auf (z.B. Abtasten der Wahrnehmungsinhalte nach Mustern, Überzeugtwerden durch Geschichten, Verteidigung von Glaubensinhalten, Ansprechen auf Neuartiges). Die Wissenschaftler-Menschen haben ihre je eigenen Brillen auf (u.a. ihre Kulturbrille, Geschlechtsbrille, Schichtbrille,

Identitätsbrille). Die Wissenschaftler-Menschen sprechen auf bestimmte Anreizstrukturen im Wissensschaffungsprozess und im Wissenschaftsbetrieb an, - nicht erfass- und bezifferbar, denn Forschungspraktiken sind komplex, da es bei jedem Forschungsprojekt unzählige Mikro-Entscheidungen zu fällen gilt, welche miteinander wechselwirken. Das Zustandekommen von Forschungspraktiken ist subtil und in seiner Erfahrungsetränktheit nicht einmal vom Forschenden selbst vollständig erfass- oder beschreibbar.

Seitenlange Methodenberichte in Forschungsartikeln, welche das jeweilige Vorgehen transparent machen sollen, würden – abgesehen davon, dass kaum jemand sie lesen möchte – wahrscheinlich entscheidende Mikro-Entscheidungen oder Kontextinformationen nicht enthalten. Man steht hier vor der Herkulesaufgabe, pralle Wirklichkeit auf ein paar Seiten Text einzufangen. Als ein weiteres ‚Heilmittel‘ wurde die sogenannte Präregistrierung von Studien vorgeschlagen, und sie wird in Einzelfällen bereits exerziert: Eine Forschende soll dabei mit vorab schriftlich festgelegter, dann nur noch abzuarbeitender Verfahrensweise die Daten erheben, auswerten, berichten und interpretieren. In seiner Starrheit und Blindheit wird man jedoch den Daten und ihrem lebensweltlichen Zustandekommen oftmals nicht gerecht. Präregistrierung vermindert vielleicht manche derzeit häufigen Fehler, verursacht aber auch neue. Überhaupt führt jede (Über-)Bürokratisierung – wie sie einer noch ausführlicheren Methodendarlegung innewohnt – dazu, dass neue Schlupflöcher entstehen. Die Dynamik des Wettrüstens ist aus Biologie und Gesellschaft hinlänglich bekannt. Abgesehen vom fraglichen Gesamtnutzen einer noch weitergehenden Methodendarlegungspflicht würde dann vermutlich nicht nur mir das wissenschaftliche Arbeiten weniger Freude bereiten.

Wie soll das System Wissenschaft nun reagieren? Womöglich hilft deutlich und nachhaltig nur ein fachöffentlicher Reputationsmonitor, bei dem Forschende mit einem Vertrauensvorschuss starten. Abschreckend ausgedrückt ist dies ein Pranger, an dem Forschende für ihre nicht replizierbaren veröffentlichten Ergebnisse mit Reputationseinbußen und/oder Kürzung ihrer Forschungsmittel bestraft werden. Ansprechend ausgedrückt handelt es sich um eine *Hall of Fame*. Die vorhersehbaren Nachteile eines solchen Reputationsmonitors (u.a. Denunziationen, Fehden, tatsächlich belastbare Erkenntnisse kommen später an die Öffentlichkeit) gälte es durch geschickte Konzeption möglichst klein zu halten. Beispielsweise könnte man ein Stück der verlorenen Geschwindigkeit bei der Veröffentlichung von Befunden zurückgewinnen durch eine Kennzeichnung des eingeschätzten Evidenzgrades der berichteten Befunde durch die Autoren selbst, etwa auf einer Skala von ‚1=spekulativ‘ bis ‚10=höchstwahrscheinlich stabil‘. Für Ähnliches gibt es bereits Evidenzbewertungsschemata, die bei Metaanalysen für die Bewertung von einfließenden Primärstudien verwendet werden. Ein Reputationsverlust durch Nichtreplikation von als spekulativ gekennzeichneten Befunden wäre gering bis nichtig. Allerdings wäre auch der Reputationsgewinn geringer, wenn sich ein als spekulativ gekennzeichneteter Befund (zu dem man ja vergleichsweise schnell gelangt) erhärtet, als wenn dies bei einem Befund geschieht, für dessen Gediegenheit sich eine Autorin verbürgt hatte (und für welchen sie wahrscheinlich eine längere Studienreihe durchführen musste).

Auf diese Weise könnte das Reputationssystem verschiedenen Naturellen von Wissenschaftlerinnen und Wissenschaftlern gerecht werden: Einige können sich im schnellen und breiten Kartieren von wissenschaftlichem Neuland ausleben, andere eng und tief forschen, wieder andere sich auf die Überprüfung der Ideen und Befunde Anderer verlegen. Vermutlich würden sich auch Fachzeitschriften ausdifferenzieren in solche, die vorwiegend als spekulativ gekennzeichnete Werke veröffentlichen, und solche, die nur Befunde mit hohem Evidenzgrad herausbringen. Als Begleiteffekt der mit einem Reputationsmonitor einhergehenden Aufwertung von Qualität zu Lasten von Quantität würde die Artikelflut eingedämmt werden. Heutzutage wird der Artikelschwemme selbst eine lesewütige Wissenschaftlerin in ihrem eigenen kleinen Forschungsfeld nicht mehr Herrin; auch belastet die Artikelschwemme die Begutachtungssysteme von Zeitschriften.

Jedenfalls finde ich, dass individuelle Wissenschaftler und Wissenschaftlerinnen sich nicht mehr so leicht hinter dem System verstecken können sollten. Sind wir die Erarbeitung verlässlicher Erkenntnisse und damit die Erfüllung unserer Daseinsberechtigung als Wissenschaftlerinnen und Wissenschaftler nicht der uns alimentierenden Gesellschaft schuldig? Sind wir die Beseitigung des Müllberges nicht auch dem Ethos schuldig, mit dem viele von uns ihren Weg in die Wissenschaft gefunden haben – der Suche nach Wahrheit?

Ich meine, wir haben nicht nur eine Bringschuld; wir sollten schon aus Eigeninteresse die Müllverbrennung anlaufen lassen: Wer möchte sich einerseits schon dauernd selbst täuschen über den Zustand der Psychologie (und anderer Disziplinen)? Wer kann andererseits mit Dauer-Zynismus produktiv und zufrieden arbeiten?

Es bleibt zu hoffen, dass die Befunde der Replikationsstudien, welche die Replikationskrise eingeläutet hatten, sich in nicht allzu ferner Zukunft selbst nicht mehr replizieren lassen.

Empfohlene Zitierweise:

Anja S. Göritz: Müllhalde Wissenschaft? Zur Replikationskrise in der Psychologie
In: *Muße. Ein Magazin*, 5. Jhg. 2020, Heft 2, S. 97–100
DOI: 10.6094/musse-magazin/5,8.2020.97
URL: <http://mussemagazin.de/2020/03/muellhalde-wissenschaft/>

*Gefördert durch die Deutsche Forschungsgemeinschaft (DFG)
Projektnummer 197396619/SFB 1015*