



Problem, Lösung oder Symptom?

Zur Forderung nach Replizierbarkeit von Forschungsergebnissen

| JÖRG STRÜBING | **Die Replikation von Studien ist in der Wissenschaft ein anerkanntes Mittel, um die Reliabilität standardisierter empirischer Studien zu bestimmen. Doch taugt sie als „generelles Kriterium wissenschaftlicher Erkenntnis“ bzw. als Qualitätsmaßstab der Forschung? Einige kritische Anmerkungen zur Replikationsdebatte.**

Im April 2017 hat die DFG in einer Stellungnahme zur „Replizierbarkeit von Forschungsergebnissen“ bereits sehr differenziert und kritisch auf die allmählich Fahrt aufnehmende Replikationsdebatte reagiert und betont, dass Replizierbarkeit „kein generelles Kriterium wissenschaftlicher Erkenntnis“ ist. Vielmehr sei die Feststellung der Replizierbarkeit oder Nicht-Replizierbarkeit von Forschungsergebnissen schon für sich ein „wissenschaftliches Ergebnis“, dass aber „weiterer Überprü-

fung“ bedürfe. Nicht-Replizierbarkeit könne z.B. „als Signal für die Existenz eines neuen, bisher noch unbekannten wissenschaftlichen Zusammenhangs gedeutet werden und insofern ein Hinweis auf im Sinne des wissenschaftlichen Erkenntnisfortschritts positiv riskante Forschung sein“. Man kann die DFG-Stellungnahme als eine Reaktion auf die in Wissenschaft und Öffentlichkeit verstärkt anzutreffende Stilisierung der Replikation zum zentralen Qualitätsmaßstab der Forschung verstehen. Die

AUTOR



Jörg Strübing ist Professor für Soziologie an der Universität Tübingen und vertritt dort insbesondere Methoden und Methodologien der qualitativen Sozialforschung. Von 2014 bis 2017 war er Mitglied im Rat für Sozialwissenschaftliche Daten der Bundesregierung (RatSWD).



Foto: picture-alliance

Langzeitarchivierung von Daten wird stark auch mit dem Argument der Replizierbarkeit begründet, und wissenschaftliche Zeitschriften drängen zunehmend auf die Bereitstellung der Forschungsdaten zu den eingereichten Papers, um die Plausibilität der präsentierten Ergebnisse prüfen zu können.

Replikation als Qualitäts- maßstab der Forschung?

Sicherlich, die Replikation von Studien ist in der Wissenschaft ein anerkanntes Mittel, um die Reliabilität standardisierter empirischer Studien zu bestimmen. Reliabilität ist allerdings lediglich eines von mehreren Gütekriterien und bezieht sich im Kern nur auf die Verlässlichkeit der Messung, also etwa eines Experiments oder eines Fragebogens: Eine Operationalisierung, die nicht valide ist, kann in irrelevanten Messungen resultieren, die man vielleicht stabil replizieren kann – der Mangel in der Operationalisierung wird da-

mit jedoch weder aufgedeckt noch geheilt.

Man kann Replikation auch in einem umfassenderen Sinne verstehen, als Wiederholung einer Untersuchung insgesamt, also einschließlich der Operationalisierung und der Interpretation der Analysen. Eine Güteprüfung im strengen Sinne ist das aber insofern nicht, als die Theoriebezüge, die Hypothesenkonstruktion und auch die Wahl und das Design von Instrumenten kei-

»Reliabilität ist lediglich eines von mehreren Gütekriterien und bezieht sich im Kern nur auf die Verlässlichkeit der Messung.«

nen unilinearen und zwingenden Zusammenhang bilden, der immer in gleicher Weise gewählt werden „muss“, damit die Studie „gut“ wird. Mit anderen Worten: Eine Replikation der ganzen Studie taugt nicht wirklich zur Prüfung der Qualität der Ausgangsstudie, son-

dern ist im Prinzip eine erneute Untersuchung des Gegenstandsbereichs mit der im Prinzip gleichen Forschungsfrage. Denn die Vielzahl der Entscheidungen, die im Vorfeld standardisierter Messungen zu treffen sind, führt unweigerlich dazu, dass die Replikation einer Studie am Ende eine (mehr oder weniger) andere Studie sein wird. Abweichungen in den Ergebnissen können dann *entweder* darauf zurückzuführen sein, dass (begründet und sinnvoll) un-

terschiedliche Entscheidungen im Operationalisierungsprozess getroffen wurden *oder* dass in der einen oder der anderen Studie Fehler gemacht wurden. Erst wenn *alle*

Entscheidungen im Verlauf eines Forschungsprojekts komplett und umfassend dokumentiert würden, könnte eine Replikationsstudie den Versuch unternehmen, die ursprüngliche Studie detailgenau nachzubilden. Ob Aufwand und Ertrag dann aber noch in einem

vertretbaren Verhältnis stehen, lässt sich für viele Forschungsfelder kaum ungezwungen mit ja beantworten.

Hinzu kommt, dass sich bei weitem nicht alle Studien mit dem Mittel der Replikation prüfen lassen. Wenn die standardisierte Forschung Replikation zur Kontrolle der Qualität, also der Verlässlichkeit der (im weitesten Sinne) Beobachtungen nutzt, dann steckt dahinter die axiomatische Mess-theorie mit einem sehr spezifischen (wenn auch weit verbreiteten und in vielen Feldern bewährten) Verständnis der Organisation des wissenschaftlichen Zugangs zum interessierenden Wirklichkeitsbereich. Es wird die vereinfachende Idealisierung zugrunde gelegt, dass die Realität unabhängig vom Beobachter gegeben ist und durch geeignete Messungen objektiv verfügbar gemacht werden kann. Innerhalb dieses vereinfachenden

Rahmens funktioniert das Modell des Messens leidlich stabil und zuverlässig. Eine ganze Reihe von (vorwiegend natur-)wissenschaftlich relevanten Phänomenen sind so beschaffen, dass sich mit standardisierten Messungen hinreichend genaue Aussagen treffen lassen, die dann auch mit dem Mittel der Replikation auf ihre Zuverlässigkeit geprüft werden können.

In der mir besonders vertrauten qualitativen Sozialforschung hingegen, die (sehr kurz gefasst) mit intensiver Präsenz im Untersuchungsfeld sowie mit kontextreichem Datenmaterial (Texte, Audio- u. Videoaufzeichnungen) empirische Fälle interpretiert, rekonstruiert und fallvergleichend Theorie generiert, eignet sich Replikation grundsätzlich nicht zur Prüfung der Qualität von Studien oder auch nur von einzelnen Interpretationsschritten. Das hat verschiedene Gründe. Einer davon liegt in den sozial- und wissenschaftstheoretischen Vorannahmen: Es kann hier nicht davon ausgegangen werden, dass Realität universell und vom Beobachter unabhängig einfach besteht. Datengewinnung und -analyse werden als soziale Prozesse verstanden, in die die Forschenden nicht nur unvermeidlich, sondern auf produktive Weise involviert sind. Die Wiederholung einer Studie durch ein anderes Team von Forschenden erbringt daher andere, aber nicht

notwendig bessere oder schlechtere Ergebnisse. Replikationen im engeren Sinne funktionieren hier auch deshalb nicht, weil unser Feld im Prozess seiner interaktiven Erforschung „lernt“ und sich verändert, jede weitere Untersuchung also im Prinzip auf ein anderes Feld trifft.

Woher rührt die Überhöhung?

Bedenkt man diese relativ begrenzte Aussagekraft von Replikationsstudien und das ungünstige Verhältnis von Aufwand und Ertrag, dem sie unterliegen, dann lohnt es vielleicht, den Blick ein wenig über den methodologischen und methodentechnischen Tellerrand schweifen zu lassen und sich zu fragen, warum die Forderung nach Replikation und Replizierbarkeit gerade in den letzten Jahren so nachdrücklich erhoben wird. Woher also kommt diese Überhö-

»Es scheint, als wäre der Kompass wissenschaftlichen Wohlverhaltens aus der inneren Haltung der Forschenden in ein forschungsethisches Exoskelett aus externen Agenturen und Instrumenten verlagert worden.«

hung und was versprechen sich relevante Akteure im Feld der Wissenschaften und hier insbesondere Fachzeitschriften und Fachgesellschaften davon?

Geht man dieser Frage nach, stößt man schnell auf Fälle wissenschaftlichen Fehlverhaltens, insbesondere die Veröffentlichung von Ergebnissen, die durch die verfügbaren Daten nicht gedeckt sind. Dabei reicht die Palette von irrtümlichen über absichtliche Fehlinterpretationen von Daten bis hin zu deren kompletter Fälschung. Betroffen sind häufig die medizinische Forschung sowie die weitgehend zur Naturwissenschaft mutierte Psychologie. Warum aber kommen gerade in den letzten 10 bis 20 Jahren verstärkt Fälle wissenschaftlichen Fehlverhaltens an den Tag? Man könnte behaupten, die Entdeckungsquote sei durch ausgefeiltere Kontrollmethoden gestiegen, was für die Plagiatskontrolle von Qualifikationsarbeiten sicher zutrifft und für die Aufdeckung von Datenfehlinterpretationen und -fälschungen zumindest plausibel ist. Aber reicht das als Erklärung hin? Ist „immer schon“ im gleichen Umfang gemogelt, betrogen und unsauber gearbeitet worden? Sollte es

da keinen kausalen Zusammenhang mit Veränderungen in der Organisation von Forschung und Lehre geben?

Auch ohne übertriebene Idealisierung eines Status quo ante der Wissenschaft kann man aktuell den Eindruck gewinnen, als wäre der Kompass wissenschaftlichen Wohlverhaltens, das Ethos der Wissenschaft, aus der inneren Haltung der Forschenden in ein forschungsethisches Exoskelett aus externen Agenturen und Instrumenten verlagert worden: Die sich anbahnende Industrie von Replikationsstudien ist nur ein Element in einem bunten Strauß aus Akkreditierungsagenturen, Antragsrichtlinien, Ethikkommissionen und formalen Evaluationen, in denen wissenschaftliche Praxis formatiert, vermessen und – wie wir täglich leidvoll erfahren – auch bürokratisiert wird. Dafür gibt es viele und darunter auch

einige gute Gründe. Zu einem guten Teil aber werden damit Fehlentwicklungen im Wissenschaftssystem kaschiert, denen wir unsere Aufmerksamkeit zuwenden sollten:

Die deutliche Absenkung der Grundausrüstung der Universitäten und Großforschungseinrichtungen seit den 1980er Jahren zeitigt inzwischen gravierende Konsequenzen für den Wissenschaftsbetrieb. Promotionsstellen an Lehrstühlen sind mittlerweile Mangelware und nicht-professorale Dauerstellen eine Seltenheit. Unter diesen Bedingungen werden akademische Karrieren für den wissenschaftlichen „Nachwuchs“ zum Hochseilakt unter prekären Lebensverhältnissen. Gleichzeitig sind in allen Fächern Promotionskollegs aus dem Boden gestampft worden, in denen – häufig unter stark verschulden Bedingungen – Dissertationen in großer Zahl und in kurzer Zeit entstehen. Nicht nur wird die Lage mit Blick auf anschließende Forschungsstellen nicht besser, wenn immer mehr Promovierte und Habilitierte auf ein in etwa gleichbleibendes Angebot an Professuren stoßen. Oft genug ist zu beobachten, dass auch die Ausbildung schlechter wird, wenn allmählich das Unwesen der *Credit Points* auch in die Promotionsphase Einzug hält und Geschwindigkeit und Output Vorrang vor einer vertieften Beschäftigung mit dem Forschungsgegenstand und dem Aufbau

professioneller Substanz und Souveränität erhält.

Der Soziologie Richard Münch hat mit Blick auf die von Rankings ausgelösten Verheerungen im Wissenschaftsbetrieb einmal von einer „Tonnenideologie“ in der Forschung gesprochen, die Mengen an eingeworbenen Drittmitteln oder an Publikationen in internationalen Zeitschriften zum vorrangigen Nachweis wissenschaftlicher Leistung (neudeutsch: Exzellenz) machen – mit dem paradoxen Effekt, dass in vielen Berufungskommissionen die Schriften der Bewerberinnen und Bewerber gar nicht mehr gelesen, sondern nur noch zu konvertiblen Scores aufaddiert werden, deren relative Höhe dann über die Listenfähigkeit der Kandidaten und Kandidatinnen entscheidet.

Die Sache wird nicht besser mit dem Studium à la Bolognese, das seit 1999 Schritt für Schritt die deutschen Hochschulen erobert und zu einer grundlegenden Umstellung der Motiva-

tionsstrukturen unter Studierenden geführt hat: Das Interesse an der Sache wurde weithin abgelöst durch das Interesse am möglichst aufwandsarmen Erwerb von Leistungspunkten. Das mag bei Ausbildungen für den außeruniversitären Arbeitsmarkt auch Vorteile haben, für die Heranbildung von Forscherinnenpersönlichkeiten ist es jedenfalls Gift.

Sinnentleerte formale Hülle

Alles in allem zeigt sich hier ein neo-liberales Ökonomisierungsregime, das selbst in der Wirtschaft längst an seine Grenzen gestoßen ist. Wenn möglichst schnell und billig mit am Reißbrett entworfenen Kompetenzen aufgeladene Wissenschaftler auf eine Forschungsorganisation treffen, die formale Indikatoren wichtiger nimmt als inhaltliche Vertiefung, wenn die Geschwindigkeit der Output-Produktion gegenüber der Innovativität der Ergebnisse präferiert wird, dann droht das Wissenschaftssys-

tem zu einer sinnentleerten formalen Hülle zu werden. Aus Commitment für die Hervorbringung neuen Wissens unter riskanten Bedingungen wird so ein selbstläufiges System von Regelbefolgung und Kontrolle.

Was das Problem der Replikation empirischer Studien damit zu tun hat? Jede Menge, nur anders als gedacht: Replizierbarkeit ist weniger ein Problem unsauberer Datengenerierung und -analyse, noch gar dessen Lösung. Sie ist vielmehr ein Symptom für gravierende Fehlentwicklungen im (nicht nur) deutschen Wissenschaftssystem. Gute, innovative Wissenschaft braucht Zeit, Vertrauen und Fehlertoleranz – davon entfernen wir uns gerade. Wenn ein Nobelpreisträger wie Peter Higgs konstatiert, dass er unter den aktuellen Bedingungen des Wissenschaftssystems keinen akademischen Job bekommen würde, weil er zu wenig publiziere, dann ist das vielleicht doch mehr als nur Koketterie.

Anzeige

**BIELEFELDER
WISSENSCHAFTS
PREIS VERLIEHEN
DURCH DIE STIFTUNG DER
SPARKASSE BIELEFELD
IM GEDENKEN AN
NIKLAS LUHMANN**



Die Stadt Bielefeld ist mit ihren staatlichen und privaten Hochschulen ein wichtiges Zentrum für Wissenschaft und Forschung in Nordrhein-Westfalen. Die Universität Bielefeld und ihr Zentrum für interdisziplinäre Forschung prägen nachhaltig das internationale Ansehen Bielefelds als renommierter Hochschulstandort.

Der von der Stiftung der Sparkasse Bielefeld verliehene Bielefelder Wissenschaftspreis wird 2018 zum achten Mal ausgeschrieben. Der Preis wird im Gedenken an Niklas Luhmann, den großen Bielefelder Soziologen, vergeben. Er wirkte als herausragender Forscher und Lehrer seit ihrer Gründung 1969 bis zu seinem Tod 1998 an der Universität Bielefeld. Sein Werk ist der Leistungskraft der Theorie verpflichtet und erstreckt sich auf alle Aspekte der modernen Gesellschaft. Luhmanns umfassendes juristisches, historisches und philosophisches Wissen, sein breites Interesse für Lebenswissenschaften, Pädagogik, Ethik und Ästhetik und sein zentrales Anliegen, eine umfassende Systemtheorie der gesellschaftlichen

Modernisierung zu formulieren, sollen Anstoß und Verpflichtung für das Leitbild einer interdisziplinären und problemoffenen Forschung sein.

Diesem Anliegen ist auch der Bielefelder Wissenschaftspreis verpflichtet. Mit ihm sollen deshalb solche Personen ausgezeichnet werden, deren Forschung, insbesondere im Bereich der genannten Fachrichtungen, höchsten Ansprüchen genügt. Mit dieser Auszeichnung wird um begründete Vorschläge für die Auszeichnung aus dem In- und Ausland gebeten.

Der Bielefelder Wissenschaftspreis, verliehen durch die Stiftung der Sparkasse Bielefeld im Gedenken an Niklas Luhmann, ist mit einem Preisgeld von **25.000 Euro** dotiert und wird alle zwei Jahre vergeben.

Bitte richten Sie Ihren Vorschlag **bis spätestens 31.03.2018** an die

**Stiftung der Sparkasse Bielefeld,
Bielefelder Wissenschaftspreis,
Prof. Dr.-Ing. Gerhard Sagerer, Vorsitzender der Jury
Schweriner Straße 5, 33605 Bielefeld
Telefon +49 (0) 521 294-1053**



**Stiftung der
Sparkasse Bielefeld**