

Motivationsheterogenität in Laborexperimenten mit Studierenden

Problemspezifizierung und Auswege

Der Beitrag von Alexander Lenger und Stephan Wolf (2018) zeigt, dass die Motivation von Studierenden, an sozialwissenschaftlichen Laborexperimenten teilzunehmen, heterogen ist. Das ist wenig überraschend, musste so aber auch erst einmal gezeigt werden. Darüber hinaus entwickeln die Autoren aus qualitativen Interviews – eine begrüßenswerte Methodentriangulation – eine Typologie von Experimentalt Teilnehmern¹ bezüglich ihrer Motivation und unterscheiden den strategisch-nutzenmaximierenden Typus, den ‚Interessierten Forscher‘ und den ‚Kritischen Experten‘. Zu einem korrekturbedürftigen Problem für den generalisierenden Einsatz von Laborexperimenten mit Studierenden wird dieser Befund allerdings nur unter bestimmten Umständen, die im Beitrag nicht oder nicht deutlich genug herausgearbeitet werden.

1 Verhaltensunterschiede zwischen Typen und Verallgemeinerbarkeit der Typenverteilung

Erstens muss man annehmen, dass es Unterschiede im Verhalten im Experiment zwischen den Teilnehmern je nach Motivationslage gibt. Das mag naheliegend sein, vor allem bei ökonomischen Entscheidungssituationen, in denen oft eigeninteressierte gegenüber prosozialen Zielen abgewogen werden müssen. Es ist anzunehmen, aber nicht nachgewiesen, dass ‚Interessierte Forscher‘ und ‚Kritische Experten‘ stärker zu prosozialen Optionen tendieren. Daher ist Vorsicht bei der Verallgemeinerung durchschnittlichen Verhaltens geboten, doch das ist – wenngleich gängige Praxis – ohnehin nicht die Idee eines Experiments.

Zweitens ist die Frage, ob die Verhaltensunterschiede zwischen den Typen nur die Verhaltensraten betreffen – beispielsweise das Ausmaß kooperativen Verhaltens – oder ob auch die Größe und möglicherweise sogar die Richtung der Effekte randomisierter Treatments (Kausaleffekte) sich je nach Typ unterscheiden. Da dank der zufälligen Zuordnung der Teilnehmer zu Treatment- und Kontrollgruppe – ein wesentliches Merkmal jedes Experiments, das Lenger und Wolf in ihrer Beschreibung vernachlässigen – die gleiche Anzahl an Personen jeden Motivationstyps zu erwarten ist, ist die interne Validität trotzdem weiterhin gegeben.

* Felix Bader, Universität Mannheim, Fakultät für Sozialwissenschaften, A5, 6, 68159 Mannheim, E-Mail: bader@uni-mannheim.de

1 Aus Gründen der Lesbarkeit wird stets die männliche Form verwendet, es sind jedoch Personen jeglichen Geschlechts gemeint.

Problematisch für die Generalisierung experimenteller Befunde wird die Motivationsheterogenität erst, wenn – je nach inhaltlichem Generalisierungsziel neben den Verhaltensunterschieden oder den heterogenen Treatmenteffekten – die externe Validität selbst eingeschränkt ist. Das ist nicht selbstverständlich, sondern nur der Fall, wenn die Verteilung der Typen im Experiment anders ist als in der Grundgesamtheit. Wenn in der Allgemeinbevölkerung bestimmte Motivationstypen seltener oder häufiger zu finden sind als in der Gruppe der Studierenden, die am Experiment teilgenommen haben, wäre die „Transportierbarkeit“ (Pearl/Bareinboim 2014) der experimentellen Befunde gering. Außerdem könnte die externe Validität eingeschränkt sein, wenn sich das Verhalten in der im Experiment simulierten Situation je nach Motivationstyp unterschiedlich gut auf das Verhalten außerhalb des Labors übertragen ließe. Es könnte sein, dass sich beispielsweise das Verhalten von Personen, die sich im Experiment entsprechend des Typs ‚Interessierte Forscher‘ darstellen, weniger gut auf deren Verhalten außerhalb des Labors übertragen ließe, als das Verhalten von ‚Kritischen Experten‘.

Wenn je nach Fragestellung zum einen die Motivationstypen Verhaltensunterschiede oder heterogene Treatmenteffekte aufweisen und zum anderen die Verteilung der Typen in der Grundgesamtheit anders oder die Übertragbarkeit des Verhaltens heterogen ist, lassen sich die entsprechenden Laborergebnisse nicht verallgemeinern. Es erscheint plausibel, dass ein Teil dieser Annahmen oder sogar alle diese Annahmen zutreffen. Jedoch müsste dies in größer und gezielt angelegten Folgestudien empirisch gezeigt werden, um die Relevanz der Typologie zu belegen. Zunächst müssten – wie Lenger und Wolf fordern – standardisierte Erhebungsinstrumente entwickelt werden, die es erlauben, die Teilnahmemotivation von einer großen Zahl von Experimentalteilnehmern mit heterogenem soziodemographischen Hintergrund bzw. von einer Zufallsstichprobe aus der Bevölkerung zu erheben. Dazu benötigt man experimentelle Methoden, die sich in nichtstudentischen Bevölkerungsstichproben anwenden lassen.

2 Experimente in Umfrageform und Online-Experimente als Alternativen

Der Beitrag von Lenger und Wolf behandelt Laborexperimente mit Studierenden als wären sie, wenn man experimentelle Methoden einsetzen möchte, aus Kostengründen alternativlos. Im zweiten Teil des Kommentars sollen im Gegensatz dazu Alternativen diskutiert werden.

Die Kosten von Laborexperimenten mit Studierenden sind nicht unerheblich. Es müssen vernetzte Computer angeschafft und unterhalten werden, geeignete Räume ausgestattet und bereitgestellt, Personal für die Anwerbung von Teilnehmern, Anmeldung, Durchführung und Auszahlung des Experiments beschäftigt werden

und die Auszahlungen an die Teilnehmer selbst müssen finanziert werden.² Um die Teilnahme am Experiment für Studierende interessant zu machen und dafür zu sorgen, dass die Entscheidungen im Experiment ernst genommen werden, muss die Auszahlung hoch genug sein, dass sie pro Stunde etwa dem erwarteten Lohn verfügbarer Erwerbstätigkeit entspricht. Gleichzeitig muss die Auszahlung pro Session hoch genug sein, dass sich der Aufwand zum Labor zu kommen für die Teilnehmer lohnt. Studierende sind recht günstige Teilnehmer, da sie bei geringem Durchschnittseinkommen einen niedrigen Reservationslohn haben und sich ohnehin oft an der Universität und damit in der Nähe des Labors aufhalten.

Um die Verallgemeinerbarkeit von Ergebnissen aus Experimenten mit Studierenden zu untersuchen, werden teilweise gezielt Personen als Teilnehmer gewählt, bei denen aufgrund gänzlich anderer Alltagserfahrungen davon ausgegangen werden kann, dass sie sich anders als Studierende verhalten. Neben Verhaltensexperimente mit indigenen oder anderen Personen aus nicht-westlichen Kulturen (Henrich 2001) gibt es solche mit Tieren, Kindern, älteren Menschen und speziellen Berufsgruppen wie Investmentbankern (für einen Überblick s. Fréchette 2016). Andersen et al. (2010) laden sogar zufällig aus den Registerdaten gezogene Dänen in ein Experiment ein, das in Besprechungsräumen von Hotels in verschiedenen Orten in Dänemark durchgeführt wurde. Bei solchen Experimenten müssen, da ist Lenger und Wolf zuzustimmen, deutlich höhere Auszahlungen vorgenommen und/oder höherer organisatorischer Aufwand betrieben werden.

Für viele Experimente ist es hingegen gar nicht erforderlich, dass sich die Teilnehmer physisch in einem Raum befinden. Interviewer könnten mit Tablets zu den Teilnehmern nach Hause kommen und diese über das Internet interagieren lassen. Die Umgebung während der Teilnahme wäre dann zwar heterogener, die Anonymität unter den Teilnehmern jedoch höher. Für einfache Entscheidungssituationen, zu denen neben Diktatorspiel und Ultimatumspiel unter anderem das Vertrauensspiel, Gefangenendilemma und einfache Public Goods Games gehören, ist es nicht einmal erforderlich, dass die Teilnehmer sich in Echtzeit miteinander austauschen. Die Entscheidungen der Teilnehmer können wie in Umfragen persönlich, telefonisch oder postalisch an geeigneten Bevölkerungsstichproben erhoben werden, um sie anschließend mit den Entscheidungen der Gegenspieler zusammenzufügen, die Auszahlung zu berechnen und den Teilnehmern zukommen zu lassen. Im Ultimatumspiel und dem Vertrauensspiel, in dem die Teilnehmer unterschiedliche Rollen haben und es einen zweiten Akteur gibt, der auf die Entscheidung des ersten reagiert sowie anderen einfachen interaktiven Entscheidungssituationen, lässt sich die Strategiemethode (Rauhut/Winter 2010) anwenden. Dabei geben die zweiten Akteure eine Antwort für jede mögliche Entscheidung des ersten Akteurs an. Je

2 Die Kosten für die Entwicklung der Software können wir bei diesem Vergleich vernachlässigen, da vergleichbarer Aufwand bei anderen Formen von Experimenten ebenfalls getrieben werden muss.

nach tatsächlicher Entscheidung des ersten Akteurs kommt die komplementäre Entscheidung des zweiten Akteurs zur Anwendung. Das ermöglicht nicht nur eine zeitversetzte Teilnahme, sondern erhöht auch die vom zweiten Akteur erhobene Information massiv. Die Durchführung eines solchen Experiments in Umfrageform bringt natürlich Kosten mit sich, ist jedoch zumindest bei lokalen Stichproben oder postalischer Erhebung nicht unbedingt teurer als die entsprechende Zahl Studierender im Labor teilnehmen zu lassen. Um nichtstudentische Zielpersonen zur Teilnahme zu motivieren sind vielleicht höhere Geldbeträge erforderlich, die Auszahlung muss jedoch nicht für den Weg zum Labor entschädigen, da die Teilnahme ja zu Hause erfolgen kann. Auf diese Weise führten beispielsweise Falk und Zehnder (2013) ein Experiment mit dem Vertrauensspiel an einer Bevölkerungsstichprobe der Stadt Zürich durch.

Fehr et al. (2002) führten ein Vertrauensspiel-Experiment mit einer deutschlandweiten Stichprobe in Kooperation mit dem SOEP durch. Ein großer Vorteil dieses Vorgehens ist, dass bei der Analyse Daten aus der Umfrage zugespielt werden können, um möglicherweise die Entscheidungen im Experiment erklären zu können. Es ist so nicht immer erforderlich, eine eigene Erhebung zu starten. Experimentelle Module mit Entscheidungssituationen können auch in laufende Umfragen integriert werden.

Eine weitere Alternative zu studentischen Laborexperimenten bieten Online-Experimente. Auch hier ist die Kontrolle über die Umgebung während der Teilnahme nicht wie im Labor gegeben und die Distanz zwischen den Teilnehmern größer, aber bisherige Studien finden kaum Anzeichen dafür, dass Teilnehmer sich in Online-Experimenten systematisch anders als in Laborexperimenten verhalten (Hergueux/Jacquemet 2015, Bader et al. 2018). Prinzipiell sind in Online-Experimenten computergestützte Echtzeit-Interaktionen wie in Laborexperimenten möglich. Das wurde auch in einigen Studien mit Online-Warteräumen umgesetzt (z.B. Suri/Watts 2011, Biliotti et al. 2016, Anderson et al. 2017). Allerdings kommt es aufgrund von Teilnahmeabbruch und Entscheidungsverzögerung, die bei Online-Interaktionen im Gegensatz zu physischer Laborpräsenz geringere psychische Kosten verursachen, zu Wartezeit, die wiederum die Wahrscheinlichkeit (weiterer) Teilnahmeabbrüche erhöht. Daher werden Online-Experimente mit einfachen Entscheidungssituationen meist nicht mit Echtzeit-Interaktion, sondern wie oben beschrieben mit nachträglichem Zusammenfügen der ggf. in Strategietabellen erhobenen Entscheidungen durchgeführt. Gegenüber Paper-and-Pencil-Experimenten im Labor oder in Umfrageform bieten computergestützte und damit auch Online-Experimente eine Reihe von Vorteilen. So sind Zufallsaufteilung und die Darstellung von Informationen flexibler, es lässt sich messen oder steuern, welche Informationen wie lange angezeigt werden, und die Software kann in Echtzeit auf Eingaben reagieren und gezielt Feedback geben.

Teilnehmer für Online-Experimente lassen sich auf sogenannten Crowdsourcing-Plattformen finden. Die größte dieser Plattformen ist Amazon Mechanical Turk. Hier können sich Interessierte als ‚Worker‘ registrieren und von ‚Requestern‘ eingestellte Arbeitsaufträge gegen Bezahlung erledigen. Typische Arbeitsaufträge sind die Kategorisierung von Bild- oder Videodokumenten oder die Transkription von Audiodokumenten, die nicht vollständig automatisiert erfolgen können. Daneben gibt es aber auch Aufträge aus der Marktforschung und der Wissenschaft, insbesondere der Psychologie und teilweise den Wirtschaftswissenschaften, vereinzelt auch aus der Soziologie (Keuschnigg et al. 2016, Bader et al. 2018).³ Die Kosten für die Gewinnung von Teilnehmern auf Mechanical Turk sind sehr gering, da dort üblicherweise sehr geringe Löhne gezahlt werden, so dass schon für die Einhaltung des US-Mindestlohns appelliert werden musste (Williamson 2016). Der Hauptvorteil für das vorliegende Thema ist jedoch, dass hier internationale, nichtstudentische Teilnehmer gewonnen werden können, um experimentelle Befunde auf breiterer Datenbasis replizieren zu können. Diese sind im Gegensatz zu studentischen Teilnehmern in erster Linie monetär motiviert, da es sich bei Mechanical Turk um einen Arbeitsmarkt und nicht um einen Teilnehmerpool handelt (Horton et al. 2011).

Neben Crowdsourcing erlauben online durchgeführte Experimente auch andere Personengruppen zu erreichen, da mittlerweile über 80% der Bevölkerung Deutschlands das Internet nutzen (Initiative D21 2018). Hervorzuheben sind vor allem Online Access Panels. Hier sind die Teilnehmer allerdings wieder Personen, die freiwillig an Studien teilnehmen, also vermutlich nichtmonetär und heterogen motiviert sind. Sie können jedoch wie bei forsa mit Hilfe von Zufallsstichproben aus der (mit dem Internet verbundenen) Bevölkerung rekrutiert werden und so einen weitgehend ‚repräsentativen‘ Querschnitt der Bevölkerung darstellen. Die Kosten für die Entschädigung der Teilnehmer sind relativ gering, da sie grundsätzlich bereits zur Teilnahme an Studien motiviert sind und nicht zum Labor reisen müssen, aber die Gebühr für das Institut, welches das Panel zur Verfügung stellt, ist mit einzuberechnen. Ein großer Vorteil ist die Qualität der Stichprobe, die wenn auch nicht rein zufällig und unverzerrt, so doch nach vielfältigen Kriterien zusammengestellt werden kann. Dieses Potential wird schon vielfach für Umfragen (z.B. durch das Online-Modul der German Longitudinal Election Study), jedoch bisher kaum für Experimente genutzt. Eine Ausnahme ist Bader et al. (2018).

Insgesamt möchte dieser Kommentar dazu aufrufen, experimentelle Möglichkeiten jenseits des studentischen Laborexperiments auszuschöpfen. Je nach Fragestellung lassen sich oft Alternativen zum Laborexperiment finden, mit denen sich heterogene Teilnehmer aus anderen Teilen der Gesellschaft erreichen lassen, die jeweils ihre Alltagserfahrungen mit ins Experiment bringen. Mit diesen Methoden lässt

3 Bader und Keuschnigg (2018) geben genauere Empfehlungen zur Umsetzung von Experimenten auf Mechanical Turk.

sich klären, ob und inwieweit Ergebnisse von Laborexperimenten mit Studierenden verallgemeinert werden können. Möglicherweise stellt sich dabei heraus, dass manche oder sogar alle Ergebnisse aus Experimenten mit Studierenden verallgemeinert werden können. Strategische Entscheidungen sind vermutlich weniger stark verzerrt als moralische und die Richtungen von Treatmenteffekten eher zu verallgemeinern als die genauen Erwartungswerte des Verhaltens, wie unsere Arbeit nahelegt (Bader et al. 2018). Gegebenenfalls lassen sich, wenn die Verteilungen genau genug bekannt sind, systematische Verzerrungen korrigieren, um Ergebnisse doch generalisieren zu können. Lenger und Wolf nennen an dieser Stelle Gewichtung. Doch selbst wenn sich zeigt, dass sich die Ergebnisse zwischen Experimenten mit Studierenden und mit anderen Teilen der Bevölkerung zu stark in nicht prognostizierbarer Weise unterscheiden, brächte die Erschließung experimenteller Alternativen, die sich breiter anwenden lassen, eine wesentliche Erweiterung des Methodenspektrums um Experimente, die der Kritik eingeschränkter externer Validität weniger stark ausgesetzt sind.

Literaturverzeichnis

- Andersen, Steffen / Harrison, Glenn W. / Lauc, Morten I. / Rutström, E. Elisabet (2010): Preference Heterogeneity in Experiments: Comparing the Field and Laboratory. *Journal of Economic Behavior & Organization* 73, S. 209-224.
- Anderson, Brilé / Bernauer, Thomas / Ballelli, Stefano (2017): Effects of Fairness Principles on Willingness to Pay for Climate Change Mitigation. *Climatic Change* 142, S. 447-461.
- Bader, Felix / Baumeister, Bastian / Berger, Roger / Keuschnigg, Marc (2018): On the Transportability of Laboratory Results. Unveröffentlichtes Manuskript, Institute for Analytical Sociology, Schweden.
- Bader, Felix / Keuschnigg, Marc (2018): Conducting Large-Scale Online Experiments on a Crowdsourcing Platform. *SAGE Research Methods Cases*.
- Ballelli, Stefano / Goldstone, Robert L. / Helbing Dirk (2016): Peer Review and Competition in the Art Exhibition Game. *Proceedings of the National Academy of Sciences* 113, S. 8414-8419.
- Falk, Armin / Zehnder, Christian (2013): A City-Wide Experiment on Trust Discrimination. *Journal of Public Economics* 100, S. 15-27.
- Fehr, Ernst / Fischbacher, Urs / von Rosenblatt, Bernhard / Schupp, Jürgen und Wagner, Gert G. (2002): A Nation-Wide Laboratory: Examining Trust and Trustworthiness by Integrating Behavioral Experiments into Representative Surveys. *Schmollers Jahrbuch* 122, S. 519-542.
- Fréchette, Guillaume R. (2016): Experimental Economics Across Subject Populations, in: John H. Kagel und Alvin E. Roth (Hg.): *The Handbook of Experimental Economics*, Vol. 2. Princeton, NJ, S. 435-480.
- Henrich, Joseph / Boyd, Richard / Bowles, Samuel / Camerer, Colin F. / Fehr, Ernst / Gintis, Herbert / McElreath, Richard (2001): In Search of Homo Economicus: Behavioral Experiments in 15 Small-Scale Societies. *American Economic Review* 91, S. 73-78.
- Hergueux, Jérôme / Jacquemet, Nicolas (2015): Social Preferences in the Online Laboratory: A Randomized Experiment. *Experimental Economics* 18, S. 251-283.

- Horton, John J. / Rand, David G. / Zeckhauser, Richard J. (2011): The Online Laboratory: Conducting Experiments in a Real Labor Market, in: *Experimental Economics* 14, S. 399-425.
- Initiative D21 (2018): D21-Digital-Index 2017/2018, Berlin. Online verfügbar unter https://initiated21.de/app/uploads/2018/01/d21-digital-index_2017_2018.pdf [23.02.2018].
- Keuschnigg, Marc / Bader, Felix / Bracher, Johannes (2016): Using Crowdsourced Online Experiments to Study Context-Dependency of Behavior. *Social Science Research* 59, S. 68-82.
- Lenger, Alexander und Stephan Wolf (2018): Experimente in der Soziologie? Über die systematische Verzerrung von Experimentergebnissen aufgrund strategisch agierender studentischer Teilnehmertypen. *Soziale Welt* 1/18.
- Pearl, Judea / Bareinboim, Elias (2014): External Validity: From Do-Calculus to Transportability Across Populations. *Statistical Science* 29, S. 579-595.
- Rauhut, Heiko / Winter, Fabian (2010): A Sociological Perspective Measuring Social Norms by Means of Strategy Method Experiments. *Social Science Research* 39, S. 1181-1194.
- Suri, Siddharth / Watts, Duncan J. (2011): Cooperation and Contagion in Web-Based, Networked Public Goods Experiments. *PLoS ONE* 6, e16836.
- Williamson, Vanessa (2016): On the Ethics of Crowdsourced Research. *Political Science & Politics* 49, S. 77-81.