

# Pulp Science? Zur Berichterstattung über Meinungsforschung in den Massenmedien

Ulrich Kohler und Julia C. Post

## Prolog

„Umfrage: Große Ablehnung für Methoden der Letzten Generation“ (Zeit Online, 13.05.2023). „Gewalt an Frauen laut Drittel junger Männer ‚akzeptabel‘“ (ZDF, 12.06.2023). „Mehr als 40 Prozent befürworten Öffnung der CDU zur AfD in Kommunen“ (Spiegel Online, 27.07.2023). Diese drei Überschriften aus überregional verbreiteten Qualitätsmedien haben eine Gemeinsamkeit: Sie alle basieren auf Ergebnissen von Online-Umfragen, deren Befragte sich dadurch auszeichnen, dass sie sich freiwillig regelmäßig befragen lassen (sog. Online-Access-Panel). Eine der drei Schlagzeilen rief jedoch Kritik hervor. Die vorgebrachten Einwände bezogen sich jedoch nicht auf spezifische Eigenheiten dieser Studie, sondern betrafen auch die anderen Beispiele und mit ihnen alle Umfragen, die an solchen Online-Access-Panels durchgeführt werden. Entsprechend bemerkenswert ist es, dass sich die öffentliche Aufmerksamkeit an der methodischen Kritik gegen Online-Access-Panel nur an dieser Umfrage entzündet hat.

Das Problem der Verbreitung zweifelhafter Umfrageergebnisse ist grundsätzlicher als es die punktuelle Kritik an der Umfrage zur Akzeptanz von Gewalt gegenüber Frauen erscheinen lässt. In diesem Aufsatz vertreten wir die These, dass eine Kombination aus methodischen Eigenschaften von Umfragen an Online-Access-Panels und der Orientierung von Medien am „Nachrichtenwert“ zu einem Zerrbild der öffentlichen Meinung führt. Die Begründung unserer These erfolgt in vier Schritten. Im ersten Schritt betrachten wir was Umfragen an Online-Access-Panels für die Medien attraktiv macht. Danach zeigen wir, dass die Verzerrungen solcher Umfragen größer sind, als oft erwartet. Drittens begründen wir, warum die Qualität von Umfragen aus



**Prof. Dr. Ulrich Kohler**

Inhaber des Lehrstuhls Methoden der empirischen Sozialforschung an der Universität Potsdam



**Julia C. Post**

Soziologin und wissenschaftliche Mitarbeiterin am Lehrstuhl Methoden der empirischen Sozialforschung der Universität Potsdam

Online-Access-Panels gelegentlich überschätzt wird. Abschließend fassen wir zusammen und geben Handlungsempfehlungen.

## Online-Access-Panel und die Medien

Ergebnisse von Umfragen sind fehleranfällig – unabhängig von Design und Größe der Stichprobe, der Art der Datenerhebung oder dem Unternehmen, das die Umfrage durchführt. Das gilt für „Meinungsbilder“, die sich aus Statements willkürlich ausgewählter Personen zusammensetzen, ebenso wie für groß angelegte Befragungen von zufällig ausgewählten Personen oder Umfragen an Online-Access-Panels. Genauer: Ein in einer Umfrage beobachteter Anteilswert, ist eine Mischung aus dem wahren Wert in der Bevölkerung und einer Vielzahl von anderen Faktoren, dem sog. Fehler:

$$\text{Beobachteter Anteilswert} = \text{Wahrer Anteilswert} + \text{Fehler}$$

Auf Basis dieser Einsicht unterscheiden Methodenforscher:innen Umfrageverfahren anhand von zwei Kriterien. *Erstens*: Sind die Fehler unsystematisch, so dass sie sich, wenn man die Umfrage unendlich oft wiederholen würde, gegenseitig ausgleichen? *Zweitens*: Sind die Fehler, die sich bei der unendlichen Wiederholung der Umfrage ergeben würden, im Schnitt eher klein oder eher groß? Diejenigen Verfahren, bei denen sich die Fehler ausgleichen, heißen „unverzerrt“, die anderen „verzerrt“. Diejenigen Verfahren, die im Durchschnitt kleine Fehler haben, heißen „effizient“, die anderen „ineffizient“. Das Wort „repräsentativ“, das für Umfragen oft bemüht wird, ist dagegen bedeutungslos (Schnell 2019a: 163).

Um ein möglichst realistisches und verlässliches Bild zu zeichnen, sollten Qualitätsmedien nur über Ergebnisse von Umfragen berichten, die mit unverzerrten und effizienten Verfahren erhoben werden.<sup>1</sup> In der Praxis lässt sich das jedoch nicht einfach umsetzen. Einerseits bräuchte es hierfür eine detaillierte Auseinandersetzung mit allen Aspekten der jeweiligen Umfragemethode, wofür entsprechende Expertise notwendig wäre. Andererseits spielen in der Praxis auch Faktoren wie Umsetzbarkeit, Zeit und Finanzierbarkeit eine Rolle, so dass Kompromisse unumgänglich werden.

Solche Kompromisse müssen alle finden, die Umfrageforschung betreiben. In der Medienberichterstattung greift jedoch ein Mechanismus, der gerade verzerrte Ergebnisse besonders attraktiv erscheinen lässt: Die Orientierung am sog. „Nachrichtenwert“ (Lippmann 1922, Warren 1934, Galtung und Ruge 1965, Schulz 1976).

Unter sonst gleichbleibenden Bedingungen ist der Nachrichtenwert einer Meldung umso höher, je überraschender, aktueller und neuer sie ist. Wir dürfen annehmen, dass Umfrageergebnisse als *überraschend* eingeschätzt werden, wenn sie von den Erwartungen einer Redaktion abweichen. Gleiches dürfte für die Wahrnehmung von Umfrageergebnissen seitens der Rezipient:innen zutreffen. Sollten diese Erwartungen in irgendeiner Weise durch den wahren Wert geprägt sein, sind Umfrageergebnisse umso überraschender, je größer ihre Fehler sind. Ergebnisse von verzerrten Umfrageverfahren

ren hätten dann einen tendenziell höheren Nachrichtenwert als Ergebnisse von unverzerrten Verfahren.

Eine Umfrage kann als *aktuell* bezeichnet werden, wenn in ihr unmittelbar auf aktuelle Ereignisse Bezug genommen wird. Das geht natürlich eher mit Umfragen, die schnell zu realisieren sind. Sozialwissenschaftliche Umfragen benötigen ein Jahr oder länger bis sie ins Feld gehen. Daran schließt sich eine längere Feldphase und quälende Monate zur Datenaufbereitung und Fehlerkontrolle an. Für Medien sind solche Umfragen wertlos. Aber auch wenn man Kompromisse macht, ist eine Umfrage zu einem aktuellen Thema, bei der zunächst eine Zufallsstichprobe realisiert werden muss, nicht innerhalb weniger Tage möglich.

*Neu* sind Ergebnisse von Umfragen zu Themen von gesellschaftlicher Relevanz, über die zuvor noch nicht gesprochen wurde. Wenn die Kosten von Umfragen überschaubar bleiben, kann eine Redaktion gezielt eine Umfrage lancieren, um das Thema, über das sie berichtet, selbst zu konstruieren oder anzureichern.

Die genannten Aspekte machen Online-Access-Panel für Medien zu einem attraktiven Angebot. Die Ergebnisse dieser Umfragen sind kostengünstig und stehen schnell zur Verfügung. Zwar sind sie verzerrter als Befragungen von Personen, die mit einem Zufallsverfahren ausgewählt wurden (Cornesse et al. 2020, Kohler und Post 2023), jedoch sind sie gerade deshalb anfälliger für überraschende und daher berichtenswerte Ergebnisse. Damit wollen wir nicht sagen, dass Medien gezielt verzerrte Ergebnisse verbreiten. Doch sofern in einer Redaktion die Expertise zur Beurteilung der Qualität einer Umfrage fehlt, kann die Qualität der Umfrage nicht als Auswahlkriterium fungieren. Entsprechend dürfte in diesem Fall der Nachrichtenwert die Entscheidung über die Verbreitung der Umfrageergebnisse dominieren. Unter diesen Bedingungen kommt es schnell zu einer unheiligen Allianz zwischen fehleranfälligen Umfragen und den Anreizen des Nachrichtenwertes, die allzu leicht ein Zerrbild des Meinungsbildes der Bevölkerung entstehen lässt.

## Die goldene Ausrede

Werden Umfragen aus Online-Access-Paneln aufgrund ihrer methodischen Unterlegenheit kritisiert, liest und hört man oft: „Die Studie zeigt dennoch eine Tendenz“. Vermutlich ist damit gemeint, dass sich der in der Umfrage erzielte Wert trotz der methodischen Mängel nicht so sehr vom wahren Wert unterscheidet.

Leider kann man nicht wissen, wie weit der wahre Wert und der in der Umfrage erzielte Wert auseinanderliegen. Das ginge nur, wenn der wahre Wert bekannt wäre – aber dann bräuchte man keine Umfrage. Man kann jedoch mathematisch herleiten, wie weit ein mit einem festgelegten Verfahren erzielter Wert im Schnitt über unendliche Wiederholungen des Verfahrens vom wahren Wert abweicht. Diese durchschnittliche Abweichung nennt man *Verzerrung*.

Wir wollen hier ausschließlich diejenige Verzerrung betrachten, die aus der Auswahl der Befragten einer Umfrage erwächst. Diese Verzerrung hängt von drei Faktoren ab, die multiplikativ zusammenwirken<sup>2</sup>:

$$\text{Verzerrung} = a \cdot b \cdot c$$

Bevor wir auf die drei Faktoren im Einzelnen eingehen, muss etwas zum multiplikativen Zusammenwirken gesagt werden. Zunächst: Wenn einer der Faktoren Null ist, gibt es keine Verzerrung. Und: Wenn keiner der Faktoren Null ist, so wirkt sich eine Veränderung eines der drei Faktoren umso stärker auf die Verzerrung aus, je höher das Produkt der beiden anderen Faktoren ist. Wenn  $a \cdot b$  groß ist, führen schon kleine Erhöhungen von  $c$  zu einer stärkeren Verzerrung.

Als Beispiel verwenden wir im Folgenden die eingangs erwähnte Umfrage zur Akzeptanz der Gewalt gegenüber Frauen, da gerade diese Studie die Aufmerksamkeit auf die Problematik von Online-Access-Panels auch außerhalb der Methoden-Community gelenkt hat. Wir möchten jedoch betonen, dass es uns dabei nicht darum geht, die Korrektheit dieses spezifischen Ergebnisses zu beurteilen – wir sind keine Experten zum Thema der Umfrage und kennen den wahren Wert nicht. Vielmehr geht es uns darum, das Bewusstsein für die Stärke möglicher Verzerrungen zu schärfen.

Der erste Faktor zur Bestimmung der Verzerrung,  $a$ , ist die Heterogenität des Merkmals, das man untersuchen möchte.<sup>3</sup> Wenn alle Männer in Deutschland Gewalt ablehnen würden, wäre  $a = 0$ . Wenn der berichtete Wert von 33% korrekt wäre, wäre

$$a = \sqrt{0.33 \cdot (1 - 0.33)} = 0.47.$$

Betrüge der Wert in Wirklichkeit dagegen nur 5% wäre

$$a = \sqrt{0.05 \cdot (1 - 0.05)} = 0.22.$$

Dieses Zahlenbeispiel illustriert den allgemeinen Punkt, dass die Gefahr einer Verzerrung am größten ist, wenn die wahren Prozentanteile bei 50% liegen, während die Gefahr umso kleiner wird, je weiter sich die wahren Prozentanteile von 50% entfernen.

Kommen wir zu  $b$ . Bei einem Online-Access-Panel ist dies die *relative* Heterogenität<sup>4</sup> der Wahrscheinlichkeit, Mitglied im Online-Access-Panel zu sein und befragt zu werden. Auch hier gilt: Wenn alle Menschen in Deutschland die gleiche Wahrscheinlichkeit hätten, an einer konkreten Umfrage teilzunehmen, wäre  $b = 0$ . Da die Panelmitglieder aber durch Medienkampagnen und persönliche Kontakte geworben werden, kann man dies getrost ausschließen. Leider lässt sich der Wert von  $b$  nicht bestimmen. Bekannt ist aber die Formel dafür, welchen Wert  $b$  schlimmstenfalls annehmen könnte. Mit diesem Wert lässt sich dann eine *potentielle* Verzerrung bestimmen. Hierfür benötigt man

- die Teilnahmerate, d.h. den Anteil der Mitglieder eines Online-Access-Panels am Umfang der Grundgesamtheit sowie
- die Antwortrate, d.h. den Anteil der antwortenden Personen an den zur Umfrage eingeladenen Panelmitgliedern.

In unserem Zahlenbeispiel gehen wir von der wahlberechtigten männlichen Bevölkerung in Deutschland aus, das sind ca. 30 Millionen Personen. Die Firma, die die besagte Umfrage durchgeführt hat, wirbt mit einem Access-Panel, das weltweit ca. 900.000 Personen umfasst. Wie viele davon in Deutschland lebende männliche Wahlberechtigte sind, ist nicht bekannt. Vergleichbare Anbieter werben mit Access-Paneln, die zwischen einer Million (Civey) und 90.000 (Norstat) in Deutschland lebende Personen umfassen.

Ebenfalls unbekannt ist die Antwortrate. Die Firma schreibt, 25% der Panelmitglieder würden nach 5 Tagen antworten. Diese Angabe hilft uns jedoch nicht weiter, daher errechnen wir den Faktor **b** hier *zugunsten* des Instituts mit 250.000 wahlberechtigten männlichen Panelmitgliedern und einer Antwortrate von 80%:

$$b = \sqrt{(1/(250000/30000000 \cdot 0.8) - 1)} = 12.21.$$

Zum Vergleich: Bei einer traditionellen Zufallsstichprobe mit ca. 1.000 Befragten und einer Antwortrate von nur 5%, liegt **b** bei 4.36 und damit deutlich niedriger als beim Online-Access-Panel.

Faktor **c** beschreibt den Zusammenhang zwischen der Teilnahmewahrscheinlichkeit und dem Merkmal, für das wir uns interessieren. Spätestens hier müssen wir uns ins Reich der Spekulation begeben. Für das vorliegende Beispiel schlagen wir ein Gedankenexperiment vor: Überlegen Sie einmal, warum sich manche Menschen freiwillig dazu bereit erklären, regelmäßig gegen Bezahlung an Umfragen teilzunehmen, andere aber nicht und warum dann einige davon an dieser spezifischen Umfrage teilnehmen. Notieren Sie alle Variablen, die Ihnen als mögliche Gründe einfallen auf einem Zettel. Streichen Sie dann alle Variablen von Ihrem Zettel, bei denen Sie *sicher* sind, dass sie *wirklich gar nichts* mit der Akzeptanz von Gewalt gegenüber Frauen zu tun haben. Dann streichen Sie noch die Variablen Geschlecht, Bildung und geographische Region (Nord, Ost, Süd, West) von Ihrem Zettel, denn darum hat sich das Institut bereits gekümmert. Und jetzt überlegen Sie, wie gut Sie die Gewaltakzeptanz und die Teilnahmewahrscheinlichkeit einer Person vorhersagen könnten, wenn Sie zu dieser Person alle Informationen hätten, die auf Ihrem Zettel stehen. Je besser diese Vorhersage wäre, desto höher ist **c**.

Auf unserem Zettel stehen: Alter, Einkommen, Rollenvorstellungen, Traditionalität, Haushaltsgröße, Familienstand, kultureller Hintergrund, Mediennutzungsverhalten, Beruf, Erwerbsstatus, Lebenszufriedenheit, Ortsgröße, Konfessionszugehörigkeit, Religiosität, soziales Engagement, Mitteilungsbedürftigkeit und Parteipräferenz.

Wir gehen jede Wette ein: Würden wir mit all diesen Variablen die Teilnahmewahrscheinlichkeit und die Akzeptanz von Gewalt gegenüber Frauen vorhersagen, wäre der  $R^2$ -Wert der dazu herangezogenen Regressionsmodelle im Bereich zwischen 10 und 20%. Welche  $R^2$ -Werte Sie auch immer für Ihre Liste annehmen, Sie können sich dem Faktor **c** recht gut annähern, indem Sie die beiden Werte miteinander multi-

plizieren und dann die Wurzel ziehen.<sup>5</sup> Wir machen das einmal mit den Werten 11 und 13%:

$$c = \sqrt{0.11 \cdot 0.13} = 0.1196.$$

Lassen Sie uns nun schauen, wie groß das *Potential* für die Verzerrung wäre, wenn die Gewaltakzeptanz in Wirklichkeit deutlich niedriger und nur bei 5% läge, nicht bei 33% wie von den Medien berichtet. Dazu müssen wir nur die errechneten Beispielergebnisse von **a**, **b** und **c** in die Formel für die Verzerrung einsetzen:

$$\text{Potentielle Verzerrung} = 0.22 \cdot 12.21 \cdot 0.1196 = 0.32$$

Das heißt: Unter den hier unterstellten Bedingungen könnten sich die Ergebnisse bis zu 32 Prozentpunkte vom wahren Wert unterscheiden. Anstatt der „wahren“ 5% könnte eine wie unterstellt gestaltete Umfrage auch ein Ergebnis von  $5 + 32 = 37\%$  hervorbringen und damit sogar noch etwas mehr als in der kritisierten Umfrage. Tatsächlich wären die in dieser Umfrage erzielten 33% schon ab einem Faktor *c* von knapp unter 0.11 erreichbar, was nach unserer Intuition eher zu niedrig als zu hoch angesetzt wäre. Kann man diese Ergebnisse noch als „tendenziell richtig“ ansehen?<sup>6</sup>

Zugegeben, das ist nur eine Beispielrechnung. Vielleicht tun wir dem Online-Access-Panel Unrecht und das Verhältnis zwischen Umfang des Access-Panels und der Populationsgröße ist größer als von uns unterstellt. Vielleicht ist die Antwortrate auch noch höher, als die von uns vermeintlich großzügig unterstellten 80%. Und vielleicht spielen die von uns aufgeführten Merkmale, die den Zusammenhang zwischen Teilnahmewahrscheinlichkeit und dem interessierenden Merkmal begründen, in Wirklichkeit gar keine so große Rolle. Wir vermuten aber, dass wir alle diese Faktoren sogar eher zugunsten der Studie ausgelegt haben.

Bei einer Zufallsstichprobe mit einer sehr niedrigen Ausschöpfung von 5% läge die potentielle Verzerrung unter sonst gleichbleibenden Bedingungen übrigens bei

$$\text{Potentielle Verzerrung} = 0.22 \cdot 4.36 \cdot 0.1196 = 0.11.$$

Das ist immer noch sehr viel, schließt aber 33% als mögliches Ergebnis aus.<sup>7</sup> Außerdem wurde hier nicht berücksichtigt, dass der Faktor *c* im Falle von Zufallsstichproben vermutlich niedriger ist als bei Online-Access-Panels (vgl. Kohler und Post 2023).

## Die Krake-Paul-Situation

Im Rahmen der Fußball-Weltmeisterschaft 2010 erregte das Orakeltier *Krake Paul* großes Aufsehen. Paul gelang es, bei allen sieben Begegnungen der deutschen Mannschaft richtig deren Sieg oder Niederlage vorherzusagen. Die Wahrscheinlichkeit, dass er durch bloßes Raten zu einer solchen Erfolgsserie kommt, ist jedoch extrem gering bzw. genauer 0.0078. Nach herkömmlichen Kriterien ( $\alpha = 0.05$ ) würde man die

Vermutung, dass er einfach nur geraten hat zurückweisen und die Alternativhypothese, dass er einen Riecher für den Sieger hat, akzeptieren. Aus guten Gründen widersetzt sich unsere Intuition allerdings der Versuchung, die Logik eines statistischen Signifikanztests auf den vorliegenden Fall anzuwenden.

Unsere Intuition hat Recht. Die Krake-Paul-Situation ist ein typisches Beispiel des sog. Schubladenproblems. Paul war nur eines von „tausenden“ (Wikipedia) Orakeltieren bei der WM 2010, und die Wahrscheinlichkeit, dass bei 1000 Tieren ein oder mehr Tiere alle sieben Spiele richtig tippen, liegt bei annähernd 100%. Es ist daher alles andere als überraschend, dass es einem Tier gelingt, siebenmal richtig zu liegen. Erklärungsbedürftig ist nur, warum dies eine Nachricht wert war.

Für unser Thema folgt daraus zweierlei. *Erstens:* Das Ergebnis einer defizitären Methode kann auch richtig sein. Und: Wenn es viele solcher defizitären Methoden gibt, wird eine davon trotz der methodischen Defizite zufällig richtige Ergebnisse liefern. Wir raten daher zu Skepsis gegenüber allen Argumenten, die darauf verweisen, dass ein spezifischer Anteilswert aus einer spezifischen Umfrage an einem spezifischen Online-Access-Panel sich im Nachhinein als zutreffend herausgestellt hat.

*Zweitens:* Bei einer Vielzahl von Umfragen wird es immer ein Ergebnis geben, das zu einem gewünschten Narrativ passt und darum als berichtenswert erscheint. Redaktionen sollten sich aber dringend davor hüten im Strom von Umfrageergebnissen zu suchen, bis Ihnen etwas berichtenswert erscheint. Denn so würden sie das Schubladenproblem reproduzieren. Es ist ein fundamentaler Unterschied, ob man ein Argument angesichts bestehender Daten entwickelt oder die Daten erhebt, um ein Argument damit zu prüfen. Oder anders ausgedrückt: Wenn man die Vorhersagen von Paul kennt, kann man diese leicht nutzen, um zu argumentieren, dass Kraken Fußballergebnisse erspüren können. Wenn man aber zur Stärkung desselben Arguments einen Kraken Fußballergebnisse vorhersagen lässt, wird man meistens scheitern. Nur das zweite Vorgehen schützt uns vor dem Schubladenproblem.

Natürlich ist das Schubladenproblem nicht auf Umfragen an Online-Access-Paneln beschränkt. Aufgrund der unbestreitbaren Kosten- und Geschwindigkeitsvorteile ist die Menge von Befragungen an Online-Access-Paneln jedoch deutlich gestiegen. Dies schafft eine wachsende Anzahl von Gelegenheiten, dem Schubladenproblem auf den Leim zu gehen.

## Epilog

Fassen wir zusammen: Umfrageergebnisse sind stets mit Fehlern behaftet. Die potentielle Verzerrung von Umfragen an Online-Access-Paneln ist aber um ein Vielfaches höher, als diejenige von Umfragen an Zufallsstichproben. Für Massenmedien sind Umfragen aus Online-Access-Paneln jedoch gerade deswegen besonders wertvoll. *Weil* sie verzerrter sind, liefern sie häufiger berichtenswerte Befunde. Da sie zudem kostengünstig und schnell sind, hat sich auch die bloße Anzahl der Befunde erhöht, aus denen die Berichtenswertesten rausgesucht werden können. Es besteht daher An-

lass zur Sorge, dass die Gesamtheit solcher Umfrageergebnisse ein Zerrbild der Wirklichkeit liefert.

Überzeichnungen gehören zum Geschäftsmodell der Boulevard-Berichterstattung seit Anbeginn und auch die Vorwürfe dagegen sind nicht neu. Neu ist, dass dieses Geschäftsmodell durch inflationäre Veröffentlichung potentiell verzerrter Umfrageergebnisse auch in Qualitätsmedien eingezogen ist. Vermutlich liegt dem keine beabsichtigte Auswahl verzerrter Ergebnisse zugrunde, sondern fehlendes Wissen zur Beurteilung der Methoden einer Umfrage, deren Ergebnisse einen hohen Nachrichtenwert aufweisen. Erschwerend kommt hinzu, dass die Berichterstattung durch den quasi-wissenschaftlichen Anstrich die fehlende Qualitätskontrolle verschleiern. Die Marktteilnehmer:innen heißen „Institute“ und betreiben „Meinungsforschung“. Zudem wird mit positiv besetzten pseudowissenschaftlichen Begriffen operiert („repräsentativ“). Anders als im Fall der Boulevard-Berichterstattung sind die Qualitätsprobleme der Umfragen daher schwer erkennbar.

Bei all dem ist zu berücksichtigen, dass Meinungsumfragen nicht nur bloßes *Infotainment* sind, sondern ein relevanter Teil des politischen Meinungsbildungsprozesses. Wenn der Spiegel, wie im eingangs erwähnten Beispiel, ein Umfrageergebnis lanciert, das hohe Zustimmung zur Zusammenarbeit zwischen der CDU und der AfD auf kommunaler Ebene signalisiert, übt das Druck auf Politiker:innen aus. Er entfacht damit ein Feuer vor die viel zitierte Brandmauer, mit politischen Konsequenzen, die nicht absehbar sind.

Die Verantwortung für die hier kritisierte Entwicklung liegt aber nicht nur bei den Medien, sondern auch bei der Wissenschaft. Es gibt eine große Scientific Community, die sich mehr oder weniger ausschließlich mit Qualitätsproblemen von Umfragen beschäftigt. Diese Community sollte ihr Know-how in die Öffentlichkeit tragen und auch zu Fehlentwicklungen Stellung beziehen. Wir sollten die Redaktionen der Medienhäuser dabei unterstützen, Qualitätsprobleme zu erkennen. Anstatt die Inhaltsleere des Begriffs *repräsentativ* zu beklagen, sollten wir uns die Bedeutung des Begriffs erkämpfen. Hilfreich wären klare Kriterien dafür, welche Art von Umfragen man in Deutschland als repräsentativ bezeichnen darf. Diese zu entwickeln wäre die Aufgabe großer einschlägiger Fachverbände, wie der *European Survey Research Association*, des *Konsortiums für die Sozial-, Verhaltens-, Bildungs- und Wirtschaftswissenschaften der Deutschen Gesellschaft für Soziologie*, der *Deutschen Vereinigung für Politikwissenschaft* und der *Akademie für Soziologie*.

Bleibe die Frage, welche Kriterien für solch ein Repräsentativitätssiegel herangezogen werden könnten. Das könnten z.B. die 20 Fragen sein, die das Schulungsmaterial für Journalist:innen von Schnell (2019b) auflistet. Auch dort findet sich das Kriterium, dass die Befragten mit einem Zufallsverfahren ausgewählt wurden, jedoch auch noch 19 weitere. Zur Vermeidung des Schubladenproblems würden wir ergänzen, dass turnusmäßig durchgeführte Umfragen mit festgelegtem Fragenkatalog gegenüber tagesaktuell lancierten Umfragen, die bei Bedarf aufgegriffen werden, im Vorteil sind. In solchen regelmäßig stattfindenden Umfragen, lassen sich in begrenztem Umfang auch aktuelle Themen einbinden. Gänzlich verabschieden sollte man sich von der Vorstellung, innerhalb von wenigen Tagen oder gar Stunden ein auch

nur ansatzweise aussagekräftiges Ergebnis aus einer Meinungsumfrage zu bekommen.

## Anmerkungen

- 1 Wir fokussieren uns im Folgenden auf die Verzerrung.
- 2 Die nachfolgende Darstellung beruht auf Formeln für die Verzerrung eines Mittelwertes nach Bethlehem (2010) in der Notation von Kohler und Post (2023). Die Anwendung auf Anteilswerte wird hier dadurch gerechtfertigt, dass sich diese als Mittelwerte auffassen lassen. Eine vertiefte Diskussion mit Simulationen haben wir online gestellt (<https://gitup.uni-potsdam.de/ukohler/pulpscience>).
- 3 Konkret: die Standardabweichung des Merkmals in der Grundgesamtheit.
- 4 Konkret: Der Variationskoeffizient der individuellen Teilnahmewahrscheinlichkeit.
- 5 Vgl. dazu die Simulation in den Begleitmaterialien: <https://gitup.uni-potsdam.de/ukohler/pulpscience>
- 6 Unsere Simulationen mit  $c = 0.11$  in den Begleitmaterialien zeigen, dass die durchschnittliche Verzerrung unter den gegebenen Umständen bei ca. 13 Prozentpunkten liegt. Je nach Thema könnte dieser Unterschied ausreichen, um den Nachrichtenwert über die Schwelle des berichtenswerten zu heben.
- 7 Nach unserer Simulation beträgt die durchschnittliche Verzerrung in diesem Fall sechs Prozentpunkte.

## Literatur

- Bethlehem, J. (2010): Selection bias in web surveys. *International Statistical Review*, 78: 161-188.
- Cornesse, C.; Blom, A. G.; Dutwin, D.; Krosnick, J.A.; De Leeuw, E.D.; Legleye, S.; Pasek, J.; Pennay, D.; Phillips, B.; Sakshaug, J.W.; Struminskaya, B. & Wenz, A. (2020): A Review of Conceptual Approaches and Empirical Evidence on Probability and Nonprobability Sample Survey Research. *Journal of Survey Statistics and Methodology*, 8: 4-36.
- Galtung, J. & Ruge, M.H. (1965): The Structure of Foreign News. The Presentation of the Congo, Cuba and Cyprus Crisis in Four Norwegian Newspapers. In: *Journal of Peace Research* 2: 64-91.
- Kohler, U. & Post, J.C. (2023): Welcher Zweck heiligt die Mittel? Bemerkungen zur Repräsentativitätsdebatte in der Umfrageforschung. *Zeitschrift für Soziologie*, 52: 67-88.
- Lippmann, W. (1922): *Public Opinion*. Harcourt: Brace & Co.
- Schnell, R. (2019a): *Survey-Interviews. Methoden standardisierter Befragungen*. Wiesbaden: VS Verlag.
- Schnell, R. (2019b): *Umfragen verstehen und beurteilen. Sozial- und Politikforschung für Journalisten*. [https://www.uni-due.de/imperia/md/content/soziologie/schnell/schnell\\_2019\\_umfragen\\_verstehen\\_und\\_beurteilen.pdf](https://www.uni-due.de/imperia/md/content/soziologie/schnell/schnell_2019_umfragen_verstehen_und_beurteilen.pdf) (abgerufen am 11. 8. 2023)
- Schulz, W. (1976): *Die Konstruktion von Realität in den Nachrichtenmedien*. Freiburg/München: Karl Alber.
- Warren, C. (1934): *Modern News Reporting*. New York: Harper & Brothers.