

„Den Daumen senken“: Hochmut und Kritik Eine Replik auf die Kritik des Demokratiebarometers von Jäckle, Wagschal und Bauschke

Wolfgang Merkel · Dag Tanneberg · Marc Bühlmann

Demokratie ist ein komplexes System. Seine Qualität zu messen, ist schwierig und voller Fallstricke. Es wäre Hybris anzunehmen, über keinen dieser Fallstricke je gestolpert zu sein. Dies war und ist den Autoren des Demokratiebarometers (DB) bewusst. Ebenso nehmen wir an, dass die Qualität des Demokratiebarometers gerade durch die kritische Auseinandersetzung mit diesem nur gewinnen kann. Kritik belebt, regt an, bisweilen allerdings auch auf. Nach positiver Rückmeldung zum Demokratiebarometer¹ durch Stoiber (2011), Schmidt (2008), Müller und Pickel (2008) sowie vor allem Gerardo Munck (2012) ist nun nach Kaina (2008) und Lauth (2011) eine umfangreichere kritische Diskussion des Demokratiebarometers von Jäckle et al. (2012) in der Zeitschrift für Vergleichende Politikwissenschaft erschienen. Deren Herausgeber haben uns dankenswerter Weise die Möglichkeit gegeben, darauf zu antworten.

Jäckle und Kollegen orientieren ihre Kritik am Demokratiebarometer (DB) an einem Text von Munck und Verkuilen (2002), der zu Recht zu einem Klassiker der Demokratiemessung geworden ist. Nach Munck und Verkuilen lässt sich die Güte einer Messanlage systematisch auf drei Ebenen prüfen: der Konzeptualisierung, der Operationalisierung und der Messung. Dieser Vorgabe folgen auch Jäckle und Kollegen. Wir halten diese Systematik für überzeugend und wollen daran auch unsere Replik orientieren.

1 Natürlich waren unter diesen positiven Rückmeldungen auch kritische Anmerkungen. Für deren Konstruktivität bedanken wir uns bei den Autoren.

Online publiziert: 15.02.2013

© Springer Fachmedien Wiesbaden 2013

Prof. Dr. W. Merkel (✉) · Dipl.-Pol. D. Tanneberg
Wissenschaftszentrum Berlin für Sozialforschung (WZB),
Reichpietschufer 50, 10785 Berlin, Deutschland
E-Mail: wolfgang.merkel@wzb.eu

Dipl.-Pol. D. Tanneberg
E-Mail: dag.tanneberg@wzb.eu

Dr. M. Bühlmann
Universität Bern, Institut für Politikwissenschaft,
Lerchenweg 36, 3000 Bern, Schweiz
E-Mail: marc.buehlmann@ipw.unibe.ch

1 Konzeptualisierung

Bei der Konzeptualisierung gehen die drei Kritiker auf drei Aspekte ein: auf die „grundlegende Konzeption“; auf die Frage, ob das Konzept nicht doch zu „maximalistisch“ sei, sowie schließlich auf die Indikatorenwahl. Mit Blick auf das grundlegende Konzept machen es sich die drei Autoren doch etwas einfach, indem sie erstaunlich unkritisch die Bedenken einer vierten Kritikerin schlicht paraphrasieren: Viktoria Kainas (2008) kurze Replik auf einen früheren Forumsbeitrag der Autoren des Demokratiebarometers zu einem Überblicksartikel von Müller und Pickel (2007) über unterschiedliche Konzepte der Demokratiemessung (vgl. Bühlmann et al. 2008).² Die Kritik basiert also nicht auf den aktuellen grundlegenden Veröffentlichungen zum Demokratiebarometer, die das Instrument erstmals in Gänze der *scientific community* vorstellen (Bühlmann et al. 2012a, b). Schon Kaina zitierte nicht präzise. Ihre Unschärfe steigern Jäckle et al. mit ihrer offensichtlich ungeprüften Zweitcitation noch einmal und kombinieren sie mit einem schweren logischen Fehlschluss.

Das liest sich bei den drei Kritikern dann so: „Einerseits werden sie [die drei Grundprinzipien Freiheit, Gleichheit, Kontrolle; Anm.: Merkel et al.] als normativ gleichwertig erachtet und andererseits – und das ist das entscheidende Novum zu alternativen Messansätzen – stehen sie gleichzeitig in einem interdependenten Verhältnis zueinander. Dies bedeutet jedoch auch, dass die Maximierung einer Dimension zu Lasten einer oder der anderen beiden Dimensionen geht“ (Jäckle et al. 2012, S. 106). Seit wann bedeutet Interdependenz, dass die Optimierung eines Prinzips zwangsläufig zu Lasten des anderen geht? Weder folgt dies aus unseren Texten, noch gehorcht es den fundamentalen Gesetzen der Logik. Unsere Kritiker erliegen vielmehr dem, was man in der Logik ein *non sequitur* nennt: Letzteres folgt schlicht nicht aus ersterem.

Aber selbst ein Spannungsverhältnis, wie es Tocqueville (1959 [1835]) für Gleichheit und Freiheit annimmt, bedeutet keineswegs, dass Zugewinne an Gleichheit immer auch zu Lasten der Freiheit und *vice versa* gehen. Wir teilen vielmehr die Ansicht von Amartya Sen (2000), dass der Ausbau der (positiven) Freiheit in hohem Maße auch die soziale wie politische Gleichheit fördert, ja, erstere geradezu zur Voraussetzung der letzteren macht.³ Naiv wäre es allerdings anzunehmen, dass ein theoretisches Maximum an Gleichheit mit einem theoretischen Maximum an individueller Freiheit zu erreichen wäre. Wohl aber können die Werte für Freiheit und Gleichheit in real existierenden Demokratien je gesondert gemessen werden. Und je besser diese beiden Prinzipien gemeinsam mit dem Prinzip

2 Wir halten es für ärgerlich, wenn nicht nachgerade für wissenschaftlich schludrig, dass die drei Kritiker diese Publikation von uns (Bühlmann et al. 2008) überhaupt nicht in ihrer Literaturliste aufführen, sondern wohlweislich direkte Publikationen (Bühlmann et al. 2011; 2012a, b) zum Demokratiebarometer zur weiteren Grundlage ihrer Kritik machen. Dennoch basiert der gesamte Abschnitt zur „fundamentalen Kritik an der grundlegenden Konzeption nach Viktoria Kania“ auf diesem von den drei Autoren nicht aufgeführten Text, der die ersten Eckpunkte des Demokratiebarometers in *statu nascendi* darstellt.

3 Unsere Messungen deuten im Übrigen darauf hin, dass Länder in unserem Sample der besten 30 Demokratien, die gut bei dem Prinzip Freiheit abschneiden, auch hohe Werte *in puncto* Gleichheit erzielen. Vgl. zum Zusammenhang von Demokratie und sozialer Gerechtigkeit auch Merkel und Giebler (2009).

der Herrschaftskontrolle in demokratischen Regimen erfüllt werden, desto größer ist die Qualität der Demokratie.⁴

Völlig auf dem Flugsand von Kainas unhinterfragten Deutungen ist folgende Aussage der drei Autoren gebaut: „Das Demokratiebarometer postuliert, dass sich ‚der Maßstab für die Demokratiequalität eines demokratischen politischen Systems an einem Optimum orientier[t], das sich für das Spannungsverhältnis jener drei Demokratieprinzipien sensibel zeigt‘ (Kaina 2008, S. 521). Damit liegt aber eben gerade kein gemeinsamer Maßstab vor, der an alle drei Prinzipien immer in derselben Weise angelegt werden kann und die Bildung einer Rangfolge erlauben würde“ (Jäckle et al. 2012, S. 106). Zunächst: Dass sich das Demokratiebarometer an einem ‚Optimum orientiere, das sich für das Spannungsverhältnis jener drei Demokratieprinzipien sensibel zeige‘, ist die misslungene Paraphrasierung einer Überlegung, die im Original so lautet: „Damit ein Regime als Demokratie bezeichnet werden kann, muss es Freiheit und Gleichheit garantieren und schützen. Darüber hinaus muss es die Interdependenzen dieser beiden Prinzipien vermittelt vertikaler und horizontaler Kontrolle balancieren und optimieren“ (Bühlmann et al. 2012b, S. 122). Formuliert wird hier nur eine systemische Selbstverständlichkeit, die über eine gute Demokratiequalität mit entscheidet. Warum so ein gemeinsamer Maßstab für die jeweils gesonderte Messung von Gleichheit, Freiheit und Kontrolle verloren ginge, übersteigt unser theoretisches Fassungsvermögen.

Angesichts solch irreführender Zitation von Zweitquellen und grober Fehlschlüsse nehmen wir es mit Humor, wenn Jäckle und seine Koautoren bereits nach wenigen Seiten mit cäsaristischem Gestus schreiben: „Je nachdem für wie kritisch man die angesprochenen und zumindest von Kaina als durchaus fundamental erachteten theoretisch-konzeptionellen Defizite des Demokratiebarometers betrachtet, könnte man bereits an dieser Stelle jede weitere Evaluation abrechnen und den Daumen über dieses Messinstrument senken“ (Jäckle et al. 2012, S. 107). Es war wohl am Ende entweder Barmherzigkeit oder doch die Intuition, dass eine Seite Sekundärliteratur nicht ausreicht, um ein Projekt abzufertigen, welches nach Aussage von Gerardo Munck „[the; Anm.: Merkel et al.] only one exception“ (Munck 2012, S. 5) unter den bisherigen Demokratiemessungen darstellt, die die konzeptionelle Herausforderung der Demokratie ernst nimmt, indem sie dieser mit der klassischen Tradition der Politischen Theorie (Bühlmann et al. 2012b, S. 117 f.) begegnet.

Jäckle, Wagschal und Bauschke ordnen unser expliziertes Demokratieverständnis „als eher maximalistisches Konzept“ ein, begründen dies aber nicht. Wir selbst sehen unser Demokratiekonzept in einer mittleren Position, da wir auf der polyarchischen Basis von Dahl (1971) zwar den Rechtsstaat und die konstitutionelle Gewaltenkontrolle berücksichtigen (vgl. auch Habermas 1992), aber *outputs* oder gar *outcomes* wie Sozialpolitik oder soziale Gerechtigkeit nicht aufgreifen. Insofern bleiben wir weit unterhalb des Maximalismus von Herrmann Heller (1983 [1934]) und Thomas Mayer (2005), die von „sozialer

4 Unser Messinstrument wurde für die 30 besten Demokratien entwickelt. Dort ist Freiheit und Gleichheit in einem zwar variierenden, aber hohen Maße gegeben. Innerhalb eines solchen Samples gehen wir hinsichtlich des Gesamtindex „Demokratiequalität“ von einer partiellen wechselseitigen Substituierbarkeit aus, da beide Prinzipien in dieser Länderauswahl nicht unter den für Demokratien hinreichenden Schwellenwert für die beiden Prinzipien rutschen können.

Demokratie“ sprechen. Immerhin konstatieren Jäckle et al. (2012, S. 107–108), dass wir „die drei Prinzipien Freiheit, Gleichheit und Kontrolle überzeugend begründen und damit zumindest auf der obersten Ebene des Konzeptbaumes kein überdehntes Demokratie-modell aufstellen“.

2 Operationalisierung

Das dritte Monitum bezieht sich auf die Indikatorenauswahl. Neben einigen positiven Anmerkungen – u. a. die symmetrische Berücksichtigung von Indikatoren, die sich jeweils auf die *de jure* und *de facto* Verhältnisse beziehen – konzentriert sich die Kritik auf drei Vorwürfe: Die Indikatoren seien nicht stringent und explizit abgeleitet worden, sie seien zum Teil falsch gewählt und zu zahlreich und führten deshalb zu einem *conceptual stretching* (Sartori 1970). Zumindest der erste Punkt trifft in engen Grenzen zu. In der Tat leiteten wir in unseren bisherigen Veröffentlichungen nicht alle 100 Indikatoren über all die Stufen von den Prinzipien (3) über die Funktionen (9), Komponenten (18) und Subkomponenten (51) bis hin zu den Indikatoren ab. Schon die Explikation der 3 Prinzipien und 9 Funktionen nahm in unseren bisherigen programmatischen Veröffentlichungen (Bühlmann et al. 2012a, b) einen so weiten Raum ein, dass wir ein *revise and resubmit* befürchten mussten. Die Ableitung nun für 18 Komponenten, 51 Subkomponenten und 100 Indikatoren zu fordern, ist für wissenschaftliche Publikationen witzlos. Einhundertzwanzig Seiten stehen keinem Autor zur Verfügung. Eine solche lückenlose Dokumentation gibt es bei keinem einzigen Messinstrument – bei Freedom House angefangen, weiter über Polity IV und die World Bank Governance Indicators bis hin zum Democracy Index des „Economist“. Jedoch kann man für ein Codebuch fordern, wir sollten alle Ableitungsschritte dokumentieren. Wir nehmen daher den Hinweis der Autoren für die Erweiterung unseres Codebuchs (erhältlich unter www.democracybarometer.org)⁵ auf.

Schwerer wiegt der Vorwurf „mangelnder Stringenz“ (Jäckle et al 2012, S. 109). Den drei Kritikern, die von uns eine lückenlose Dokumentation der 100 Indikatoren einfordern, genügen sage und schreibe dreieinhalb Indikatoren, um diesen Vorwurf zu belegen. Bezweifelt wird die Aussagekraft der Indikatoren für die Qualität einer Demokratie. „So erscheint die Verwendung der Mordrate und der Anzahl gewalttätiger Proteste zwar für die Gesamtkomponente ‚Recht auf körperliche Unversehrtheit‘ durchaus notwendig, aber ob diese Komponente in ihrer Gänze wirklich einen notwendigen Teilaspekt der Demokratiequalität darstellt, kann wohl bezweifelt werden“ (Jäckle et al. 2012, S. 110). Ein Blick auf den dokumentierten Ableitungsbaum hätte genügt, um unsere Argumentationskette nachzuvollziehen und sie dann konkret zu kritisieren. Die Ableitungssequenz verläuft folgendermaßen: Sie beginnt mit „Freiheit“, einem der drei Grundprinzipien unseres Demokratiekonzepts. Eine von drei Funktionen dieses allgemeinen Prinzips ist die Garantie „individueller“ Freiheiten. Von dieser werden wiederum zwei Komponenten

5 Bisher gibt es zwar auf 51 Seiten eine lückenlose Dokumentation aller Prinzipien, Funktionen, Komponenten, Subkomponenten und Indikatoren. Die geforderte Ableitung und Begründung ist bisher auf die Ebenen der Funktionen beschränkt.

abgeleitet, wovon eine „Recht auf physische Integrität“ heißt. Sie gliedert sich wiederum in drei Subkomponenten, von denen eine als „wechselseitige Akzeptanz des Rechts auf individuelle physische Integrität durch die Bürger“ firmiert. Diese Subkomponente messen schließlich die beiden Indikatoren „niedrige Mordrate“ und „keine gewaltsamen Aufstände“.

Eine seriöse Kritik hätte zumindest versuchen können, die dargelegte Ableitungskette an wenigstens einem der dokumentierten Schritte zu widerlegen. In der Sache geben wir Folgendes zu bedenken: Sind gewaltsame Aufstände, etwa der Afroamerikaner oder Pakistani in manchen Städten der USA und Großbritanniens, oder sind die hohen Mordraten in den rassistisch segregierten Inner-Cities nicht doch Ausdruck einer verfehlten Politik – einer Politik, die nicht in der Lage ist, Freiheit und physische Unversehrtheit für alle Bürger so weit wie nur möglich zu garantieren? Je besser das in einer Demokratie gelingt, umso besser werden individuelle Freiheiten geschützt, je besser diese geschützt werden, desto höher ist *ceteris paribus* die Qualität der Demokratie. Der kritische Hinweis, niedrige Mordraten könnten „auch als ein Indiz für ein autoritäres Regime erachtet werden“ (Jäckle et al., S. 110) läuft gleich zweifach ins Leere. Erstens, seit wann sind niedrige Mordraten ein Ausweis für die hohe Qualität autoritärer Regime? Machen niedrigere individuelle Mordraten im Hitlerregime dieses zu einer besseren Diktatur als jene anarchoiden Gewaltherrschaften von Charles Taylor in Liberia oder der wechselnden Autokraten im Kongo? Zudem, und dies ist der gewichtigste Hinweis, haben wir das Demokratiebarometer explizit und unübersehbar für Demokratien entwickelt und nicht für alle politischen Regime, also auch nicht für Diktaturen. Hier wird ein Popanz aufgebaut, der dann mit der Eleganz eines Don Quichote erledigt werden kann.

Unsere Kritiker bezweifeln außerdem, dass der Organisationsgrad von Gewerkschaften oder die zivilgesellschaftliche Dichte von Menschenrechts- und Tierschutzvereinen etwas mit Demokratiequalität zu tun haben. Im Einklang mit der herrschenden Meinung in der zivilgesellschaftlichen Literatur vertreten wir jedoch die Ansicht, dass die kollektive Selbstorganisation der Gesellschaft jenseits des Staates wichtige Organisations-, Aggregations- und Artikulationsfunktionen im pluralistischen Ideen- und Interessenwettbewerb erfüllt. Sie sind Teil einer vitalen Zivilgesellschaft und helfen, eine Öffentlichkeit⁶ erst zu konstituieren, die die konstitutionellen Organisationen und Institutionen demokratisch unterfüttert (u. a. Habermas 1992). Jäckle et al. (2012, S. 111) argumentieren hingegen: „Der gewerkschaftliche Organisationsgrad hängt stark mit der historischen Entwicklung des Wirtschafts- und Sozialsystems zusammen, die eben nur zum Teil durch die Politik und damit Demokratie geprägt war. Aus diesem Grund kann dieser Indikator auch nicht als Maß für die Qualität einer Demokratie dienen“. Die Logik frappt. Weil Institutionen und Organisationen auch von der „historischen Entwicklung des Wirtschafts- und Sozialsystems“ geprägt worden sind, sollen sie nicht bei der Qualitätsbeurteilung der Demokratie herangezogen werden dürfen? Was bliebe da noch übrig für die Beurteilung? Die meisten der politischen und gesellschaftlichen Institutionen, Organisationen, Werte, Einstellungen und Verhaltensweisen in Demokratien sind auch vom Wirtschafts- und Sozialsystem geprägt. Die politischen Parteien, das Wählerverhalten, die Zivilgesellschaft, die

6 Die Kritiker nennen „Öffentlichkeit“ eine „Komponente“ (Jäckle et al. 2012, S. 111); wir weisen sie in unserem Konzeptbaum jedoch klar als eine „Funktion“ aus.

Medienlandschaft, das Rechtssystem, der Föderalismus: Sie alle sind keineswegs allein „durch die Politik und damit die Demokratie“ geprägt.

Wir wissen, dass nicht alle Indikatoren hundertprozentig valide sein können, und man darüber diskutieren kann und muss. Drei aus 100 Indikatoren herauszuziehen, um damit mangelnde Stringenz und Aufblähung unserer Konzeptualisierung zu belegen, zeugt von einer gewissen Unerschrockenheit. Aber selbst zu diesen drei Indikatoren werden logisch brüchige oder schlicht sinnwidrige Begründungen vorgelegt. Möglicherweise lassen sich Eifer und Fahrlässigkeit der Kritiker mit Schopenhauer erklären: „Imponieren, Verdutzen, Mystifizieren, dem Leser durch allerlei Kunstgriffe Sand in die Augen streuen ist die Methode geworden, und durchgängig leitet statt der Einsicht die Absicht den Vortrag.“ (Schopenhauer 1989 [1839], S. 610).

3 Messung

Ihre methodologischen Einwände eröffnen die Autoren mit der Frage, ob die Vielzahl reliabler Indikatoren des Demokratiebarometers deren vorgeblich fragliche Validität aufwiege. Sie unterstellen dem Barometer eine „Überbetonung des Reliabilitätskriteriums bei gleichzeitiger Vernachlässigung des eigentlich noch wichtigeren Validitätskriteriums“ (Jäckle et al. 2012, S. 112). Diese Kritik am Demokratiebarometer zielt vor allem auf die Relativierung unserer Skepsis gegenüber Expertenbefragungen (vgl. Bollen 1990; Bollen und Paxton 2000. Jäckle et al. (2012: 112) behaupten, gerade bei komplexen Angelegenheiten wie der Demokratiequalität könne „auf das Wissen von Experten eventuell nicht verzichtet werden“. Es müssten nur hinreichend viele von ihnen befragt werden, um eine objektiven Daten vergleichbare Reliabilität zu gewährleisten. Beide Einwände sind für uns nur schwer nachvollziehbar.

So weist bereits eine einflussreiche Einführung in die Sozialwissenschaft darauf hin, dass Reliabilität eine notwendige Bedingung für die Validität eines Messinstruments darstellt (Diekmann 2008, S. 267): Je geringer die Zuverlässigkeit einer Messung ausfällt, desto weniger kann ihre Gültigkeit erwartet werden (vgl. Carmines und Zeller 1979). Gleichzeitig erstaunt die Unbekümmertheit, mit der Jäckle et al. die Probleme von Expertenbefragungen allein auf deren Teilnehmerzahl verengen. Expertenmeinungen sind jedoch nur unter anspruchsvollen Voraussetzungen valider als andere Erhebungsformen. Hierzu zählen standardisierte Erhebungsinstrumente, die weder fehlerhaft konstruiert noch missverständlich formuliert wurden und deren Elemente nicht mehrdimensional sind (Giebler 2012, S. 513). Daran scheitert es oft und infolgedessen steigt mit der Teilnehmerzahl meist weder die Reliabilität noch die Validität solcher Umfragen. Die breite, klaren Kriterien unterworfenen Quellenwahl ist folglich eine wesentliche Stärke des Demokratiebarometers, da sie systematischen oder zufälligen Verzerrungen des Messergebnisses vorzubeugen und die Unwägbarkeiten von Expertenbefragungen zu vermeiden sucht. Hiervon profitieren Reliabilität wie Validität des Barometers gleichermaßen.

Wichtiger erscheinen uns die Einwände der drei Kritiker gegen das Skalierungsverfahren des Demokratiebarometers. Sie behaupten einerseits, das Best-Practice-Verfahren bedeute nichts anderes als „die Problematik der theoretischen Bestimmung der Maxima und Minima auf ein theoretisch unreflektiertes ‚der Wert sollte möglichst groß sein‘ auszu-

lagern“ (Jäckle et al. 2012, S. 114). Andererseits verletze unser Skalierungsverfahren das Transitivitätsgebot, wonach jede gültige Transformation der Rohdaten die ursprüngliche Rangfolge der Untersuchungseinheiten unberührt lassen muss. Folglich sei die Rangfolge der im Barometer erfassten Länder ein Artefakt technischer Mängel und nicht Ausdruck einer unterschiedlichen Demokratiequalität. Unseres Erachtens missverstehen die Kritiker sowohl die Prämissen als auch die Eigenschaften unseres Skalierungsverfahrens.

Nach welchem Maßstab lässt sich die Qualität einer Demokratie beurteilen? In Anbetracht der Unzulänglichkeiten anderer Ansätze (vgl. Bühlmann et al. 2011) versucht das Demokratiebarometer, mit Hilfe des Best-Practice-Verfahrens aus dem empirischen Material selbst Richtwerte zu gewinnen (Bühlmann et al. 2012b, S. 131 ff.). Eben hierzu dient das Blueprint-Sample der 30 besten Demokratien. Die darin erfassten Länder erfüllen fraglos sämtliche Funktionen der Demokratie – auf unterschiedlichem Niveau. Sie geben daher eine geeignete Richtschnur für die Spannweite der Qualität etablierter Demokratien vor. Genau dieses Kontinuum fängt das Best-Practice-Verfahren ein.

Der Vorwurf, diese Skalierung verstoße gegen das Transitivitätsgebot, ignoriert die fundamentale Annahme des Demokratiebarometers, alle Elemente auf derselben Aggregationsstufe seien gleich wichtig (Bühlmann et al. 2011, S. 10; 2012b, S. 134). Zunächst bedeutet dies eine Absage an explizite Gewichtungsschemata. Allerdings muss im selben Moment einer impliziten Gewichtung des Datenmaterials vorgebeugt werden. Letztere resultiert unter anderem aus unterschiedlichen Spannweiten der verwendeten Indikatoren: Je breiter die Messwerte auf der Skala eines Indikators X im Vergleich zu einem Indikator Y unter sonst gleichen Bedingungen streuen, desto geringer fällt das Gewicht von X bei Aggregation von X und Y aus. Das Best-Practice-Verfahren trägt dem Rechnung, indem es die Spannweite jedes Indikators zu dessen Standardisierung heranzieht.⁷ Jäckle et al. berücksichtigen das Problem impliziter Gewichtungen hingegen weder in dem von ihnen gegebenen Zahlenbeispiel noch im Zuge der von ihnen vorgenommenen Z-Transformation, woraus die unterstellten „Mängel“ erst resultieren.

Vorwürfe richten sich auch gegen das Aggregationsverfahren des Demokratiebarometers. Die Kritiker unterstellen zunächst einen theoretisch unvermittelten Bruch bei der Verdichtung von Komponenten zu Funktionen und behaupten anschließend, der von uns bei der Aggregation von Komponenten zu Funktionen, von Funktionen zu Prinzipien und von Prinzipien zum Gesamtindex verwendete Arkustangens erzeuge künstliche Varianz. Jäckle et al. schlagen daher vor, statt des Arkustangens das geometrische Mittel zur Aggregation einzusetzen. Ihr Vorschlag führt in die Irre.

So handelt es sich bereits bei der Kennzeichnung des Aggregationsprozesses als eines auf notwendigen und hinreichenden Bedingungen beruhenden Verfahrens (Jäckle et al. 2012, S. 116) um eine Fehlinterpretation. Jene logischen Operatoren kamen lediglich bei der Bestimmung der Blueprint-Länder zur Anwendung. Es galt gemäß der Forderung

7 Technisch handelt es sich um eine Min-Max-Normalisierung. Es sei x_{\min} das beobachtete Minimum und x_{\max} das beobachtete Maximum eines beliebigen Indikators x . Im Anschluss an die Normalisierung soll das neue Minimum 0, das neue Maximum 100 betragen. Erreicht wird dies durch folgende Transformation:

$$x_{norm} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \cdot (x_{\max}^{neu} - x_{\min}^{neu}) + x_{\min}^{neu} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \cdot 100$$

Sartoris (1970, 1991), die Klassifikation eines homogenen Samples aller Quantifizierung voranzustellen. Infolgedessen erreichen notwendige und hinreichende Bedingungen keine weitere Differenzierung der Länderauswahl (s. oben; vgl. Bühlmann et al. 2011, S. 9 f.). Jedoch fällt es schwer, bereits auf der Ebene der Indikatoren und Subkomponenten theoretische Annahmen über deren Effekt auf die Demokratiequalität zu treffen. Eben diese Schwierigkeit erklärt den Einsatz des arithmetischen Mittels (Bühlmann et al. 2011, S. 13). Ab der Funktionsebene liegt hingegen ein klarer theoretischer Rahmen vor, welcher systematisch umgesetzt wurde.

So interpretieren wir die gleichwertige Ausprägung der höheren Demokratieelemente als Beleg ihrer Balance. Asymmetrische Ausprägungen unterlaufen diese Balance und beeinträchtigen die Demokratiequalität. Wie Jäckle et al. (2012, S. 116) zutreffend herausarbeiten, unterstellt das Demokratiebarometer daher „eine begrenzte Substituierung ungleicher Werte“. Darüber hinaus sollen Ungleichgewichte bestraft werden: „Punishment for equal degrees of disequilibrium should be punished equally, larger disequilibrium more than smaller disequilibrium. This implies progressive discount the larger the disequilibrium“ (Bühlmann et al. 2011, S. 10). Den Schlussstein setzt die Annahme abnehmender Grenzerträge. Jenseits einer gewissen Schwelle tragen Verbesserungen oder Verschlechterungen eines Demokratieelements weniger zur Steigerung respektive Minderung der Demokratiequalität bei. Methodisch löst die Multiplikation die „begrenzte Substituierbarkeit“ ein, während der Arkustangens sowohl die progressive Bestrafung etwaiger Ungleichgewichte als auch die nach oben *und* unten abnehmenden Grenzerträge abbildet (vgl. Bühlmann et al. 2011, S. 8 ff.).

Jäckle et al. behaupten nun, im Vergleich zu arithmetischem und geometrischem Mittel bestrafe der Arkustangens uneinheitliche Indikatorenausprägungen über Gebühr und erzeuge deshalb künstliche Varianz. Wir fragen uns jedoch, weshalb das arithmetische Mittel unserem Aggregationsverfahren überhaupt zum Maßstab gereichen sollte. Weder schränkt es die Substituierbarkeit ungleicher Messwerte ein, noch erlaubt es die progressive Bestrafung bestehender Asymmetrien, noch setzt das arithmetische Mittel den Gedanken abnehmender Grenzerträge um. Kurz: Der klassische Mittelwert bricht auf ganzer Linie mit unseren theoretischen Ansprüchen und führt in der Diskussion nicht weiter. Das geometrische Mittel nimmt zwar den Gedanken begrenzter Substituierbarkeit auf, beschreibt aber eine deutlich andere funktionale Form der Demokratiequalität. Es formuliert keine abnehmenden Grenzerträge, sondern eine lineare Qualitätszunahme. Ob das geometrische Mittel uneinheitliche Messwerte angemessener bestraft, ist dann nicht mehr von Belang, denn theoretisches Konzept und empirische Messung klaffen auseinander. Genau davor warnt Goertz (2006, S. 96): „if the measure has a different structure from the concept then it is a measure of a different concept“. Folgten wir also dem Rat unserer Kritiker, begingen wir einen methodologischen Kardinalfehler.

4 Schluss

Am Ende ihrer bisweilen harsch formulierten Kritik vollziehen Jäckle et al. eine geradezu kopernikanische sprachliche Wende. Sie bescheinigen dem DB „Optimierungen gegenüber vorhergegangenen Messvorschlägen“, auch wenn es „noch eine Reihe

an Schwächen sowohl in theoretisch-konzeptioneller als auch in praktischer Hinsicht“ (2012, S. 124) zeige. Von Optimierung wagen wir nicht zu sprechen. Dennoch glauben wir mit dem Barometer ein Instrument vorgelegt zu haben, dass zum Ersten die gar nicht so feinen Unterschiede der etablierten Demokratien erfasst; zum Zweiten die Auswahl der Indikatoren erstmals auf ein deduktiv expliziertes Demokratiekonzept zurückführt und zum Dritten eine Datensammlung zur Verfügung stellt, die für vielfältige Analysen der Entwicklung der 30 besten Demokratien verwendet werden kann.

Literatur

- Bollen, Kenneth A. 1990. Political democracy: Conceptual and measurement traps. *Studies in Comparative International Development* 25 (1): 7–24.
- Bollen, Kenneth A., und Pamela Paxton. 2000. Subjective measures of liberal democracy. *Comparative Political Studies* 33 (1): 58–86.
- Bühlmann, Marc, Wolfgang Merkel, Lisa Müller, und Bernhard Wessels. 2008. Wie lässt sich Demokratie am besten messen? Zum Forumsbeitrag von Thomas Müller und Susanne Pickel. *Politische Vierteljahresschrift* 49 (1): 114–122.
- Bühlmann, Marc, Wolfgang Merkel, Lisa Müller, Heiko Giebler, und Bernhard Wessels. 2011. *Democracy Barometer: Methodology*. Aarau: Zentrum für Demokratie. <http://democracybarometer.org/Papers/Methodology%201990–2007.pdf>. Zugegriffen: 9. Jan. 2012.
- Bühlmann, Marc, Wolfgang Merkel, Lisa Müller, und Bernhard Wessels. 2012a. The democracy barometer: A new instrument to measure the quality of democracy and its potential for comparative research. *European Political Science* 11 (4): 519–536.
- Bühlmann, Marc, Wolfgang Merkel, Lisa Müller, Heiko Giebler, und Bernhard Wessels. 2012b. Demokratiebarometer: ein neues Instrument zur Messung von Demokratiequalität. In *Indizes in der Vergleichenden Politikwissenschaft*. ZfVP-Sonderheft 2/2012, Hrsg. Gert Pickel und Susanne Pickel, 115–159. Wiesbaden: SpringerVS.
- Carmines, Edward G., und Richard A. Zeller. 1979. *Reliability and validity assessment*. Thousand Oaks: Sage.
- Dahl, Robert. 1971. *Polyarchy: Participation and opposition*. London: Yale University Press.
- Diekmann, Andreas. 2008. *Empirische Sozialforschung: Grundlagen, Methoden, Anwendungen*. 19. Aufl. Hamburg: Rowohlt Taschenbuch.
- Giebler, Heiko. 2012. Bringing methodology (back) in: Some remarks on the contemporary democracy measurements. *European Political Science* 11 (4): 509–518.
- Goertz, Gary. 2006. *Social science concepts. A user's guide*. Princeton: Princeton University Press.
- Habermas, Jürgen. 1992. *Faktizität und Geltung*. Frankfurt a. M.: Suhrkamp.
- Heller, Hermann. 1983 [urspr. 1934]. *Staatslehre*. 6. bearbeitete Auflage, hrsg. von Gerhart Niemeyer. Tübingen: Mohr Siebeck.
- Jäckle, Sebastian, Uwe Wagschal, und Rafael Bauschke. 2012. Das Demokratiebarometer: „basically theory driven“? *Zeitschrift für Vergleichende Politikwissenschaft* 6 (1): 99–125.
- Kaina, Viktoria. 2008. Die Messbarkeit von Demokratiequalität als ungelöstes Theorieproblem. *Politische Vierteljahresschrift* 49 (3): 518–524.
- Lauth, Hans-Joachim. 2011. Qualitative Ansätze der Demokratiemessung. *Zeitschrift für Staats- und Europawissenschaften* 9 (1): 49–77.
- Mayer, Thomas. 2005. *Theorie der sozialen Demokratie*. Wiesbaden: VS Verlag für Sozialwissenschaften.

- Merkel, Wolfgang, und Heiko Giebler. 2009. Measuring social justice and sustainable governance in the OECD. In *Sustainable governance indicators 2009, policy performance and executive capacity in the OECD*, Hrsg. Bertelsmann Stiftung, 187–215. Gütersloh: Verlag Bertelsmann Stiftung.
- Munck, Gerardo L., und Jay Verkuilen. 2002. Conceptualizing and measuring democracy. Evaluating alternative indices. *Comparative Political Studies* 35 (1): 5–34.
- Munck, Gerardo L. 2012. *Conceptualizing the quality of democracy. The framing of a agenda for comparative politics*. Dis Working Paper Series WP/2012/23. Budapest: Central European University, Center for the Study of Imperfections in Democracies.
- Müller, Thomas, und Susanne Pickel. 2007. Wie lässt sich Demokratie am besten messen? Zur Konzeptqualität von Demokratie-Indizes. *Politische Vierteljahresschrift* 48 (3): 511–539.
- Müller, Thomas und Susanne Pickel. 2008. Antwort auf die Replik von Marc Bühlmann, Wolfgang Merkel, Lisa Müller und Bernhard Wessels zum Forumsbeitrag von Thomas Müller und Susanne Pickel. *Politische Vierteljahresschrift* 49 (1): 123–126.
- Sartori, Giovanni. 1970. Concept misformation in comparative politics. *American Political Science Review* 64 (4): 1033–1053.
- Sartori, Giovanni. 1991. Comparing and miscomparing. *Journal of Theoretical Politics* 3 (3): 243–257.
- Schmidt, Manfred G. 2008. *Demokratietheorie. Eine Einführung*. 4. Aufl. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Schopenhauer, Arthur. 1989 [urspr. 1839]. Über die Freiheit des Willens. In *Kleinere Schriften*, Hrsg. Arthur Schopenhauer, 521–631. Frankfurt a. M.: Suhrkamp.
- Sen, Amartya. 2000. *Ökonomie für den Menschen: Wege zu Gerechtigkeit und Solidarität in der Marktwirtschaft*. München: Carl Hanser.
- Stoiber, Michael. 2011. *Die Qualität von Demokratien im Vergleich. Zur Bedeutung des Kontextes in der empirisch vergleichenden Demokratietheorie*. Baden-Baden: Nomos.
- Tocqueville, Alexis de. 1959 [urspr. 1835]. *Über die Demokratie in Amerika*. Stuttgart: Reclam.