

Experimentelle Rechtstatsachenforschung im lauterkeitsrechtlichen Prozess: Beweismöglichkeiten jenseits der Demoskopie

Alexander Egberts*

Vor etwas mehr als einem Jahrhundert führte Arthur Nussbaum den Begriff der Rechtstatsachenforschung ein, um damit die Ermittlung von Tatsachen, die „für ein volles Verständnis und eine sachgemäße Anwendung der Normen erforderlich“ sind, zu bezeichnen.¹ Der damit verbundene programmatische Aufruf, sich weniger einer „leeren Dogmatik“ zu bedienen und vermehrt tatsächliche Lebensumstände zu berücksichtigen, ist bis heute nicht abgeklungen.² Im Kern geht es dabei um die Ermittlung der „für das Recht notwendigen Empirie“.³ Diese Ermittlung ist im Grundsatz methodenoffen,⁴ beschränkt sich in Deutschland allerdings überwiegend auf beobachtende Studien mit beschreibendem Erkenntnisinteresse.⁵ Anders ist

* LL.M., M.A.; den Teilnehmer:innen der 33. Jahrestagung der GJZ sei für die hilfreichen Diskussionsbeiträge und persönlichen Rückmeldungen gedankt. Dank für wertvolle Kommentare am Manuskript gilt zudem Frederike Zufall, Hanjo Hamann, Johannes Rottmann, Johannes Kruse, Fabian Jäkel, Thomas Holzhausen und Arian Henning. Zur Verbesserung der Lesbarkeit werden determinierende Kompositionsglieder nicht gegendert.

1 A. Nussbaum, *Rechtstatsachenforschung: Ihre Bedeutung für Wissenschaft und Unterricht*, Tübingen 1914, S. 6. Erste Nennung des Begriffs auf S. 8: „will man [die in die Jurisprudenz gehörigen Tatsachen] mit einem Schlagwort bezeichnen, so wird dafür vielleicht der Ausdruck ‚Rechtstatsachen‘ gewählt werden können“.

2 Prominent N. Petersen, *Braucht die Rechtswissenschaft eine empirische Wende?*, *Der Staat* 49 (2010), 435 (455).

3 A. Chiotellis / W. Fikentscher, *Einführung*, in: Chiotellis/Fikentscher (Hrsg.), *Rechtstatsachenforschung*, Köln 1985, I (3).

4 So auch T. Raiser, *Rechtstatsachenforschung und Rechtsfortbildung*, in: Plett/Ziegert (Hrsg.), *Empirische Rechtsforschung zwischen Wissenschaft und Politik*, Tübingen 1984, S. 29; A. Pflüger, *Rechtstatsachenforschung in der Praxis*, in: Brand/Strempel (Hrsg.), *Soziologie des Rechts: FS Erhard Blankenburg*, Baden-Baden 1998, 561 (562).

5 H. Hamann, *Evidenzbasierte Jurisprudenz*, Tübingen 2014, S. 41; N. Petersen, *Empirische Wende* (Fn. 2), 435: „Dagegen ist die Beschäftigung mit [...] kausalen Zusammenhängen [...] sowohl in der juristischen Praxis als auch in der Rechtswissenschaft bisher nicht sonderlich stark ausgeprägt“. Zu neueren Ansätzen H. Hamann, *Evidenz, Empirie und Data Science in der Rechtswissenschaft. Ein „Scoping Review“ zur Taxonomie*

dies beispielsweise in den USA, wo das Feld der *empirical legal studies* auf einen breiten Strauß von Forschungsdesigns und Methoden zurückgreift.⁶

Rechtstatsachenforschung lässt sich über die von ihr angesprochenen Rezipienten in drei Bereiche einteilen: die rechtspolitische Tatsachenforschung wird im Kontext von Gesetzgebungsverfahren herangezogen, um als „wesentliche Entscheidungsgrundlage“⁷ darüber zu informieren, welche Rolle das Recht für die Verwirklichung sozialer Lebensverhältnisse spielt (und umgekehrt!);⁸ die rechtsdogmatische Tatsachenforschung berücksichtigt empirische Befunde als unmittelbaren Bestandteil der Rechtsanwendung, beispielsweise im Rahmen der teleologischen Normauslegung, der Ermittlung der Wortlautgrenze sowie zur empirischen Überprüfung von (Privat)Rechtsprinzipien;⁹ die rechtspraktische Tatsachenforschung zielt dagegen darauf ab, Lebenssachverhalte empirisch zu ermitteln, die für gerichtliche Einzelfallentscheidungen von Bedeutung sind. Mit letzterer befasst sich diese Untersuchung. Wenn auch wiederholt vorgeschlagen,¹⁰ finden empirische Untersuchungen als Beweismittel im Zivilprozess bisher tatsächlich wenig Verwendung.¹¹ Eine prominente Ausnahme bildet das Beweismittel der demoskopischen Befragung zur Ermittlung der Verkehrsauffassung im gewerblichen Rechtsschutz.¹²

quantitativer Rechtsforschung, in: Wörner/Wilhelmi/Glockner/et al., Digitalisierung des Rechts, Berlin 2024, 41 (44 ff.).

6 Das mag unter anderem in der dort vorherrschenden Wahrnehmung der Rechtswissenschaft als Teil der Sozialwissenschaften liegen (*legal realism*), vgl. W. Bayer, Privatrechtsdogmatik und Rechtstatsachenforschung, in: Auer/Grigoleit/Hager/et al. (Hrsg.), FS Claus-Wilhelm Canaris, Berlin 2017, 319 (322) m.w.N.

7 D. Stempel, Empirische Rechtsforschung, in Broda/Deutsch/Schreiber/Vogel, FS Rudolf Wassermann, Neuwied 1985, 223 (232).

8 Da diese Untersuchungen größtenteils aus der sozialwissenschaftlichen Forschung kommen, gilt die Kritik des eingeschränkten Methodengebrauchs hier ausdrücklich nicht.

9 W. Bayer, Privatrechtsdogmatik und Rechtstatsachenforschung (Fn. 6), 319 (321 f.) m.w.N.

10 Bspw. für die Ermittlung des Handelsbrauchs P. Oestmann, Die Ermittlung von Verkehrssitte und Handelsbräuchen, JZ 2003, 285 oder für das Transparenzgebot i.R.d. § 307 Abs. 1 S. 2 BGB A. Stöhr, Die Bestimmung der Transparenz, AcP 216 (2016), 558.

11 H. Hamann, Marken- und Wettbewerbsrecht als Vorbilder für die Vertragsauslegung, in: Christandl/Laimer/et al., Jahrbuch Junger Zivilrechtswissenschaftler, Innsbruck 2017, S. 135 (147): „... hat das Zivilrecht von der Demoskopie bisher kaum Kenntnis genommen“.

12 Dieser sei „Vorreiter für Demoskopie in der Rechtspraxis“, T. Müller, Mutmaßungen über Verkehrsauffassungen, JR 1992, 8 (11) m.w.N.

Das Lauterkeitsrecht zeigt sich im Ansatz offen gegenüber der Verwendung rechtspraktischer Tatsachenforschung. Daher untersucht dieser Beitrag, ob und inwieweit sich andere, insbesondere experimentelle, Studienformen zur Ermittlung lauterkeitsrechtlich relevanter Sachverhaltselemente (besser) eignen. Dieser Frage nachgehend soll zunächst die Verwendung demoskopischer Gutachten in lauterkeitsrechtlichen Prozessen dargestellt werden (A.). Im Anschluss werden die Erkenntnisgrenzen der demoskopischen Methode im Lauterkeitsrecht untersucht (B.), um sodann Mehrwert und Anwendungsmöglichkeiten experimenteller Untersuchungen zu erklären (C.). Der Beitrag schließt mit einer Zusammenfassung (D.).

A. Demoskopische Ermittlung lauterkeitsrechtlicher Generalklauseln

Das Lauterkeitsrecht richtet sich gegen Handlungen von Unternehmen im Wettbewerb, die sich nachteilig auf Mitbewerber, Verbraucher oder sonstige Marktteilnehmer auswirken. Konzeptionell lässt sich das maßgeblich im Gesetz gegen den unlauteren Wettbewerb (UWG) kodifizierte Lauterkeitsrecht als Deliktsrecht der Unternehmen verstehen,¹³ welches dem Interesse der Allgemeinheit am unverfälschten Wettbewerb dient, § 1 Abs. 1 S. 2 UWG.¹⁴ Das UWG enthält eine Reihe von „per se“-Verboten, die gem. § 3 Abs. 3 i.V.m. Anhang I UWG stets zur Unzulässigkeit führen, ohne dass es der Prüfung ihrer Auswirkungen bedarf. Daneben enthalten die §§ 3 – 7a UWG weitere Vorschriften, unter ihnen auch mehrere Generalklauseln, welche im Fokus dieser Untersuchung stehen.¹⁵

Darunter fallen die Verbrauchergeneralklausel (§ 3 Abs. 2 UWG) und die Verbote von aggressiven (§ 4a UWG) sowie irreführenden (§ 5, 5a UWG) geschäftlichen Handlungen. Die Vorschriften eint ihr struktureller Aufbau: Sie verbieten Handlungen,¹⁶ die ein „Unwertkriterium“¹⁷ erfüllen, das sich an der Verkehrsauffassung orientiert (I.). Zudem müssen diese sich spürbar

13 So V. Emmerich / K. Lange, Lauterkeitsrecht, 12. Aufl., München 2022, § 1 Rn. 5.

14 Daneben gibt es zahlreiche Nebengesetze, die in dieser Untersuchung jedoch unberücksichtigt bleiben.

15 § 3 Abs. 1 UWG und § 3a UWG bleiben wegen ihrer fehlenden europarechtlichen Determinierung unberücksichtigt, vgl. T. Lettl, Die geschäftliche Relevanz, WRP 2019, 1265 (1273 f.).

16 Der Begriff der geschäftlichen Handlung (§ 2 Abs. 1 Nr. 2 UWG) wird dabei in europarechtskonformer Auslegung sehr weit verstanden, vgl. H. Köhler, in: Köhler/Bornkamm/Feddersen (Hrsg.), UWG-KO, 41. Aufl., München 2023, § 2 Rn. 2.7.

17 So bspw. H. Köhler (Fn. 16), § 3 Rn. 5.16.

auf das Verkehrsverhalten auswirken („Relevanzkriterium“) (II.). Beide Kriterien werden in der gerichtlichen Praxis sowohl normativ als auch durch einen Rückgriff auf demoskopische Untersuchungen ermittelt (III.).

I. Das Unwertkriterium als Verkehrsauffassung

Das Unwertkriterium erfordert, dass das als unlauter angegriffene Verhalten gegen die „unternehmerische Sorgfalt“ verstößt.¹⁸ Die Verbrauchergeneral Klausel stellt als Auffangtatbestand unmittelbar auf dieses Kriterium ab, § 3 Abs. 2 UWG. Bei § 4a UWG und §§ 5, 5a UWG liegen dagegen spezielle Sorgfaltsverstöße durch aggressives bzw. irreführendes Geschäftsverhalten vor; der Sorgfaltsverstoß i.S.d. § 2 Abs. 1 Nr. 9 UWG wird sodann gesetzlich vermutet.¹⁹

Zur Ermittlung des Unwertkriteriums wird überwiegend auf die maßgebliche Verkehrsauffassung abgestellt. Welche Fachkenntnis und Sorgfalt vom Unternehmen erwartet werden darf, ist anhand der „berechtigten Erwartungen eines Durchschnittsverbrauchers“ zu konkretisieren;²⁰ das erforderliche unternehmerische Verhalten um eine „informationsgeleitete und freie, mithin rationale Entscheidung“ zu ermöglichen bestimmt sich aus dessen Sicht.²¹ Und auch für die Beurteilung der Irreführungsgefahr ist auf das Verständnis der angesprochenen Adressat:innen abzustellen:²² maßgeblich ist, ob die Angabe nach Auffassung des adressierten Verkehrskreises geeignet ist, in hinreichendem Umfang eine Fehlvorstellung hervorzurufen.²³ In beiden Fällen ist damit die Verkehrsauffassung für die

18 Legaldefiniert als „Standard an Fachkenntnissen und Sorgfalt“ den Verbraucher:innen vom Unternehmen billigerweise erwarten dürfen, § 2 Abs. 1 Nr. 9 UWG.

19 EuGH GRUR 2013, 1157 Rn. 37 ff.; GRUR 2015, 600; dazu auch *KH. Fezer*, in Fezer/Büscher/Obergfell (Hrsg.), UWG-KO, 3. Aufl., München 2016, § 2 Abs. 1 Nr. 7, Rn. 30 f. Rn. 176.

20 EuGH, GRUR 2016, 1180 Rn. 34 f. zum Begriff des Durchschnittsverbrauchers unten, A.III.

21 BGH NJW 2014, 2279, Rn. 23. Kritisch zu diesem Vorgehen *R. Podszun*, in Harte-Bavendamm/Henning-Bodewig (Hrsg.), UWG-KO, 5. Aufl., München 2021, § 3 Rn. 78.

22 So die st. Rspr. BGHZ 13, 244 (253), BGHZ 28, 1 (6), BGHZ 156, 250 (252), jüngst BGH GRUR 2021, 1315, Rn. 15 und die Literatur, siehe nur *H. Köhler* (Fn. 16), § 5 Rn. 1.57 ff, m.w.N.

23 *A. Lubberger*, in Gloy/Loschelder/Danckwerts (Hrsg.), Hdb. Wettbewerbsrecht, 5. Aufl., München 2019, § 40 Rn. 97 f. Dagegen sind für die Beurteilung der „Unwahrheit“ von Angaben objektive Kriterien heranzuziehen, *G. Dreyer*, in Harte-Baven-

Beurteilung des Unwertkriteriums ausschlaggebend.²⁴ Grundsätzlich sind demoskopische Untersuchungen dazu geeignet, die Anschauung und Meinung eines Verkehrskreises zu ermitteln. Etwas anderes gilt lediglich bei der Beurteilung der Aggressivität geschäftlicher Handlungen i.S.d. § 4a UWG.²⁵

II. Das Relevanzkriterium als Verkehrsverhalten

Des Weiteren muss die Realisierung des Unwertkriteriums dazu geeignet sein, „den Verbraucher oder den sonstigen Marktteilnehmer zu einer geschäftlichen Entscheidung zu veranlassen, die er andernfalls nicht getroffen hätte“, § 3 Abs. 2 i.V.m. § 2 Abs. 1 Nr. 11 Hs. 1 UWG; § 4a Abs. 1 UWG; § 5 Abs. 1 UWG; § 5a Abs. 1 Nr. 2 UWG. Dies wird als „geschäftliche Relevanz“ bezeichnet.²⁶ Betonung verdient, dass der Wortlaut des Merkmals über die hier untersuchten Generalklauseln hinweg quasi identisch ist.²⁷ Inhaltlich geht es dabei um die Frage nach dem Kontrafaktischen, nämlich dem hypothetisch-alternativen Verbraucherverhalten.²⁸ Die geschäftliche Entscheidung wird dabei als maßgeblicher Anknüpfungspunkt (§ 2 Abs. 1 Nr. 1 UWG) weit ausgelegt: Erfasst sind nicht etwa nur direkte Kaufentscheidungen, sondern auch dem Kauf vor- oder nachgelagertes Verhalten.²⁹

Das Relevanzkriterium untersucht, ob die angegriffene Geschäftshandlung sich auf das Verbraucherverhalten auswirken kann. Denn verboten ist nicht allein die „unwerte“ Geschäftspraktik als solche; diese muss sich auch auf wettbewerblich relevante Weise auf das Verbraucherverhalten aus-

damm/Henning-Bodewig (Hrsg.), UWG-KO, 5. Aufl., München 2021, § 5 Rn. 144, 309 ff.

24 Präzise P. Bähr, in Ahrens, Der Wettbewerbsprozess, 9. Aufl., Köln 2021, Kap. 26 Rn. 9: „Die Verkehrsauffassung ist nichts real existierendes, sondern eine gedankliche Konstruktion zur Erfassung eines gesellschaftlichen Phänomens, nämlich einer mehr oder minder gleichgerichteten Anschauung.“

25 Die hier erforderliche „erhebliche Eignung zur Beeinträchtigung der Entscheidungsfreiheit“ spricht jedenfalls mittelbar das potenziell hervorgerufene Verkehrsverhalten an (dazu sogleich).

26 H. Köhler (Fn. 16), § 3 Rn. 3.22; § 4a Rn. 1.36.; R. Podszun (Fn. 21), § 3 Rn. 92.

27 In der Verbrauchergeneralklausel sind lediglich „sonstige Marktteilnehmer“ nicht erfasst.

28 Dazu B. I.

29 Wie bspw. das Betreten eines Geschäfts (EuGH, GRUR 2014, 196 Rn. 35 f.) oder das Klicken auf ein Produkt zur weiteren Informationseinholung (BGH GRUR 2019, 746 Rn. 29).

wirken können.³⁰ Deren geschäftliche Entscheidung muss gerade Folge der „unwerten“ Einflussnahme auf die Entscheidungsfindung sein.³¹ Im Rahmen der Verbrauchergeneralklausel und des Verbots aggressiver Geschäftshandlungen wird die Realisierung des Unwertkriteriums damit an das hervorgerufene Verbraucherverhalten geknüpft, wobei unterschiedliche Anforderungen an die Erheblichkeit der Einflussnahme gestellt werden.³² Bei den Irreführungstatbeständen geht es dagegen um die Frage, ob und inwieweit die Irreführung geschäftliche Entscheidungen der Verbraucher:innen beeinflussen kann.

Anders als das Unwertkriterium stellt das Relevanzkriterium damit nicht auf die Verkehrsauffassung, sondern auf das hypothetische Verkehrsverhalten ab. Ob der Adressatenkreis glaubt, potenziell von der Geschäftspraktik beeinflusst werden zu können, ist nicht maßgeblich; es gilt zu ermitteln, mit welcher Wahrscheinlichkeit er sich in der konkreten Situation hätte beeinflussen lassen können. Für diese Ermittlung sind die spezifischen Umstände der Entscheidungssituation zu berücksichtigen, nicht aber das tatsächlich-individuelle Verbraucherverhalten im Einzelfall. Vielmehr ist die „typische Reaktion“ in Form des antizipierten Verkehrsverhaltens maßgeblich.³³

Damit handelt es sich um ein abstraktes Kausalitätserfordernis, das sich vom klassisch-juristischen Konzept der deterministischen Kausalität unterscheidet: ob eine konkrete Verbraucherin wegen einer konkreten Einflussnahme in einer konkreten Situation eine konkrete Entscheidung getroffen hat, ist gerade nicht fallentscheidend.³⁴ Stattdessen ist die Wahrscheinlichkeit zu ermitteln, mit der die Praktik Verbraucher:innen in ihrem Verhalten als abstrakt betroffene Marktakteure beeinflussen kann. Damit stellt das

30 G. Dreyer (Fn. 23), § 5 Rn. 334.

31 In der Literatur wird daher z.T. zwischen Einwirkung und Auswirkung der Einflussnahme unterschieden, vgl. H. Köhler (Fn. 16), § 3 Rn. 3.22. Diese Differenzierung kann jedoch für den weiteren Verlauf der Untersuchung vernachlässigt werden, da inhaltlich i.E. beide Punkte nach der objektiven Wahrscheinlichkeit der Verhaltensbeeinflussung fragen.

32 Während das Eingriffskriterium i.R.d. § 2 Abs. 1 Nr. 1 UWG als Bagatellschwelle verstanden wird, muss bei § 4a UWG ein *erheblicher* Einfluss vorliegen.

33 EuGH GRUR 2012, 1269.

34 Dies kann allenfalls eine indizielle Bedeutung entfalten. Etwas anderes gilt im Rahmen des neuen Schadensersatzanspruches für Verbraucher, § 9 Abs. 2 UWG, der den Nachweis einer Einzelfallkausalität erfordert, vgl. H. Köhler (Fn. 16), § 9 Rn. 2.23.

Erfordernis auf ein probabilistisches Kausalitätsverständnis ab, das dem der sozialwissenschaftlichen Wirkungsforschung entspricht.³⁵

In der lauterkeitsrechtlichen Entscheidungspraxis wird für die Ermittlung dieses Kriteriums überwiegend auf eine normative Prognose der Tatrichter:innen abgestellt.³⁶ Maßgeblich für deren Entscheidung ist, wie sich ein durchschnittlich informiertes, verständiges und der Situation entsprechend aufmerksames Mitglied des angesprochenen Verkehrskreises verhalten würde.³⁷ Das typische Verkehrsverhalten ist dabei anhand der Leitlinien des europäisch determinierten Leitbildes des Durchschnittsverbraucher zu ermitteln.³⁸ Auch die Ermittlung dieses Kriteriums steht der demoskopischen Untersuchung offen.³⁹

III. Rechtsdemoskopie in der lauterkeitsrechtlichen Praxis

„Demoskopie“ beschreibt die systematische Ermittlung von Meinungen oder Einstellungen in der Gesellschaft (oder einer festgelegten Zielgruppe) anhand von Umfragen.⁴⁰ Dafür wird eine repräsentative Stichprobe der relevanten Bevölkerungsgruppe ausgewählt und mit einem standardisierten Fragebogen konfrontiert.⁴¹ Die „Rechtsdemoskopie“⁴² nutzt solche Umfragemethoden zur Feststellung rechtsrelevanter Tatsachenfragen, insbesondere zur Ermittlung der Verkehrsauffassung.⁴³ Mittels demoskopischer Untersuchungen lassen sich Ansichten der Befragten sammeln und zusammenfassen. In lauterkeitsrechtlichen Prozessen wird von dieser Möglichkeit seit

35 Dazu C.I. Nur der Vollständigkeit halber sei hier darauf hingewiesen, dass dies auch nicht der Kategorie eines Gefährdungstatbestandes entspricht, welcher zwar kein Verschulden voraussetzt, aber durchaus auf einen individuellen Kausalzusammenhang abstellt.

36 Statt aller *KH. Fezer* (Fn. 19), § 2 Abs. 1 Nr. 7, Rn. 30 f. Rn. 179.

37 *G. Dreyer* (Fn. 23), § 5 Rn. 344.

38 EuGH GRUR 2012, 1269 Rn. 53.

39 Vgl. dazu B.II.2.

40 Aus dem altgriechischen δῆμος (dēmos) „Volk“, σκοπεῖν (skopeín) „spähen“.

41 Die Auswahl einer repräsentativen Stichprobe kann dabei eine erhebliche methodische Herausforderung darstellen, *N. Döring / J. Bortz*, Forschungsmethoden und Evaluation, 5. Aufl., Wiesbaden 2016, S. 292 ff., 305 ff.

42 Begriff erstmals bei *H. Eichman*, Gegenwart und Zukunft der Rechtsdemoskopie, GRUR 1999, 939.

43 Zu Anwendungsgebieten auch jenseits des Lauterkeitsrechts *A. Pflüger*, in *Gloy/Loschelder/Danckwerts* (Hrsg.), Hdb. Wettbewerbsrecht, 5. Aufl., München 2019, § 42 Rn. III ff.

langer Zeit Gebrauch gemacht. Bereits 1933 zog das Reichsgericht erstmals ein demoskopisches Gutachten heran.⁴⁴ Zwischenzeitlich sah der BGH diese unter bestimmten Voraussetzungen als zwingendes Beweismittel an.⁴⁵ Doch auch wenn Richter:innen sich mittlerweile grundsätzlich auf ihre individuelle Einschätzung der Verkehrsauffassung verlassen dürfen,⁴⁶ werden solche Gutachten in der Rechtspraxis regelmäßig als Sachverständigenbeweis⁴⁷ oder – häufiger – als qualifizierter Sachvortrag berücksichtigt.⁴⁸ Dabei befasst sich die lauterkeitsrechtliche Praxis ganz überwiegend mit der Ermittlung der Irreführungsgefahr i.R.d. §§ 5, 5a UWG.

Damit eine Studie den relevanten Ausschnitt der Realität widerspiegeln kann, müssen Stichprobenauswahl, Fragestellung, Durchführung und Auswertung methodisch korrekt erfolgen.⁴⁹ Gerade bei der Frageformulierung bestehen auch explizite Vorgaben aus der Rechtsprechung.⁵⁰ Nur eine ordnungsgemäß ausgewählte Stichprobe erlaubt die Ausdehnung der Erkenntnisse auf die maßgebliche Bevölkerungsgruppe, nur eine methodengerechte Durchführung bringt valide Ergebnisse hervor.⁵¹ Dabei gibt es eine Hand-

44 RG GRUR 1934, 76.

45 Nach der „Bärenfangdoktrin“ musste für die Ablehnung der Irreführungsgefahr stets eine demoskopische Untersuchung eingeholt werden, vgl. BGH GRUR 1963, 270 (272 f.).

46 Seit BGH GRUR 2002, 550 (552); dies gilt auch wenn sie nicht zum angesprochenen Verkehrskreis gehören, eine Einschätzung aber aufgrund beruflicher Erfahrung treffen können.

47 Aufgrund eines parteilichen Beweisantritts i.S.d. § 403 ZPO oder von Amts wegen, § 144 Abs. 1 S. 1 ZPO. Prozessuale Einzelheiten bei A. Pflüger / H. Dobel, in: Hasselblatt, MAH GewRS, 6. Aufl., München 2022, § 9 Rn. 83 – 102.

48 BGH GRUR 2011, 1043 Rn. 50. Bei korrekter Durchführung muss sich das Gericht sodann i.R.d. § 286 ZPO zwingend mit diesen auseinandersetzen, siehe nur BGH MDR 2003, 45 (46); BGH GRUR 2013, 631 Rn. 48.

49 Im Einzelnen H. Dobel, Verkehrsauffassung und demoskopische Gutachten im Marken- und Wettbewerbsrecht, Baden-Baden 2014, S. 106 ff.; instruktiv A. Pflüger (Fn. 43), § 42 Rn. 13 – 44.

50 So sei grds. eine Kombination aus offenen Fragen und nachfassenden geschlossenen Fragen zur Aufklärung erforderlich, vgl. BGH GRUR 1993, 920 (922); BGH GRUR 1992, 66 (68); BGH GRUR 1991, 680 (681). Auch sollen Fragen so formuliert sein, dass sich eine Suggestivwirkung ausschließen lässt, vgl. BGH GRUR 1989, 440 (442); BGH GRUR 1992, 70 (71).

51 F. Faulbaum, Methodische Grundlagen der Umfrageforschung, Wiesbaden 2019, S. 171 ff., 393 ff., 407 ff. Kritik an der Stichprobenauswahl ist dabei nicht selten, vgl. für das Facebook-Verfahren J. Kruse, Der Nutzer – das unbekannte Wesen, NZKart 2019, 418 (420 f.).

voll auf Rechtsdemoskopie spezialisierte Institute, die regelmäßig mit der Ausführung derartiger Untersuchungen beauftragt werden.⁵²

Im Zuge der europäischen Determinierung des Verbraucherleitbildes ist bezweifelt worden, ob empirische Erkenntnisse zu dessen Beurteilung überhaupt herangezogen werden dürfen.⁵³ Denn bei der Ermittlung von Verkehrsauffassung und Verkehrsverhalten ist das normative Leitbild des Durchschnittsverbrauchers zu berücksichtigen. Dieser ist „angemessen gut unterrichtet [...] aufmerksam und kritisch“.⁵⁴ Das frühere Leitbild des „flüchtigen Verbrauchers“, nach dem Verbraucher:innen sich nicht ausführlich mit Werbeaussagen auseinandersetzen und diese nur flüchtig wahrnehmen,⁵⁵ fand spätestens mit der „Gut Springenheide“-Entscheidung des EuGH ein Ende. Dieser stellt nunmehr auf die „mutmaßliche Erwartung eines durchschnittlich informierten, aufmerksamen und verständigen Durchschnittsverbrauchers“ ab.⁵⁶ Der BGH hat sich diesem Leitbild mit der „Orient-Teppichmuster“-Entscheidung angeschlossen.⁵⁷ Die Gültigkeit dieses Leitbildes steht für das Lauterkeitsrecht außer Frage.⁵⁸

Damit ist jedoch noch nichts darüber gesagt, wie die Verkehrsauffassung im konkreten Fall zu ermitteln ist. Zwar legt ErwGr. 18 S. 5 UGP-RL fest, dass das Europäische Verbraucherleitbild „nicht auf einer statistischen Grundlage“ beruhe. Die Frage, inwieweit daneben auch empirische Untersuchungen herangezogen werden können, spaltet die lauterkeitsrechtliche Literatur: ein Teil spricht sich für eine ausschließlich normative Ermittlung der Verkehrsauffassung durch entscheidende Richter:innen aus;⁵⁹ der andere hält eine empirische Ermittlung der Verkehrsauffassung grundsätzlich

52 Vgl. A. Pflüger / H. Dobel (Fn. 47), § 9 Rn. 141.

53 Bspw. G. Dreyer (Fn. 23), § 5 Rn. 149; I. Scherer, Normative Bestimmung von Verwechslungs- und Irreführungsgefahr, GRUR 2000, 273; HW. Micklitz / M. Namysłowska in Heermann/Schlinghoff (Hrsg.), Münchner KO UWG, 3. Aufl. München 2020, Art. 6 UGP-RL Rn. 27 ff.

54 Vgl. ErwGr. 18 S. 2 UGP-RL, zurückgehend auf EuGH GRUR Int 1998, 795 Rn. 37.

55 Soweit ersichtlich von BGH GRUR 1957, 128 bis einschließlich BGH GRUR 1995, 60 (61).

56 EuGH GRUR Int 1998, 795 Rn. 31.

57 BGH GRUR 2000, 619 (612). Umfangreiche Nachweise bei J. Bornkamm, / J. Feddersen, in Köhler/Bornkamm/Feddersen, UWG-KO, 41. Aufl., München 2023, § 5 Rn. 071.

58 Statt aller A. Lubberger (Fn. 23), § 40 Rn. 14. Der deutsche Gesetzgeber hat dieses Leitbild explizit übernommen, vgl. BT-Drucks. 16/10145 S. 40.

59 Vgl. Fn. 53. Dabei ist zweifelhaft, ob eine rein normative Ermittlung überhaupt möglich ist. Denn die Frage nach der Verkehrsauffassung lässt sich nur faktisch beantworten. Entscheidend ist eher, welche *Art von Fakten* das Gericht heranzieht:

für zweckmäßig oder gar geboten.⁶⁰ Die Darstellung als sich gegenseitig ausschließende Ermittlungsmethoden ist jedoch verkürzt. Vielmehr sollten in einem ersten Schritt die tatsächlichen Elemente der Verkehrsauffassung ermittelt werden – basierend auf Parteianhörung, richterlicher Lebenserfahrung und auch empirischen Untersuchungen – um dann in einem zweiten Schritt das europäisch-normative Verbraucherleitbild heranzuziehen.⁶¹ Der Einwand, ein solches Vorgehen sei europarechtswidrig, verfährt nicht:⁶² auch hier wird die normativ determinierte Entscheidungshoheit letztlich gewahrt; die Einholung empirischer Beweismittel erweitert oder verkürzt diesen Entscheidungsspielraum nicht.⁶³ Zudem stellt das europäische Verbraucherleitbild auch an anderen Stellen auf die tatsächlichen Marktgeschehnisse ab, sodass eine empirische Ermittlung zumindest nicht grundsätzlich systemfremd scheint.⁶⁴ Schließlich hält der EuGH selbst explizit fest, dass nationale Gerichte zur Ermittlung der Irreführungsgefahr „ein Sachverständigengutachten einholen oder eine Verbraucherbefragung in Auftrag geben“ können.⁶⁵

Auch in der Entscheidungspraxis des BGH wird die Heranziehung empirischer Erkenntnisse weiterhin als zulässig erachtet. Zwar gesteht dieser auch die Ermittlung anhand eigener Lebenserfahrung und Sachkunde zu.⁶⁶ Daneben bleibt es jedoch bei der Einordnung als Tatsachenfrage, die dem Beweis – auch dem empirischen – zugänglich ist.⁶⁷ Obwohl sich die Praxis vor allem mit der Irreführungsgefahr der §§ 5, 5a UWG befasst, lässt sich diese Erkenntnis auch auf die empirische Ermittlung weiterer Tatbestandsmerkmale übertragen.

richterliche Lebenserfahrung (anekdotische Evidenz) oder empirische Befunde? Vgl. dazu auch *J. Kruse*, *Ökonomik vor Gericht*, Baden-Baden 2024, S. 10 f., 181 ff.

60 *A. Lubberger* (Fn. 23), § 40 Rn. 24, § 41 Rn. 7 ff.; *J. Glöckner*, in *Harte-Bavendamm/Henning-Bodewig* (Hrsg.), *UWG-KO*, 5. Aufl., München 2021, Einl. B Rn. 484 f.; *HJ. Ahrens*, *Verwirrtheiten juristischer Verkehrskreise zum Verbraucherleitbild*, WRP 2000, 812 (814 f.). Instruktive Darstellung bei *H. Dobel*, *Verkehrsauffassung* (Fn. 49), S. 44–57, 63 ff.

61 So auch *A. Lubberger* (Fn. 23), § 41 Rn. 1–9; *P. Ruess*, in *Heermann/Schlingloff* (Hrsg.), *Münchner KO UWG*, 3. Aufl. München 2020, § 5 Rn. 261 ff. m.w.N.

62 So aber *G. Dreyer* (Fn. 23), § 5 Rn. 152.

63 So auch *A. Lubberger* (Fn. 23), § 41 Rn. 1, 11; *J. Glöckner* (Fn. 60), Einl. B Rn. 485.

64 *J. Glöckner* (Fn. 60), Einl. B Rn. 440 ff.; *A. Lubberger* (Fn. 23), § 41 Rn. 23.

65 EuGH GRUR Int 1998, 795 Rn. 35.

66 BGH GRUR 2004, 244 (245).

67 Vgl. *O. Sosnitza*, in *Ohly/Sosnitza*, *UWG-KO*, 8. Aufl. München 2023, § 5 Rn. 151 ff. m.w.N.

B. Erkenntniswert und -grenzen demoskopischer Untersuchungen

Demoskopische Untersuchungen sind ein hilfreiches empirisches Werkzeug, verfügen jedoch über inhärente Erkenntnisgrenzen. Aus methodischer Perspektive eignen sie sich vor allem für die Ermittlung der Verkehrsauffassung, nicht aber zur Ermittlung des Verkehrsverhaltens. Um das zu veranschaulichen, soll zunächst eine kurze Einordnung der Demoskopie in das Spektrum sozialwissenschaftlicher Methoden erfolgen (I.), bevor die konkreten Erkenntnisgrenzen im Lauterkeitsrecht genauer erläutert werden (II.).

I. Demoskopie im Methodenspektrum empirischer Sozialforschung

Allgemein wird in der empirischen Sozialwissenschaft auf methodischer Ebene zwischen **beobachtenden** und **experimentellen Studien** unterschieden.⁶⁸ Bei beobachtenden Studien greifen Forschende nicht in das Geschehen ein, sondern machen die vorhandenen Informationen lediglich verfügbar, indem sie Daten sammeln und auswerten. In experimentellen Studien geht es dagegen gerade darum, die Auswirkungen eines bestimmten Eingriffes zu untersuchen. Dafür werden die Beobachtungen in einem klar definierten Umfeld gesammelt, in dem ein Teil der beobachteten Einheiten dem Eingriff („treatment“) unterliegt, der andere Teil jedoch nicht („control“).⁶⁹

Im Hinblick auf die Zielsetzung empirischer Forschung wird zudem zwischen **beschreibenden** und **kausalen Studien** differenziert.⁷⁰ Diese unterscheiden sich hinsichtlich des mit ihnen verfolgten Erkenntnisinteresses. Beschreibende Studien abstrahieren Beobachtungen eines konkreten Lebenssachverhalts und fassen diese anhand bestimmter Kennzahlen (z.B. dem Durchschnittswert) oder grafischer Darstellungen zusammen. Damit beschreiben sie Tatsachen oder Phänomene in der Population, ohne notwendigerweise die zugrundeliegenden Ursachen zu identifizieren. Kausale Studien zielen dagegen auf die Identifikation genau dieser Ursachen ab und

68 Instruktiv H. Hamann, Evidenzbasierte Jurisprudenz (Fn. 5), S. 137 ff. m.w.N.

69 N. Döring / J. Bortz, Forschungsmethoden (Fn. 41), S. 193 f.

70 Regelmäßig wird zwischen beschreibendem, erklärendem und erkundendem Erkenntnisinteresse differenziert, vgl. N. Döring / J. Bortz, Forschungsmethoden (Fn. 41), S. 193 f. Dabei lässt sich das erkundende Erkenntnisinteresse als Unterfall des Beschreibenden verstehen, mit dem sich neuen Phänomenen initial angenähert wird.

ersuchen, die Wirkungsbeziehungen zwischen verschiedenen Variablen zu identifizieren.⁷¹

Damit lässt sich der aus kausalen Studien ermittelte Erkenntniswert als weitreichender verstehen:⁷² Einerseits nehmen diese den Schritt vom Beschreiben zum Erklären, indem sie zugrunde liegende Wirkungsmechanismen offenlegen. Andererseits befähigen sie dazu, wahrscheinlichkeitsbasierte Angaben über zukünftige oder hypothetische Geschehensabläufe zu treffen. Denn wenn die kausale Auswirkung einer Intervention feststeht, lassen sich Einschätzungen über hypothetische Geschehensabläufe treffen. Anders formuliert: beobachtende Studien beschreiben das Faktische; kausale Studien ermöglichen eine Einschätzung über das Kontrafaktische, also den Zustand, der sich nicht beobachten lässt.

Vor diesem Hintergrund drängt sich die Frage auf, wo sich die Demoskopie im Koordinatensystem von Studienform (beobachtend/experimentell) und Erkenntnisinteresse (beschreibend/kausal) verorten lässt. Die Antwort vermag Sozialwissenschaftler:innen und Jurist:innen zu versöhnen, denn: es kommt darauf an. Im Ausgang ist die Demoskopie lediglich ein Werkzeug zur Datenerhebung.⁷³ Die Studienform und das damit verfolgte Erkenntnisinteresse hängen von der konkreten Ausgestaltung der Untersuchung ab. Bei einer üblichen Meinungsumfrage beantworten Proband:innen einen standardisierten Fragebogen. Die so erfassten Ergebnisse werden quantifiziert und statistisch aufbereitet, um einen Überblick über die erteilten Antworten zu geben, der sich idealiter auf die untersuchte Population ausweiten lässt.⁷⁴ Im Grundsatz lässt sich die Demoskopie damit als beobachtende Untersuchung mit beschreibendem Erkenntnisinteresse einstufen.

Darüber hinaus können jedoch auch kausale Erkenntnisinteressen verfolgt werden, wenn der Aufbau der Befragung der Logik randomisierter Kontrollgruppenstudien folgt.⁷⁵ Dafür werden Proband:innen zufallsbasiert

71 N. Döring / J. Bortz, Forschungsmethoden (Fn. 41), S. 194 f.

72 Dazu D. Berg-Schlusser, Epistemologische Grundlagen, in: Wagemann/Goerres/Siewert (Hrsg.), Hdb. Methoden der Politikwissenschaft, Wiesbaden 2019, S. 1 – 18.

73 Eine Darstellung, die der Komplexität der methodischen Anforderungen gerecht wird, kann hier nicht erfolgen. Zur Vertiefung N. Döring / J. Bortz, Forschungsmethoden (Fn. 41), S. 356 ff. und 398 ff.

74 Bereits diese Reinform der Demoskopie ist nie *ausschließlich* beobachtend, weil unmittelbar mit dem Forschungssubjekt interagiert wird. Klassischerweise versucht die demoskopische Forschung aber, den so ausgeübten Einfluss möglichst gering zu halten, vgl. ebenda, S. 360 ff.

75 Näher dazu C.I.

in mehrere Gruppen eingeteilt, die in den untersuchungsrelevanten Aspekten unterschiedliche, ansonsten jedoch identische Fragebögen erhalten. Neben der beobachtenden Informationserhebung durch den Fragebogen können damit auch kausale Schlussfolgerungen gezogen werden: Wenn die Antworten der Proband:innen in dem abweichend formulierten Befragungsteil signifikant von der Kontrollgruppe abweichen, lässt dies Rückschlüsse auf die kausale Wirkung der Formulierung zu. Dieser Mehrwert experimenteller Befragungsformen ist der (rechts)demoskopischen Literatur nicht unbekannt.⁷⁶ Im rechtsdemoskopischen Kontext wird diese als „Kontrollgruppentest“⁷⁷ oder „Splitbefragung“⁷⁸ bezeichnet. Bereits 1967 regte der BGH eine Prüfungspraxis in Form von Kontrollgruppentests an,⁷⁹ ließ dies aber in seiner weiteren Rechtsprechung unberücksichtigt. Zwar finden sich immer wieder vereinzelt Urteile, in denen Gutachten mit Kontrollgruppentests durchgeführt wurden⁸⁰ und auch in der Literatur wird dies vereinzelt gefordert.⁸¹ Solche Experimentalumfragen haben sich in der rechtsdemoskopischen Praxis jedoch nicht ersichtlich durchgesetzt – vielmehr scheint die Begutachtungspraxis von einfachen Informationsabfragen dominiert zu werden.⁸² In der amerikanischen Rechtsdemoskopie sei die Verwendung von Kontrollgruppen dagegen „seit Ende der 80er Jahre [...] allgemeiner Konsens“.⁸³

76 A. Niedermann / E. Noelle, Die Bedeutung der gegabelten Befragung mit Kontrollgruppen, in Keller/Passmann/Falck (Hrsg.), FS Winfried Tilmann, Köln 2003, 857.

77 A. Niedermann, in Fezer/Büscher/Obergfell, UWG-KO, 3. Aufl., München 2016, S 18 Rn. 50.

78 A. Lubberger (Fn. 23), § 41 Rn. 40.

79 BGHZ 46, 305.

80 Bspw. BGH GRUR 1987, 535 (538); OLG Nürnberg, NJOZ 2001, 2169.

81 E. Noelle-Neumann / C. Schramm, Umfrageforschung in der Rechtspraxis, Weinheim 1961, 41 ff.; H. Raffée, Verkehrsauffassung und Verbrauchererwartung aus Sicht der empirischen Marktforschung, ZLR 1991, 281 (290); C. Gerstenmaier, Umfrageforschung in der wettbewerbsrechtlichen Praxis, Wiesbaden 2003, S. 128 f.

82 Eine empirische Auswertung der verwendeten Studiendesigns liegt dem Verf. nicht vor. Wenn aber der Darstellungsumfang in der (zu großen Teilen von Angehörigen rechtsdemoskopischer Institute verfassten) Kommentarliteratur einen Rückschluss auf die Praxis zulässt, ist von einer untergeordneten Bedeutung auszugehen, vgl. die überblicksartigen Darstellungen bei A. Pflüger (Fn. 43), § 42 Rn. 152; A. Pflüger / H. Dobel (Fn. 47), § 9 Rn. 35–37; Brüning, in Harte-Bavendamm/Henning-Bodewig (Hrsg.), UWG-KO, 5. Aufl., München 2021, § 12 Rn. 197; H. Köhler (Fn. 16), § 12 Rn. 1.83; P. Ruess (Fn. 61), § 5 Rn. 267. Ausführlicher bei A. Niedermann (Fn. 77), S 18 Rn. 50–64 m.w.N.

83 A. Niedermann (Fn. 77), S 18 Rn. 54.

II. Erkenntnisgrenzen demoskopischer Untersuchungen im Lauterkeitsrecht

Dass die rechtsdemoskopische Praxis in Deutschland überwiegend auf beobachtende demoskopische Untersuchungen abstellt, ist aus methodologischer Perspektive zu kritisieren. Denn dies führt regelmäßig dazu, dass das mit ihr verfolgte Erkenntnisinteresse – die Verkehrsauffassung und das hypothetische Verkehrsverhalten – nicht vom Erkenntnisgehalt der Untersuchungen gedeckt ist. Bereits bei der Feststellung des Irreführungsgehalts zielt die Logik einfacher Befragungsmethoden eigentlich auf ein kausales Erkenntnisinteresse ab, das nur einen mittelbaren Schluss auf die Verkehrsauffassung gestattet (1.). Besonders prägnant ist diese Diskrepanz im Rahmen des Relevanzkriteriums („zu einer geschäftlichen Entscheidung zu veranlassen, die er andernfalls nicht getroffen hätte“), das durch beobachtende Untersuchungen grundsätzlich nicht ermittelt werden kann (2.).

1. Demoskopische Irreführungsprüfung

Die rechtsdemoskopische Praxis im Lauterkeitsrecht befasst sich vor allem mit der Ermittlung der Irreführungsgefahr geschäftlicher Handlungen. Im Rahmen der Befragung wird ermittelt, welche Vorstellungen eine bestimmte Äußerung bei Angehörigen des relevanten Verkehrskreises hervorruft.⁸⁴ Die so ermittelten Angaben werden dann anhand des objektiven Wahrheitsgehaltes bewertet: deckt sich die subjektive Wahrnehmung nicht mit der Beschaffenheit, liegt eine Irreführungsgefahr vor; deckt sie sich, besteht diese Gefahr nicht. Die sog. Irreführungsquote (das Ergebnis der Befragung) beschreibt den Anteil der befragten Personen, die die infragestehende Äußerung unrichtig auffassen.⁸⁵ Wenn also 30 % der Befragten ein Verständnis preisgeben, das sich nicht mit den objektiven Tatsachen deckt, wird der geschäftlichen Handlung eine Irreführungsquote von 30 % attestiert. Dieses Vorgehen ruft grundlegende methodische Zweifel hervor.

Einerseits lässt sich bei *jeder* Umfrage ein gewisses Maß an Irrtum oder Verwirrung feststellen – die sog. Konfusionsrate.⁸⁶ Da die Befragten

84 Zur Ermittlung dieses Verkehrskreises A. Pflüger / H. Dobel (Fn. 47), § 9 Rn. 20 – 30.

85 A. Niedermann (Fn. 77), S 18, Rn. 65.

86 E. Noelle-Neumann / C. Schramm, Umfrageforschung (Fn. 81), S. 59 f.; A. Niedermann (Fn. 77), S 18 Rn. 66 geht bei Umfragen in der Gesamtbevölkerung von einer Konfusionsrate von 5 – 10 % aus.

heterogen sind (unterschiedlichen Alters- und Berufsgruppen, unterschiedliche kommunikative und kognitive Fähigkeiten), ist es nur konsequent, dass standardisierte Fragenformulierungen zu unterschiedlichen Interpretationen führen und auch Fehlvorstellungen verursachen können. Zudem können sichtbare und nicht-sichtbare Verhaltensweisen der Interviewenden einen Einfluss auf das Antwortverhalten haben („Versuchsleitereffekt“).⁸⁷ Dies gilt insbesondere bei persönlichen Befragungen in Interviews, die bei rechtsdemoskopischen Untersuchungen den Regelfall bilden.⁸⁸ All diese Umstände wirken sich auf die gemessene Irreführungsquote aus. Sie können den Unternehmen jedoch nicht zur Last gelegt werden und sollten daher bei der lauterkeitsrechtlichen Bewertung unberücksichtigt bleiben. Zum Teil wird dem mit Forderungen nach höheren Irreführungsquoten begegnet.⁸⁹ Dieses Vorgehen lässt sich aber eher als pauschalisierte Daumenregel verstehen. Eine genauere Berechnung der auf die geschäftliche Handlung zurückgehenden Irreführung wäre dagegen mittels Kontrollgruppenbefragungen möglich: Konfusionsrate und Versuchsleitereffekte sind sowohl in der Experimental- als auch der Kontrollgruppe zu erwarten. Die Differenz der Quoten beider Gruppen würde daher eine genauere Einschätzung der Irreführungswirkung der untersuchten Angabe ermöglichen.⁹⁰

Zudem lässt sich bezweifeln, ob das Erkenntnisziel einer Irreführungsprüfung überhaupt von einer bloßen Beobachtung erfasst werden kann. Zwar lässt sich die Verkehrsauffassung grundsätzlich im Wege der Demoskopie ermitteln. Bei der Ermittlung der Irreführungsquote wird jedoch nicht lediglich beobachtend eine Meinung erhoben, sondern das Ergebnis der Auseinandersetzung mit einer konkreten Aussage erfasst. Hierfür wer-

87 Etwa, indem sie Befragten (unterbewusst) das Gefühl vermitteln, eine bestimmte Antwort sei „richtiger“, vgl. *M. Häder*, Empirische Sozialforschung, 3. Aufl., Wiesbaden 2015, S. 222 ff. Beispiele bei *A. Niedermann* (Fn. 77), S 18, Rn. 63.

88 *A. Pflüger* (Fn. 43), § 42 Rn. 71 ff, 43.

89 Dies lässt sich in jüngeren BGH-Entscheidungen beobachten, scheint seine Ursache aber überwiegend in der Anpassung an das europäische Verbraucherleitbild zu haben, vgl. *O. Sosnitza* (Fn. 67), § 5 Rn. 157 f.

90 Als Beispiel OLG Nürnberg, NJOZ 2001, 2169: Hier ging es um die Frage, ob der Zusatz „am Stadtpark“ eine besonders ruhige Lage eines Hotels suggeriere. In der Experimentalgruppe („Hotel Garni am Stadtpark“) nahmen 31 % der Befragten eine besonders ruhige Lage an. Jedoch wurde dieser Eindruck auch bei 15 % der Kontrollgruppe („Hotel Garni Maier“) erweckt. Ohne Kontrollgruppe hätte man von einer Irreführungsquote von 31 % ausgehen müssen. Die eigentliche Auswirkung lag aber in der Differenz (16 %).

den Proband:innen mit einer konkreten geschäftlichen Handlung konfrontiert und nach ihrer Interpretation befragt. Damit wird aber nicht die Verkehrsauffassung im Sinne einer individuellen Ansicht oder Meinung erhoben („Wie **finden** Sie diese Aussage?“), sondern der Irreführungserfolg als Ergebnis einer geistigen Auseinandersetzung („Wie **verstehen** Sie diese Aussage?“) gemessen.⁹¹ Streng genommen geht das Erkenntnisziel der Untersuchung damit aber über eine bloße Beobachtung hinaus und untersucht den *Einfluss* einer vermeintlich irreführenden Angabe. Damit zielt die Untersuchung im Kern darauf ab, den Effekt einer Intervention zu prüfen – das Erkenntnisziel ist also ein kausales. Sodann besteht jedoch Grund zur Kritik aus methodischer Perspektive: Denn wenn die Auswirkung der Intervention nur anhand einer Experimentalgruppe ermittelt wird, ohne eine Kontrollgruppe hinzuzuziehen, fehlt es an einem neutralen Referenzpunkt zur Beurteilung des Einflusses. Es gibt keine Informationen über die Vorstellungen, die eine neutralere Angabe ohne das vermeintlich irreführende Element bei Adressat:innen hervorgerufen hätte. Ein derartiges Vorgehen kann daher dem eigentlich erforderlichen Erkenntniswert nicht gerecht werden.

Beide Kritikpunkte können adressiert werden, indem statt rein beobachtender demoskopischer Untersuchungen Kontrollgruppentests verwendet werden. Einerseits kann damit methodisch konsequent die tatsächliche (kausale) Auswirkung der infrage stehenden geschäftlichen Handlung überprüft werden. Andererseits können durch den Vergleich von Experimental- und Kontrollgruppe die nicht zu berücksichtigenden Einflüsse auf die Irreführungsquote – wie Konfusionsrate und Versuchsleitereffekte – herausgerechnet werden und somit eine präzisere Angabe über die tatsächlich verursachte Irreführung ermöglichen. Eine konsequente Berücksichtigung experimenteller Methodik sollte mithin dazu führen, im Rahmen demoskopischer Ermittlungen *stets* eine Kontrollgruppe hinzuzuziehen.⁹²

91 Das ist nicht *per se* fernliegend: Der Irreführungserfolg lässt einen Rückschluss auf die Irreführungsgefahr zu: je mehr Personen getäuscht werden, desto größer ist auch die Täuschungsgefahr. Zudem würde die Abfrage einer Selbsteinschätzung („Finden Sie diese Äußerung irreführend?“) regelmäßig ungenaue Ergebnisse erzielt, dazu so gleich.

92 So auch H. Raffée, Verkehrsauffassung und Verbrauchererwartung (Fn. 81), 281 (290). In der Tendenz zustimmend A. Niedermann (Fn. 77), S 18 Rn. 53 f.

2. Demoskopische Relevanzprüfung

Regelmäßig wird in der Rechtsprechung von dem Vorliegen einer hinreichend hohen Irreführungsquote auf die geschäftliche Relevanz der Irreführung geschlossen, sodass es hierfür keiner gesonderten Beweiserhebung bedarf.⁹³ Etwas anderes gilt, sofern begründete Zweifel an der Relevanz bestehen.⁹⁴ Eine empirische Feststellung der Relevanz entfaltet daher einerseits für Anspruchsteller:innen Bedeutung, wenn mangels hinreichend hoher Irreführungsquoten kein Rückschluss auf die geschäftliche Relevanz gestattet ist. Andererseits kann dies für Anspruchsgegner:innen hilfreich sein, um trotz hinreichend hoher Irreführung Zweifel an der geschäftlichen Relevanz zu erregen. In der rechtsdemoskopischen Praxis wird daher im Anschluss an die Ermittlung des Unwertkriteriums grundsätzlich auch deren geschäftliche Relevanz ermittelt.⁹⁵ Dies geschieht in der Regel durch die explizite Abfrage, ob die konkrete geschäftliche Handlung dazu geeignet sei, die geschäftliche Entscheidung der befragten Person zu beeinflussen, bzw. ob diese „irgendwie wichtig“ für die Kaufentscheidung sei.⁹⁶ Der BGH hat sich vermehrt zu der konkreten Formulierung dieser Fragestellung geäußert, jedoch die Ermittlung der geschäftlichen Relevanz mittels Umfragen an sich nicht infrage gestellt.⁹⁷ Dabei ist gerade die demoskopische Ermittlung des Relevanzkriteriums aus methodischer Perspektive zu kritisieren.⁹⁸ Erneut lassen sich zwei Kritikpunkte identifizieren.

Erstens wird bei der Abfrage der Bedeutung der geschäftlichen Handlung für die eigene Kaufentscheidung eine erhebliche Introspektion der angesprochenen Verbraucher:innen gefordert, die von Befragten in der Regel jedoch kaum zu leisten sein wird.⁹⁹ Konsumententscheidungen setzen sich aus einer Vielzahl verschiedener Verhaltens- und Entscheidungsmotive zusammen, die sich der bewussten Kontrolle des Individuums weitgehend

93 BGH GRUR 1981, 71 (73); GRUR 1991, 215; BGH GRUR 1991, 852 (855).

94 BGH GRUR 1993, 920 (922); BGH GRUR 1992, 70 (72).

95 Zumindest bei Proband:innen, die irreführt worden sind, vgl. A. Pflüger (Fn. 43), § 42 Rn. 142; A. Niedermann (Fn. 77), S 18 Rn. 43.

96 BGH GRUR 1987, 535 (537); BGH WRP 1998, 1002 (1005).

97 Bspw. BGH GRUR 1981, 71 (73 f.); BGH GRUR 1987, 535 (537).

98 Im Ergebnis ähnlich deutlich W. Tilmann, Die Verkehrsauffassung im Wettbewerbs- und Warenzeichenrecht, GRUR 1984, 716 (720): „Ungeeignet für Umfragen ist schließlich in aller Regel die [...] wettbewerbliche Relevanz“.

99 Dazu auch J. Kruse, Ökonomik vor Gericht (Fn. 59), S. 29 f.

entziehen.¹⁰⁰ Zahlreiche Untersuchungen der Verhaltensforschung legen nahe, dass Menschen ihr hypothetisches Verhalten oder die dafür maßgeblichen Faktoren regelmäßig nicht wahrheitsgetreu einschätzen können.¹⁰¹ Zudem kann eine unmittelbare Abfrage der Entscheidungsrelevanz erhebliche Suggestiveffekte haben.¹⁰² Die von der demoskopischen Abfrage der geschäftlichen Relevanz hervorgebrachten Ergebnisse sind daher potenziell kaum aussagekräftig. Diese Einwände haben sowohl bei einer direkten Abfrage („*Hat das Merkmal zur Kaufentscheidung beigetragen?*“) als auch bei einer indirekten Abfrage („*Ist das Merkmal für Sie wichtig?*“) Bestand.

Daneben besteht ein zweiter Kritikpunkt, der in der rechtlichen Literatur bisher keine ersichtliche Aufmerksamkeit gefunden hat: Die Befragung als empirisches Werkzeug eignet sich schon vom methodischen Ansatz her nicht zur Ermittlung des hypothetischen Verkehrsverhaltens. Denn als beschreibende Studienform werden hiermit nur Meinungsbilder sichtbar gemacht. Das Relevanzkriterium ist aber – ganz explizit – eine Frage nach dem Kontrafaktischen: Ob die Geschäftshandlung „zu einer geschäftlichen Entscheidung“ veranlasst, die „andernfalls nicht getroffen“ worden wäre, erfordert letztlich die Einschätzung eines hypothetischen Kausalverlaufes. Will man diese Einschätzung auf empirischen Erkenntnissen basieren, kommen nur kausale Studien in Betracht. Denn in der realen Welt kann für jedes Individuum nur *einmal* das Verhalten in einer konkreten Situation wahrgenommen werden. Wie sich dessen Entscheidung in einer hypothetischen Alternativsituation gestaltet hätte, entzieht sich logischerweise dem Bereich des Beobachtbaren und kann daher im Wege der Beschreibung schlicht nicht ermittelt werden. In der Ökonometrie spricht man in diesem Zusammenhang vom „Fundamentalproblem kausaler Inferenz“.¹⁰³ Um auf dieses Problem zu reagieren, müssten Untersuchungen mit kausalem Erkenntnisinteresse eine „Identifikationsstrategie“ verfolgen. Zwar gibt es auch Identifikationsstrategien, die aus Beobachtungsdaten kausale Inferenz-

100 M. Schmidt, in: Ahrens, Der Wettbewerbsprozess, 9 Aufl., Köln 2021, Kap. 27 Rn. 31 m.w.N.

101 Grundlegend R. Peterson / R. Kerin, The Quality of Self-reported Data, Review of Marketing 1981, 5.; R. Fischer / J. Katz, Social-Desirability Bias and the Validity of Self-Reported Values, Psychology & Marketing 2000, 105.

102 Sinngemäß, wenn auch auf die konkrete Formulierung abstellend, BGH GRUR 1981, 71 (73 f.).

103 Grundlegend D. Rubin, Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies, Journal of Educational Psychology 1974, 688.

ziehung zulassen;¹⁰⁴ im Kontext des Relevanzkriteriums kommt jedoch vor allem eine Identifikation anhand experimenteller Studien in Betracht, da es regelmäßig an Beobachtungsdaten für alternative Geschäftspraktiken fehlt. Eine methodisch konsequente empirische Ermittlung des Relevanzkriteriums müsste daher auf randomisierte Kontrollgruppenstudien abstellen (dazu sogleich), welche Verbraucherverhalten vergleichen. Damit ließe sich das Erkenntnisziel der probabilistischen Wahrscheinlichkeit hypothetischer Kausalverläufe überhaupt erst ermitteln.

Auch der (fehleranfälligen) Selbsteinschätzung von Proband:innen bedarf es dann nicht mehr. Die Auswirkung auf geschäftliche Entscheidungen lässt sich experimentell offenlegen, ohne dass es die Aufdeckung der individuellen Entscheidungsprozesse voraussetzt. Im Rahmen des Relevanzkriteriums geht es nicht um die *Verkehrsauffassung* über die potenzielle Relevanz, sondern um das hypothetische *Verkehrsverhalten*. Um diesem Problem methodisch konsequent zu begegnen, müssten experimentelle Kontrollgruppenstudien durchgeführt werden, die das Verhalten der Proband:innen vergleichen. Im weiteren Verlauf soll daher untersucht werden, inwieweit experimentelle Erkenntniswerkzeuge zur Ermittlung des Relevanzkriteriums in der praktischen Rechtstatsachenforschung fruchtbar gemacht werden können.

C. Ergänzung des Methodenkanons rechtspraktischer Tatsachenforschung

Um darzulegen, wie experimentelle Tatsachenforschung im Rahmen der Ermittlung des Relevanzkriteriums berücksichtigt werden kann, sollen zunächst die zugrunde liegenden Erkenntnisstrategien randomisierter Kontrollgruppenstudien Darstellung finden (I.). Im Anschluss wird erklärt, wie Ergebnisse dieser Studienformen zur Beurteilung der Relevanz beachtet werden können (II.). Es lassen sich Einwände gegen den Vorschlag antizipieren, die es zu adressieren gilt (III.). Schließlich wird kurz auf die besondere Eignung des experimentellen Ermittlungsansatzes im Kontext digitaler Geschäftspraktiken eingegangen (IV.).

104 Zugängliche Einleitung bei J. Angrist / J.-S. Pischke, *Mostly Harmless Econometrics*, Princeton 2009; ausführlich L. von Auer, *Ökonometrie*, 6. Aufl., Heidelberg 2013.

I. Die Logik randomisierter Kontrollgruppenstudien¹⁰⁵

Unter den Strategien zur Ermittlung kausaler Wirkungszusammenhänge gilt die randomisierte Kontrollgruppenstudie als „Goldstandard“.¹⁰⁶ Dabei werden Teilnehmer:innen zufällig („randomisiert“) in mindestens zwei Gruppen eingeteilt, wodurch Selektionseffekte vermieden werden. Eine Gruppe, die Kontrollgruppe, gilt als Ausgangspunkt; die andere Experimentalgruppe, erfährt eine Intervention. Abgesehen von dieser Intervention müssen die Umstände in der Kontroll- und Experimentalgruppe gleich sein („*ceteris paribus*“-Annahme). Sofern sich zwischen den beiden Gruppen eine statistisch signifikante¹⁰⁷ Abweichung im Verhalten beobachten lässt, kann innerhalb dieses Forschungsdesigns mit hinreichender Wahrscheinlichkeit angenommen werden, dass eine solche Abweichung *kausal* von der Intervention verursacht wurde.

Regelmäßig werden derartige Studien in einem Verhaltenslabor vorgenommen, um den Entscheidungskontext so neutral wie möglich zu halten. Solche Laborexperimente zeichnen sich durch ein erhöhtes Maß an Abstraktion aus. Statt sich auf eine konkrete Entscheidungssituation zu konzentrieren, wird menschliches Entscheidungsverhalten in einem möglichst neutralen Umfeld untersucht, um externe Einflüsse (sog. Störvariablen) so gering wie möglich zu halten. Kehrseite dieses Vorgehens ist, dass die künstlichen Situationen regelmäßig Zweifel an der Übertragbarkeit auf tatsächliche Lebenssachverhalte hervorrufen.¹⁰⁸ Damit eignet sich dieses Verfahren also vor allem für die Überprüfung von Verhaltenstheorien: Wenn eine bestimmte Intervention bereits in einem neutralen Umfeld nicht die vorhergesagte Wirkung entfaltet, so wird sie dies erst recht nicht in einer konkreten Lebenssituation tun, in der zusätzliche Einflussquellen vorliegen.

105 Hier kann lediglich ein Überblick über die zugrundeliegende *ratio* des Erkenntniswerkzeuges vermittelt werden. Detailliertere Darstellung der Einzelheiten bei N. Döring / J. Bortz, *Forschungsmethoden* (Fn. 41), S. 205 ff. Einleitung für Jurist:innen bei K. Chatziathanasiou / M. Leszcynska, *Experimentelle Ökonomik und Recht*, RW 2017, 314 (317).

106 Siehe nur L. Bothwell / J. Greene / S. Podolsky / D. Jones, *Assessing the Gold Standard – Lessons from the History of RCTs*, in *New England Journal of Medicine* 2016, 21275.

107 Zur Signifikanz unter C.II.

108 Aus juristischer Perspektive bspw. H. Fleischer / U. Schmolke / D. Zimmer, *Verhaltensökonomik als Forschungsinstrument für das Wirtschaftsrecht*, in *Fleischer/Zimmer, Beiträge der Verhaltensökonomie*, Frankfurt a.M. 2011, S. 1 (45 f.).

Randomisierte Kontrollgruppentests lassen sich jedoch auch außerhalb des Labors vornehmen, beispielsweise in sog. Feldexperimenten.¹⁰⁹ Hier wird das Verhalten bestimmter Individuen in der tatsächlichen Lebenswelt untersucht – die so ermittelten Befunde lassen sich einfacher auf das Verhalten in der echten Welt übertragen. Für die Durchführung eines Feldexperimentes wird ein Teil der Bevölkerung einer Intervention ausgesetzt, ein anderer jedoch nicht. In der Regel ist den Versuchspersonen dabei nicht bewusst, dass sie an einem Experiment teilnehmen; es handelt sich um „echte Beobachtungen“ im „echten Leben“. Hier fällt es jedoch schwerer, für potenzielle Störvariablen zu kontrollieren. Zudem muss „Spillover“-Effekten vorgebeugt werden: Es könnte das beobachtete Verhalten verfälschen, wenn Mitglieder der Kontrollgruppe erfahren, dass eine Experimentalgruppe existiert, die anders behandelt wird (und *vice versa*). In der Regel erfordern Feldexperimente einen besonders hohen zeitlichen und finanziellen Aufwand.¹¹⁰ Weniger aufwändig sind Feldexperimente jedoch im digitalen Kontext, beispielsweise wenn unterschiedlichen Nutzern einer Internetseite abweichende Nutzungsoberflächen angezeigt werden (dazu unter IV.)

Dabei lassen sich Laborexperimente (maximale Kontrolle) und Feldexperimente (maximale Übertragbarkeit) als zwei Enden eines Spektrums verstehen, in dem mehrere Zwischenstufen existieren. Die unterschiedlichen Ausgestaltungsformen haben jeweils eigene Vor- und Nachteile und sind je nach der konkreten Forschungsfrage auszuwählen.¹¹¹ Als Ergebnis produzieren randomisierte Kontrollstudien eine Einschätzung über den *durchschnittlichen* kausalen Effekt einer Intervention. Unter bestimmten Voraussetzungen lassen sich die so gewonnen Erkenntnisse auf die Gesamtbevölkerung ausdehnen.¹¹² Von statistischem Interesse ist dabei regelmäßig die Differenz der Durchschnittswerte zwischen Kontroll- und Experimentalgruppe: Treffen beispielsweise 20 % der Kontrollgruppe eine bestimmte Entscheidung, die in der Experimentalgruppe zu 70 % getroffen wird, liegt

109 J. Weimann / J. Brosig-Koch, Einführung in die experimentelle Wirtschaftsforschung, Berlin 2019, S. 33 f., 37.

110 S. Levitt / J. List, Field Experiments in Economics, European Economic Review 2009, 1 – 18.

111 J. List / S. Sadoff / M. Wagner, So you want to run an experiment, now what?, Experimental Economics 2011, 439.

112 Dies erfordert eine repräsentative Stichprobe, Generalisierbarkeit des experimentellen Umfelds, Robustheit und Konsistenz der Ergebnisse und die Kontrolle für mögliche Störvariablen, vgl. N. Jacquemet / O. L'Haridon, Experimental Economics, Cambridge 2018, S. 243 ff.

der kausale Effekt in der Erhöhung ebendieser Entscheidungsfindung um 50 %-Punkte.¹¹³

Damit soll deutlich werden, dass das Kausalitätsverständnis der Sozialforschung ein probabilistisches ist. Es werden Wahrscheinlichkeitsangaben ermittelt, inwieweit sich ein Einfluss auf das Verhalten auswirkt. Damit unterscheidet dieses sich vom deterministischen Kausalitätsverständnis im Sinne einer *Conditio-sine-qua-non*-Formel.¹¹⁴ Zum Nachweis einer solchen Einzelfallkausalität eignen sich empirische Studien daher nicht unmittelbar; sie können allenfalls für die Bildung von Erfahrungssätzen in der rechtlichen Entscheidungspraxis herangezogen werden.¹¹⁵

II. Verwendung zur Ermittlung des Relevanzkriteriums

Vor diesem Hintergrund wird deutlich, warum die experimentelle Untersuchung geschäftlicher Handlungen im Kontext des Relevanzkriteriums besonders geeignet ist. Denn auch hier wird im Kern nach einem probabilistischen Kausalitätsverständnis gefragt. Entscheidend ist nicht der Einzelfall, sondern inwieweit die geschäftliche Handlung das Entscheidungsverhalten der Marktgegenseite im Allgemeinen beeinflussen kann. Die Ergebnisse experimenteller Kausalitätsforschung liefern genau diesen Einblick, indem sie eine probabilistische Einschätzung der Auswirkung einer Intervention auf die untersuchte Gesamtbevölkerung erlauben. Aus erkenntnistheoretischer Perspektive wären sie daher das passende Werkzeug zur Ermittlung des Relevanzkriteriums.

Sodann stellt sich die Frage, wie sich solche Ergebnisse im Rahmen einer gerichtlichen Entscheidung berücksichtigen ließen. Die Ergebnisse solcher Untersuchungen bestehen, vereinfacht formuliert, aus Effektgrößen und deren Signifikanz.¹¹⁶ Die Effektgröße bezeichnet den Unterschied im

113 Darauf basierend kann die genaue Effektstärke berechnet werden. Deren Ermittlung hängt vom konkreten Studiendesign ab, vgl. *N. Jacquemet / O. L'Haridon*, *Experimental Economics*, Cambridge 2018, S. 59 ff. Verbreitet ist z.B. die Darstellung durch *Cohen's d*, vgl. Fn. 122.

114 Vertiefend *F. Schauer / B. Spellmann*, *Probabilistic Causation in the Law*, in: Engel/Schweitzer (Hrsg.), *Causality in the Law and in the Social Sciences*, 4 (6–9).

115 *B. Beutel*, Möglichkeiten und Grenzen von Erfahrungssätzen, WRP 2017, 513 (517 f.). Zur Berücksichtigung experimenteller Erkenntnisse im Lauterkeitsrecht allgemein *M. Leistner*, *Behavioral Economics und Lauterkeitsrecht*, ZGE 2009, 3 (31 ff.).

116 *J. Weimann / J. Brosig-Koch*, *Experimentelle Wirtschaftsforschung* (Fn. 109), S. 254 ff.

beobachteten Verhalten zwischen Kontroll- und der Experimentalgruppe. Die Signifikanz bezeichnet das Maß an Sicherheit, mit der ein beobachteter Effekt nicht auf einer zufälligen Datenvariation beruht und sich somit auf die repräsentierte Grundgesamtheit verallgemeinern lässt. Sofern keine methodischen Bedenken hinsichtlich der Durchführung bestehen und eine hinreichende Signifikanz vorliegt,¹¹⁷ kann die entscheidende Richterin die ermittelte Effektgröße heranziehen, um zu beurteilen, ob eine geschäftliche Relevanz der untersuchten Einflussnahme vorliegt. Zwar bestehen bei jedem signifikanten Ergebnis Anzeichen dafür, dass die geschäftliche Handlung das Verhalten der Marktgegenseite beeinflusst. Hinsichtlich der juristischen Bewertung sollte jedoch zwischen drei Bereichen von Effektgrößen unterschieden werden.

Theoretisch denkbar, wenn auch praktisch unwahrscheinlich, wäre die Ermittlung eines negativen Effekts (< 0). Dies würde bedeuten, dass die untersuchte Intervention die Wahrscheinlichkeit des beobachteten Verhaltens verringert. Untersucht man den Einfluss einer geschäftlichen Handlung auf die Kaufentscheidung, würde diese bedeuten, dass das Verhalten des Unternehmers dazu führt, dass *weniger* Adressat:innen einen Kauf abschließen. Zwar beinhaltet auch dies im Ergebnis eine Eignung, den Verbraucher zu einer „Entscheidung zu veranlassen, die er andernfalls nicht getroffen hätte“ (denn auch das Ablassen vom Kauf ist eine geschäftliche Entscheidung). Im Lichte des Schutzzweckes des Lauterkeitsrechts sollte dies jedoch nicht die in §§ 8 ff. UWG normierten Rechtsfolgen nach sich ziehen dürfen.¹¹⁸

Daneben muss notwendigerweise zwischen hinreichend großen und kleineren Effektgrößen differenziert werden. Zwar bedeutet auch ein sehr kleiner Effekt, dass die adressierte Population von der Handlung „zu einer geschäftlichen Entscheidung“ veranlasst wird. Allerdings gebietet es der Verhältnismäßigkeitsgrundsatz, dass nicht jede noch so kleine Einflussnahme zur Feststellung der Unlauterkeit führen kann. Denn wenn eine geschäftliche Handlung zwar einen erheblichen Anteil der Adressat:innen in die Irre führt, dies jedoch zu einer kaum spürbaren Veränderung des Marktverhaltens führt, ist der Unrechtsgehalt der hier untersuchten lauterkeitsrechtlichen Verbotsnormen gerade nicht erfüllt.¹¹⁹ Das gilt insbesondere deshalb, weil sich durch die Erhöhung der Stichprobengröße auch die

117 In der empirischen Sozialforschung hat sich ein Signifikanzniveau von 5 % als Konvention weitestgehend etabliert.

118 Ähnlich im Kontext der Irreführung P. Ruess (Fn. 61), § 5 Rn. 218 f.

119 J. Bornkamm, / J. Feddersen (Fn. 57), § 5 Rn. 0.8 f.

Wahrscheinlichkeit erhöht, ein statistisch signifikantes Ergebnis zu erzielen.¹²⁰ Bei einem statistisch signifikanten aber nicht hinreichend großem Effekt wäre die geschäftliche Relevanz damit abzulehnen.

Nur bei einem statistisch signifikanten Effekt in hinreichender Größe kann angenommen werden, dass das Marktverhalten durch die geschäftliche Handlung spürbar beeinflusst wird und sich der Unrechtsgehalt der lauterkeitsrechtlichen Normen erfüllt. Die Festlegung einer starren Grenze verbietet sich dabei: es kommt grundsätzlich auf die konkreten Umstände des Einzelfalles an. Dies liegt insbesondere daran, dass sich die Eignung zur Marktbeeinträchtigung erst in der Gesamtbetrachtung mit dem festgestellten Unwertkriterium ergibt: bei einer hohen Irreführungsquote (i.e.: die relative Anzahl getäuschter Verbraucher) würde schon ein kleiner Effekt (i.e.: die Wahrscheinlichkeit der Verhaltensänderung) zu einer spürbaren Beeinträchtigung des Marktverhaltens führen.

Da aber die geschäftliche Relevanz bei hohen Irreführungsquoten ohnehin vermutet wird,¹²¹ wäre ein empirischer Nachweis vor allem dann von Bedeutung, wenn eine niedrigere Irreführungsquote vorliegt. Hier dürfte eine höhere Effektgröße erforderlich sein. Wenn nur eine geringe Anzahl von Adressat:innen getäuscht wird, müssten diese überwiegend zu einer geschäftlichen Handlung veranlasst werden, um das Verkehrsverhalten spürbar zu beeinträchtigen. Einen anfänglichen Richtwert könnten dabei die in der empirischen Sozialforschung konventionell etablierten Bewertungen der Effektgröße darstellen.¹²²

Zur Klarstellung sei hier noch erwähnt, dass dies eine demoskopische Ermittlung des Unwertkriteriums nicht ersetzt. Denn die geschäftliche Relevanz erfordert gerade eine Veränderung des Verhaltens *infolge* der irreführenden Handlung. Auch bei der umfragebasierten Ermittlung des Relevanzkriteriums wird deswegen nur auf diejenigen Proband:innen abgestellt, die zuvor eine unzutreffende Vorstellung äußerten.¹²³ Daher sollte auch im Rahmen der experimentellen Ermittlung des Relevanzkriteriums

120 Nicht zuletzt deshalb ist *a priori* eine Teststärkenberechnung (power calculation) durchzuführen, vgl. N. Döring / J. Bortz, Forschungsmethoden (Fn. 41), S. 105.

121 BGH GRUR 1993, 920 (922). Dieser Rückschluss lässt sich indes durchaus kritisieren; Irreführung und Relevanz stellen inhaltlich unterschiedliche Fragen dar.

122 Bspw. *Cohen's d*, das insbesondere für t-Tests verwendet wird (Differenz der Mittelwerte geteilt durch die gemeinsame Standardabweichung). Dabei wird ein Wert von 0,5 bis 0,8 als mittlerer und ein Wert von über 0,8 als starker Effekt anerkannt. Zurückgehend auf J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2. Aufl., New Jersey 1988, S. 8 ff.

123 BGH GRUR 1995, 354 (357); OLG Köln MD 1998, 1269 (1273).

erhoben werden, ob die Proband:innen einer Fehlvorstellung unterliegen. Denn andernfalls ließe sich nicht darlegen, dass gerade die irreführenden Adressat:innen auch diejenigen sind, die ihr Verhalten verändert haben. Eine solche Ermittlung kann im Anschluss an die experimentelle Untersuchung mittels eines Fragebogens erfolgen.

III. Einwände

Hinsichtlich des Untersuchungsgegenstandes (tatsächliches Entscheidungsverhalten) und des Erkenntnisziels (probabilistische Kausalitätseinschätzung) eignen sich experimentelle Untersuchungen also konzeptionell und methodisch besonders gut zum Nachweis lauterkeitsrechtlicher Relevanzkriterien. Das soll jedoch mitnichten bedeuten, dass diese Ermittlungsform auch in jeder Konstellation geeignet ist oder gar zwingend herangezogen werden sollte. Zwar verfangen die hier zu erwartenden rechtstheoretischen Einwände nicht ohne weiteres (1.), jedoch gibt es überzeugende rechtspraktische Einwände (2.).

1. Rechtstheoretische Einwände

Da zum Teil bereits die Zulässigkeit demoskopischer Beweismittel infrage gestellt wird, lassen sich vergleichbare Einwände auch gegenüber der Heranziehung experimenteller Untersuchungen antizipieren. Ähnlich den Vorbehalten zur Demoskopie¹²⁴ ließe sich auch im Rahmen des Relevanzkriteriums die Ansicht vertreten, es handele sich um eine reine Rechtsfrage, deren Beurteilung allein auf einer individuell-normativen Einschätzung der Tatrichter:innen beruhe.¹²⁵ Eine empirische Ermittlung des Verhaltens widerspreche dem europäisch determinierten Leitbild des „Durchschnittsverbrauchers“ (ErwGr. 18 UGP-RL). Hierfür wird regelmäßig auf die „Purely Creative“-Entscheidung des EuGH hingewiesen, nach der sich nationale Gerichte „bei der Beurteilung der Frage, wie der Durchschnittsverbraucher in einem gegebenen Fall typischerweise reagieren würde, auf ihre eigene Urteilsfähigkeit verlassen“ müssen.¹²⁶

124 Oben, A.III.

125 So z.B. G. Dreyer (Fn. 23), § 5 Rn. 349.

126 EuGH GRUR 2012, 1269 Rn. 53.

Dass es letztlich einer normativen Entscheidung bedarf, soll hier nicht infrage gestellt werden. Vorschlag dieser Untersuchung ist nicht, diese normative Entscheidung durch experimentelle Untersuchungen zu *ersetzen*. Dies wäre schon nicht in allen Konstellationen zweckmäßig oder überhaupt möglich (dazu 2.). Jedoch können empirische Erkenntnisse in bestimmten Konstellationen *ergänzend* herangezogen werden, um Tatrichter:innen in ihrer Entscheidung zu unterstützen. Im Kontext der Demoskopie wird dies vor allem dann befürwortet, wenn Kernaussagen zur Unternehmensleistung angegriffen werden¹²⁷ oder Tatrichter:innen sich nicht zum angesprochenen Verkehrskreis zählen¹²⁸ und sich eine verallgemeinerungsfähige Einschätzung der Verkehrsauffassung wegen mangelnder Spezialkenntnisse nicht zutrauen. In all diesen Konstellationen ließe sich auch eine experimentelle Ermittlung des hypothetischen Verbraucherverhaltens befürworten. Wenn das angegriffene geschäftliche Verhalten zum Kerngeschäft des Unternehmens gehört und dessen Verbot erhebliche wirtschaftliche Konsequenzen nach sich zögen, kann es durchaus angemessen sein, empirische Grundlagen ergänzend heranzuziehen. Ebenfalls wird es Tatrichter:innen, die selbst nicht zum Adressatenkreis gehören und keine einschlägige berufliche Spezialkenntnis haben regelmäßig schwerfallen, das hypothetische Verhalten der Marktgegenseite einzuschätzen. In solchen Konstellationen bliebe allein die Spekulation über das Verkehrsverhalten; sodann sollte die Heranziehung experimenteller Studien als belastbarere Alternative nicht nur möglich, sondern geboten sein.

Wie bereits im Rahmen der Demoskopie dargestellt (A.III.), widerspräche dies auch nicht europarechtlichen Vorgaben. Die Rechtsprechung des EuGH lässt sich so verstehen, dass eine empirische Ermittlung die normative Beurteilung der Tatrichter:innen zwar nicht ablösen soll, eine ergänzende Heranziehung aber grundsätzlich möglich sei. Zwar spricht der EuGH in diesem Zusammenhang explizit nur von „Sachverständigengutachten [...] oder [...] Verbraucherbefragungen.“¹²⁹ Jedoch ist nicht ersichtlich, warum für die Heranziehung eines anderen – methodisch besser geeigneten – empirischen Werkzeugs strengere Maßstäbe gelten sollten. Da der BGH auch die Heranziehung demoskopischer Untersuchungen gestattet, wäre eine Zulässigkeit experimenteller Untersuchungsmethoden daher nur konsequent.

127 A. Lubberger (Fn. 23), § 41 Rn. 30.

128 BGH GRUR 2002, 550 (552); jünger auch BGH GRUR-RR 2014, 201 (205).

129 EuGH GRUR Int 1998, 795 Rn. 35, Hervorhebungen vom Verf.

2. Rechtspraktische Einwände

Weitere Einwände lassen sich aus rechtspraktischer Perspektive erwarten. Bereits im Zusammenhang mit demoskopischen Untersuchungen wird darauf hingewiesen, dass diese wegen des finanziellen und zeitlichen Aufwands regelmäßig nicht prozessökonomisch seien.¹³⁰ Damit kommen demoskopische Untersuchungen vor allem für Hauptsacheverfahren mit höheren Streitwerten in Betracht. Jedoch können auch zur Glaubhaftmachung im einstweiligen Rechtsschutz (der im Lauterkeitsrecht traditionell eine große Rolle spielt) „Blitzumfragen“ herangezogen werden, deren Erstellung bis zu zwei Wochen beansprucht.¹³¹ Nichtsdestotrotz wird die Einholung eines demoskopischen Gutachtens insbesondere in kleineren Wettbewerbsfällen für die beteiligten Parteien wegen der damit verbundenen Kosten regelmäßig nicht prozessökonomisch sein.¹³²

Diese Einwände sind grundsätzlich anzuerkennen und gelten erst recht im Zusammenhang mit experimentellen Studien, die hier als fallbasierte Untersuchungen konkreter Geschäftspraktiken diskutiert wird. In einem Laborexperiment, in dem es gerade um die Neutralisierung der Entscheidungsumgebung und um die Abstraktion der Befunde geht, kann dies in aller Regel bereits nicht erreicht werden. Zwar können die so ermittelten abstrakten Verhaltenserkenntnisse im Wege von Erfahrungssätzen berücksichtigt werden. Ein Verhaltensexperiment in neutralem Umfeld kann jedoch keine unmittelbaren Aussagen über die geschäftliche Relevanz einer konkreten Geschäftspraktik ermöglichen, da die Umstände der individuellen Entscheidungssituation bewusst ausgeklammert werden.

Will man aber die reale Entscheidungssituation unmittelbar berücksichtigen, bedarf es der Durchführung eines Feldexperiments. Nur hier lässt sich das Verbraucherverhalten in der tatsächlichen Entscheidungssituation beobachten, so wie es vom Relevanzkriterium gerade erfordert wird. Feldexperimente verursachen jedoch regelmäßig einen besonders großen zeitlichen und finanziellen Aufwand, sodass sie regelmäßig kaum für einzelne Gerichtsverfahren realisierbar sind. Will man auf diesem Weg die kausale Auswirkung einer geschäftlichen Handlung überprüfen, müsste eine kon-

130 Demoskopische Untersuchungen kosten i.d.R. 15.000 bis 50.000 EUR und erfordern eine Dauer von 6 Wochen bis zu mehreren Monaten, vgl. A. Pflüger (Fn. 43), § 42 Rn. 6–8; A. Niedermann (Fn. 77), S 18 Rn. 16.

131 Ebenda.

132 A. Lubberger (Fn. 23), § 41 Rn. 30.

krete Entscheidungssituation einmal mit und einmal ohne die infrage stehende Geschäftspraktik nachgebildet werden. Geht es beispielsweise um die Frage, ob der Aufdruck einer italienischen Flagge auf einer Brotverpackung geschäftliche Relevanz entfaltet,¹³³ so könnte der Artikel in zufällig ausgewählten Filialen durch eine Verpackung ohne diesen Aufdruck ersetzt werden. Durch den Vergleich der Absatzvolumina zwischen „Experimentalfilialen“ und den „Kontrollfilialen“ (in denen das Produkt weiter mit Italienflagge angeboten würde) ließe sich die Auswirkung dieses Elements auf die geschäftliche Entscheidung ermitteln. Sodann müssten Kund:innen nach Kaufabschluss im Rahmen einer demoskopischen Befragung zu ihren Vorstellungen hinsichtlich der Herkunft des Produkts befragt werden, um festzustellen, ob auch eine Irreführung vorlag.

Der mit einem solchen Vorgehen verbundene Aufwand wird jedoch regelmäßig außerhalb des prozessökonomisch Möglichen liegen. Zudem ließen sich potenzielle Störvariablen und „Spillover“-Effekte hier nur schwer ausschließen: Kund:innen können das Produkt bereits vor dieser Anpassung in der Originalfassung wahrgenommen haben oder andernorts mit einer Werbung mit Italienbezug konfrontiert worden sein. Ein solches Vorgehen verspräche daher trotz erheblichen Aufwands nicht zwingend eindeutige Ergebnisse. Regelmäßig wird eine experimentelle Ermittlung des Relevanzkriteriums in der analogen Welt daher aus methodischen und prozessökonomischen Gründen nicht in Betracht kommen.

IV. Ausblick: Digitale Geschäftspraktiken als geeigneter Anwendungsfall

Etwas anderes gilt jedoch für digitale Geschäftspraktiken, insbesondere die Gestaltung von Nutzungsoberflächen. Diese eignen sich besonders gut zur Durchführung einzelfallbasierter Feldexperimente. Denn die grafischen Benutzeroberflächen, mit denen die Nutzer:innen sich konfrontiert sehen, können ohne großen Aufwand angepasst und auf Zufallsbasis angezeigt werden. Damit ließe sich – im Sinne eines „echten“ Feldexperimentes – das tatsächliche Verbraucherverhalten in Reaktion auf die konkrete Gestaltungspraktik beobachten und mit einer Kontrollgruppe vergleichen. Beispielsweise könnte die Website einmal mit und einmal ohne ein vermeintlich irreführendes Informationselement angezeigt oder mit und ohne einem

133 Entfernte Anlehnung an BGH GRUR 2016, 406.

vermeintlich aggressiven Pop-Up Fenster. Bei der in Reaktion darauf zu beobachtenden Systeminteraktion ist zu berücksichtigen, dass der BGH den Begriff der geschäftlichen Entscheidung weit auslegt: Bereits das Klicken auf bestimmte Elemente einer Website zur weiteren Informationseinholung über ein Produkt wird als geschäftliche Entscheidung eingestuft.¹³⁴ In diesem Kontext könnte der experimentelle Ermittlungsansatz also einen veritablen Erkenntnisvorteil für rechtlich relevantes Adressatenverhalten ermöglichen, ohne einen unvertretbaren Ermittlungsaufwand zu verursachen.

Erkennbar liegt die Möglichkeit, solche Untersuchungen durchzuführen, vor allem bei den Unternehmen, die die konkret angegriffenen Gestaltungsformen verwenden. Diesen wird es leichtfallen, auf ihrer eigenen Website eine randomisierte Kontrollstudie über die Auswirkung der Oberflächengestaltung vorzunehmen. Die Anspruchsinhaberin wird hierzu regelmäßig nicht ohne weiteres in der Lage sein, da sie keinen Zugriff auf die Oberflächengestaltung der Anspruchsgegnerin hat. Die Simulation der entsprechenden Entscheidungssituation in einem Laborkontext wäre nur mittelbar (als Erfahrungssatz) aufschlussreich, da sie nicht im konkreten Entscheidungsumfeld erfolgen.

Zudem lässt sich davon ausgehen, dass viele Unternehmen ohnehin bereits genauere Kenntnisse über der Auswirkungen ihrer Designelemente besitzen, weil sie derartige Feldexperimente bereits vorgenommen haben: Im Bereich des digitalen Marketings entspricht es der gängigen Praxis, vor maßgeblichen Anpassungen des Designs sog. A/B-Tests durchzuführen.¹³⁵ Dabei werden Nutzer:innen (in der Regel basierend auf randomisierter Zuteilung) mit unterschiedlichen Oberflächengestaltungen konfrontiert und ihr Verhalten wird im Hinblick auf eine relevante Zielmetrik ausgewertet, wie der Rate weiterführender Klicks, der Verweildauer auf der Website oder der Anzahl von Vertragsabschlüssen.¹³⁶

Die für die Beurteilung des Relevanzkriteriums entscheidungserheblichen Informationen liegen den Unternehmen daher regelmäßig bereits vor oder können mit vertretbarem Aufwand erhoben werden. Daher wird eine Beweiserbringung der Anspruchsgegnerin leichter fallen als der Anspruchstellerin. Vor diesem Hintergrund scheint zumindest eine sekundäre Darle-

134 BGH GRUR 2019, 746 Rn. 29.

135 M. Witzel, Quick Guide A/B-Testing, Wiesbaden 2021, S. 2 f.

136 Überblick und weitere Zielmetriken bei F. Casalegno, A/B Testing: A Complete Guide to Statistical Testing, 2021, abrufbar auf www.towardsdatascience.com.

gungslast geboten.¹³⁷ Es überrascht daher kaum, dass sich zunehmend auch Stimmen für eine Beweislastumkehr in digitalen Lauterkeitsfällen aussprechen.¹³⁸

Dabei kommt insbesondere dem Bereich digitaler Geschäftspraktiken absehbar eine wachsende Bedeutung zu. Die wirtschaftliche Tragweite digitaler Interaktionen ist im Bereich von Privattransaktionen in den letzten Jahren erheblich gestiegen.¹³⁹ Unter Stichworten wie „Dark Patterns“,¹⁴⁰ „Behavioral Microtargeting“¹⁴¹ und „Digitale Manipulation“¹⁴² erfahren unlautere Geschäftspraktiken im digitalen Bereich jüngst gesteigerte Aufmerksamkeit. Gerade in diesem Bereich können experimentelle Untersuchungen ein hilfreiches Instrument darstellen, um juristische Entscheidungen mit realisierbaren und methodisch validen empirischen Befunden zu unterstützen.

D. Zusammenfassung

Die rechtspraktische Tatsachenforschung befasst sich mit der empirischen Ermittlung von Lebenssachverhalten, die in konkreten Rechtsstreitigkeiten von entscheidungserheblicher Bedeutung sind. Im Lauterkeitsrecht hat dieses Vorgehen Tradition: Vor allem im Rahmen der Irreführungsprüfung (§§ 5, 5a UWG) findet die Ermittlung der maßgeblichen Verkehrsauffassung regelmäßig demoskopisch, also anhand von Meinungsumfragen, statt. Daran hat sich auch durch die europäische Determinierung des „durchschnittlichen Verbrauchers“ nichts geändert. Zwar ist dieses Leitbild grundsätzlich normativ zu ermitteln; regelmäßig scheint die zusätzliche Heran-

137 Vgl. bereits für die analoge Geschäftswelt H. Köhler, Zur „geschäftlichen Relevanz“, WRP 2014, 259 (261 f.).

138 Vgl. N. Helberger / O. Lynkey / HW. Micklitz / et al., EU Consumer Protection 2.0, 2021 (75 ff., 109, 118), abrufbar auf annualreport.beuc.eu; F. Esposito / TMC. Ferreira, Addictive Design as an Unfair Commercial Practice, European Journal of Risk Regulation 2024, 1 (10).

139 Das Nettovolumen digitaler Kaufumsätze betrug 2009 noch 15,6 Mrd., 2021 bereits 86,7 Mrd. EUR, Institut für Handelsforschung Köln, Online Monitor 2023, S. 6, abrufbar unter www.einzelhandel.de.

140 Q. Weinzierl, Dark Patterns als Herausforderung für das Recht, NVwZ – Extra 2020, 1; M. Martini / I. Kramme / P. Seeliger, „Nur noch 30 Minuten verfügbar“, VuR 2022, 123.

141 M. Ebers, Beeinflussung und Manipulation von Kunden, MMR 2018, 423.

142 M. Denga, Digitale Manipulation und Privatautonomie, ZfDR 2022, 229.

ziehung empirischer Befunde zur richterlichen Entscheidungsfindung jedoch geboten.

Dies verdient im Ansatz Zustimmung, bietet in der Durchführung aber auch Anlass zur Kritik: Einerseits wäre es methodisch konsequenter, bereits im Rahmen der Irreführungsprüfung grundsätzlich eine Kontrollgruppe heranzuziehen, um die irreführende Wirkung präziser offenzulegen. Zudem wird auch das Relevanzkriterium regelmäßig im Wege der Befragung ermittelt. Inhaltlich stellt dieses Kriterium jedoch auf eine probabilistische Einschätzung der Beeinflussung des Verkehrsverhaltens ab; als Frage nach dem Kontrafaktischen lässt sich dies nicht allein im Wege von Beobachtungen beantworten, sondern bedarf aus methodischer Sicht einer experimentellen Ermittlungsform. Dazu sind insbesondere randomisierte Kontrollgruppenstudien gut geeignet. Weil dabei die Umstände der konkreten Entscheidungssituation zu berücksichtigen sind, bieten sich vor allem Feldexperimente als Ermittlungswerkzeug an. Der damit verbundene zeitliche und finanzielle Aufwand lässt sich für die analoge Durchführung jedoch regelmäßig als Realisierungshindernis verstehen. Bei digitalen Geschäftspraktiken sind derartige Untersuchungen dagegen mit geringerem Aufwand möglich. Insbesondere in diesem wirtschaftlich zunehmend bedeutenden Bereich sollte daher künftig eine experimentelle Ermittlung des Relevanzkriteriums erwogen werden.

