

# GRUR

## Gewerblicher Rechtsschutz und Urheberrecht

Zeitschrift der Deutschen  
Vereinigung für  
gewerblichen Rechtsschutz  
und Urheberrecht

[www.grur.de](http://www.grur.de)

Horst-Peter Götting	<b>Der Begriff des Geistigen Eigentums</b>	353
Peter W. Heermann	<b>Ambush-Marketing anlässlich Sportgroßveranstaltungen</b>	359
Anne Niedermann	<b>Empirische Erkenntnisse zur Verkehrsdurchsetzung</b>	367
Volker Deutsch	<b>Der BGH-Beschluss zur unberechtigten Schutzrechtsverwarnung und seine Folgen für die Praxis</b>	374
Andreas Klein/Alexander Insam	<b>Telefonische Abwerbung von Mitarbeitern am Arbeitsplatz und im Privatbereich nach neuem UWG</b>	379
Andreas Schrell/Nils Heide	<b>Zu den Grenzen des „product-by-process“-Patentanspruchs im Erteilungs- und Verletzungsverfahren</b>	383
EuGH	<b>Eintragungsfähigkeit einer im Eintragungsstaat nicht beschreibenden Wortmarke – Matratzen Concord/Hukla</b>	411
BGH	<b>Prüfungspflichten der Presse bei Schlankheitsmittelanzeigen – Schlank-Kapseln</b>	429

### 5/2006

Seiten 353 bis 440 · 108. Jahrgang · Mai 2006

**Verlag C. H. Beck**  
München · Frankfurt am Main

## Empirische Erkenntnisse zur Verkehrsdurchsetzung

Anne Niedermann\*

*„Verkehrsdurchsetzung“ ist eine der häufigsten Begutachtungsaufgaben aus dem Markenrecht für Umfrageforscher. Es besteht Anlass, etablierte Abfragemodelle und Wertungsverfahren neu zu betrachten und weiter an der Verständigung zwischen „demoskopischer und juristischer Denkungsart“ (Elisabeth Noelle, in: Noelle-Neumann/Schwarzenauer, Markenartikel 1973, S. 331) zu arbeiten. Aus demoskopischer Sicht sind einige aktuelle Entwicklungen in der Rechtsprechung zu beleuchten, bei denen es um die Beurteilung der Tragweite einwandfrei erhobener demoskopischer Befunde zur Durchsetzung geht. Ausgangspunkt werden die BGH-Urteile „Kinder“ und „Telekom“ sein. Es wird auch zu prüfen sein, wie das HABM bisher dem Umfragebeweis gegenübersteht, welche Vorstellungen dort von „guten“ und „schlechten“ Umfragen herrschen. Bewährte Messinstrumente sollten sich auch auf europäischer Ebene durchsetzen.*

### I. Das etablierte Fragemodell zur Verkehrsdurchsetzung

Das Abfragemodell, das sich hier in Deutschland durchaus sinnvoll etabliert hat, ist ein Dreischritt, d. h. es

werden drei Beweiselemente zu einer Frageserie integriert<sup>1</sup>: (Schritt 1) Allgemeine Bekanntheit des Zeichens. (Schritt 2) Der Kern des Beweises ist die anschließende Ermittlung der Markenfunktion in Form von Herstellerzuordnung. Unter Angabe des Warenbereiches ist hierbei zu klären: Geht von dem Zeichen ein Herstellerhinweis aus, in welchem Maß wird das Testobjekt exklusiv nur einer einzigen verantwortlichen Herkunftsquelle zugeordnet? Der dritte Schritt, traditionell immer mit dabei, ist die Identifikation der Herkunftsquelle bzw. des Herstellers, am häufigsten über Namensnennung. Alternative In-

---

\* Dr., Institut für Demoskopie Allensbach, Leitung Umfrageforschung für die Rechtspraxis. – Der Beitrag beruht auf einem Vortrag anlässlich des 3. Kölner Symposiums zum Wettbewerbsrecht am 10. 3. 2005 in Köln.

1) Zum Dreischritt-Modell s. z. B. Pflüger, GRUR 2004, 654 ff.; Niedermann/Schneider, sic! 2002, 821; Eichmann, GRUR 1999, 944; Synnerstad, marknadsundersökningar som bevismedel i varumärkesrättsliga mål och ärenden, Stockholm 1992; Schulz, Markenartikel 1984, 143; Noelle-Neumann/Schramm, GRUR 1966, 70.

diktoren wie Beschreibung sonstiger Produkte des Herstellers oder des Zielpublikums sind durchaus möglich.

Das Modell sei auf ein Beispiel angewandt. Angenommen, es sei die Durchsetzung eines Bildmotivs zu prüfen, einer Trachtenfigur auf einer Bierflasche. Zuerst wird die Bekanntheit unter Personen, die wenigstens gelegentlich Bier trinken oder kaufen, ermittelt: „*Hier sehen Sie einen Bierdeckel. Haben Sie schon einmal Bier oder Bierdeckel gesehen mit diesem Bild hier darauf? Ich meine, kennen Sie dieses Bild, oder sehen Sie es gerade zum ersten Mal?*“ Wer das Bild kennt, wird weiter befragt, wem das Bild unbekannt ist, scheidet aus. Dann erfolgt im zweiten Schritt die Kernfrage: „*Wenn Sie dieses Bild hier sehen, wissen Sie dann schon gleich, ob das von einer ganz bestimmten Biermarke oder Brauerei ist, oder kann man das allein von dem Bild her nicht sagen, kommen dafür mehrere, ganz verschiedene Brauereien in Frage?*“ Angepasst an die Vorantwort „*von einer ganz bestimmten*“ oder aber „*von mehreren, ganz verschiedenen Brauereien*“ folgt im dritten Schritt entweder die Nachfrage zur Herstelleridentifikation im Singular: „*Von welcher Biermarke oder Brauerei?*“ oder im Plural „*An welche Biermarken oder Brauereien denken Sie?*“ Der entscheidende Prozentsatz wird errechnet aus der Zuordnung zu einer ganz bestimmten Brauerei (Anteil der Personen, die das Zeichen kennen und es gleichzeitig mit nur einem einzigen Unternehmen verbinden) in Abhängigkeit von der Bekanntheit. Als ein zusätzlicher Pluspunkt wäre es zu werten, wenn auf die Nachfrage hin (Schritt 3) noch ein ansehnlicher Prozentsatz diese Biermarke, diese Brauerei tatsächlich identifizieren kann. Diese Frage stellt sicher, dass diejenigen, die bei Frage 2 mit „ja“ antworten, verschiedene Hersteller meinen.

### 1. Begriffsklärungen

Beim Bezug auf demoskopische Befunde zur Verkehrsdurchsetzung herrscht in Kommentarliteratur, Urteilen, Entscheidungen und auch vielen Schriftsätzen derzeit Begriffsverwirrung. Es wird ungenügend zwischen den Ebenen des dreistufigen Abfragemodells unterschieden. Infolgedessen ist vielfach unklar, um was es jeweils geht – um Bekanntheit, um den Herstellerhinweis oder um die Identifikation? Es sind in Literatur und Rechtsprechung eine Vielzahl unterschiedlicher Bezeichnungen für diese drei Ebenen im Umlauf, beispielsweise in der BGH-Entscheidung „Telekom“. Diese Entscheidung aus 2004 befasste sich mit der Verwechslungsgefahr zwischen den Firmenschlagworten „Telekom“ und „01 051 Telecom“. Eine Verwechslungsgefahr wurde im Ergebnis verneint. Dass die Bezeichnung „Telekom“ eine gesteigerte Kennzeichnungskraft hat, wollte die Klägerin mit Hilfe eines Umfragegutachtens nachweisen. Soweit sich die Entscheidung damit befasst, ist die Begriffsverwendung unklar. Zuerst heißt es dort, aus dem Privatgutachten „*ergebe sich eine Verkehrsbekanntheit der Kennzeichnung ‚Telekom‘ von 75 Prozent. Die Zuordnung der Bezeichnung ‚Telekom‘ zur Klägerin nehmen nach dem Gutachten aber nur 60 Prozent der befragten Verkehrskreise vor*“<sup>2</sup>. Mit 60 Prozent ist eindeutig die namentliche Identifikation des Markeninhabers gemeint. Weiter unten im Urteilstext heißt es dann aber: „*Der Bekanntheitsgrad bei den allgemeinen Verkehrskreisen lag demnach bei 60 Prozent, einen höheren Bekanntheitsgrad hat die Klägerin nicht dargelegt*.“ Das ist unzutreffend. Es existierte ja mindestens der weitere Prozentsatz von 75 Prozent. Dieser Wert, 75 Prozent, war der Grad der abstrakten Herstellerzuordnung (Schritt 2 im Abfragemodell). Die Bekanntheit muss noch höher gelegen haben als bei 75 Prozent. Nur

Schritt 3 im Abfragemodell, die Identifikationsrate, die in dieser Entscheidung zum wichtigsten Kriterium avancierte, betrug 60 Prozent. In der BGH-Entscheidung „Kinder“ wird ebenfalls ein Umfragebefund, der sich eindeutig auf die namentliche Identifikation bezieht, als „Bekanntheitsgrad“ bezeichnet<sup>3</sup>. Auch dort wurde allein der Grad zutreffender Identifikation in den Mittelpunkt der Wertung gestellt. In beiden Fällen wurden die erzielten Werte als zu niedrig eingestuft.

### 2. Überzogene und realistische Erwartungen an die Ergebnisse

Die Begriffsverwirrung wirkt sich problematisch auf die Wertung aus und führt zu überzogenen Ansprüchen an Umfragebefunde. Zur Rechtfertigung von Höchstwerten zur Verkehrsdurchsetzung von deutlich über 80 Prozent wird etwa vielfach und unsachgemäß die „Avon“-Entscheidung herangezogen, eine ältere Entscheidung des BGH zum Beweisthema berühmte Marke<sup>4</sup>. Dort wird ein demoskopisch ermittelter Wert von 80 Prozent erwähnt, außerdem von anderen Marken erzielte Vergleichswerte. Diese Werte zur Rechtfertigung von Höchstansprüchen bei Durchsetzung heranzuziehen, beruht auf einer Gleichsetzung von Bekanntheitsgrad und Herstellerzuordnung, einer Ebenenverwirrung. Außerdem ist das Abfragemodell beim Nachweis einer bekannten Marke grundlegend anders als beim Nachweis von Durchsetzung. Die Höhe der ermittelten Prozentwerte ist nicht vergleichbar. Man kann also von Ergebnissen zur bekannten Marke keine Ansprüche an Werte ableiten, die nach dem – für die Befragten schwierigeren – Abfragemodell zur Messung von Durchsetzung erzielt wurden.

Es wird übersehen, dass der Umfrage, die dem „Avon“-Urteil zu Grunde liegt, wegen des grundlegend anderen Abfragemodells gar kein Prozentwert zur Herstellerzuordnung zu entnehmen ist. Ein solcher wurde bei dem „Avon“-Gutachten überhaupt nicht ermittelt, sondern die Umfrage erhob bloß die allgemeine Bekanntheit der Bezeichnung. Damit lassen sich keine Anforderungen an Herstellerzuordnungen begründen. Der BGH bezieht sich in dieser Entscheidung zwar mehrfach richtig auf den nachgewiesenen „Bekanntheitsgrad“, spricht aber unter Bezug auf die Ergebnisse zur Bekanntheit auch von „Durchsetzungsgrad“.

Es wäre viel gewonnen, wenn man sich auf drei Stichworte einigt und diese Begriffe zur Bezeichnung der drei Ebenen des Abfragemodells in Zukunft bewusst und eindeutig benutzen würde: Bekanntheitsgrad, Herstellerzuordnung und Herstelleridentifikation. Es sollte stets klar erkennbar sein, auf welche Ebene man sich bei der Wertung von Prozentsätzen bezieht.

Die Bewertung eines Umfragebefunds im Gesamtbild der Umstände und sonstigen Beweismittel eines Falls ist zweifelsohne die Kompetenz und Aufgabe des Richters oder Prüfers. Werden allerdings von der Rechtsprechung Ansprüche formuliert, ohne wichtige inhärente Effekte der Befragungsmethode zu beachten, dann muss sich der Demoskop äußern, ob eine solche Beurteilung methodisch vertretbar ist. Dies erscheint dringend notwendig im Hinblick auf weit verbreitete Vorstellungen der Rechtsprechung davon, welche Höchstwerte für Kennzeichnungskraft erreicht werden könnten.

2) BGH, GRUR 2004, 514 (516) – Telekom.

3) BGH, GRUR 2003, 1040 (1043) – Kinder.

4) BGH, GRUR 1991, 863 – Avon; für die Forderung, die Durchsetzung müsse „deutlich über 80%“ betragen, vgl. z. B. Beschl. der Markenabteilung 3.4 des DPMA v. 20. 5. 1999 – Milchschnitte.

In der herrschenden juristischen Lehre werden für „einhellige Verkehrsdurchsetzung“ immer wieder Prozentwerte von 80 Prozent und darüber gefordert<sup>5</sup>. Teilweise ist eindeutig der Bekanntheitsgrad gemeint<sup>6</sup>, wogegen keine methodischen Bedenken seitens der Umfrageforschung bestehen. Teilweise ist jedoch unklar, ob sich die Forderungen auf den Bekanntheits- oder Zuordnungsgrad beziehen, teilweise wird ausdrücklich eine Herstellerzuordnung von über 80 Prozent gefordert<sup>7</sup>. Letzteres ist aus umfragemethodischer Sicht ein überzogener Anspruch, den bereits 1966 Noelle-Neumann und Schramm zurückgewiesen haben: „Was die Rechtsprechung derzeit bezüglich der Höhe der Verkehrsgeltung verlangt, ist demoskopisch nicht erfüllbar“<sup>8</sup>.

Vielfach wird von juristischer Seite darauf hingewiesen, inzwischen würden viel mehr Marken als noch in den Siebziger oder in den Achtziger Jahren Bekanntheitsgrade von über 90 Prozent erreichen. Dies könne doch höhere Anforderungen rechtfertigen. Dabei wird impliziert, parallel zum Bekanntheitsgrad hätten sich gleichfalls Herstellerzuordnungen und Identifikationsgrade erhöht. Teilweise sind die Bekanntheitsgrade von Marken tatsächlich gestiegen. Umfrageforscher beobachten, dass Verbrauchern heute ein breiteres Repertoire an Markennamen geläufig ist als noch vor einigen Jahrzehnten. Das Markenrepertoire ist allerdings oberflächlicher und seltener mit substanziellem weiteren Wissen unterfüttert. Dies wird unter anderem auf die Vermehrung der erhältlichen Marken zurückzuführen sein. Die Sicherheit bei der Herstellerzuordnung und Herstelleridentifikation hat sich keineswegs gesteigert, ist nicht gleichermaßen mitgewachsen. Während dem Erreichen von Bekanntheitsgraden von deutlich über 80 Prozent kein Methodeneffekt entgegensteht (und solche Ergebnisse heutzutage tatsächlich auch von nicht wenigen Marken erreicht werden), werden in der Praxis nur extrem selten Zuordnungsgrade (also Herstellerhinweiswerte, nicht Bekanntheitsgrade) in der Größenordnung von 80 Prozent oder darüber festgestellt<sup>9</sup>. In allen 46 Gutachten des *Instituts für Demoskopie Allensbach* nur zu diesem Thema seit 1985 wurden Werte über 75 Prozent Herstellerzuordnung nur zweimal gefunden<sup>10</sup>. Sämtliche anderen Befunde lagen unter 75 Prozent, und zwar größtenteils sehr erheblich darunter.

Das liegt nicht an den Marken oder den Vermarktungsstrategien der jeweiligen Hersteller, sondern vorrangig an der Struktur des Messinstruments, nämlich an der Tatsache, dass beim Dreischrittmodell hintereinander geschaltet abgefragt wird. Je nach der Vorantwort verengen sich die befragten Kreise von Schritt zu Schritt stetig, bis am Ende nur noch ein kleinerer Teil der Ausgangsmenge überhaupt mit der dritten Frage zur Herstelleridentifikation konfrontiert wird. Bei demoskopischen Testfragen bleibt grundsätzlich ein gewisser Anteil der Befragten unentschieden oder gibt gar keine Antwort, und zwar bei jeder Frage etwa 8 bis 10 Prozent, bei schwierigen Fragestellungen auch darüber, um 15 Prozent. So gehen bei jeder Phase im Dreischritt durchschnittlich 10 Prozent als Unentschiedene ohne konkrete Antwort verloren. Selbst bei Bekanntheitswerten von 98 oder 99 Prozent kann die nachfolgende Herstellerzuordnung daher maximal ca. 80 Prozent erreichen. Das ist als Extremwert zu begreifen, der so gut wie nicht erreichbar ist. Es handelt sich sozialwissenschaftlich ausgedrückt um einen so genannten „Ceiling-Effekt“ (von engl. „ceiling“ = Decke). Die Messwerte werden gleichsam „unter der Zimmerdecke“ zusammengedrückt, können nicht weiter steigen. Auf Grund der Gestaltung des Messinstruments sind keine höheren Messwerte möglich<sup>11</sup>.

70 Prozent Herstellerzuordnung wären als außerordentlich guter Wert aufzufassen. Wenn zum Beispiel in der BGH-Entscheidung „Stonsdorfer“ von 1974 ein Zuordnungsgrad von 74 Prozent bei noch viel höherer Bekanntheit festgestellt wurde<sup>12</sup>, spricht dies entgegen der damaligen Auffassung des BGH klar für das Vorliegen „einhelliger Verkehrsdurchsetzung“, weil höhere Werte kaum erreichbar sind. Bei der einst von Noelle-Neumann/Schramm vorgeschlagenen und viel zitierten Skala der Kennzeichnungskraft mit vier Stufen war die höchste Stufe (bezeichnet als „notorische bzw. berühmte Marke“, „überragende Verkehrsgeltung“) sachgerecht ab 63 bis 65 Prozent Herstellerzuordnung angesetzt (wohlgermerkt Herstellerzuordnung, nicht Bekanntheit), also wenn etwa zwei Drittel der Verkehrsbeteiligten in dem jeweiligen Zeichen eine Marke erblicken<sup>13</sup>. Das ist vom demoskopischen Standpunkt aus nach wie vor ein brauchbarer Leitwert für die Bejahung von Durchsetzung, wenn erhöhte Anforderungen gestellt werden.

Insbesondere bei den neueren, nicht-traditionellen Markenformen müssen aus demoskopischer Sicht die Ansprüche an die höchstens erreichbaren Prozentsätze korrigiert, sogar deutlich heruntergeschraubt werden<sup>14</sup>. Diese Marken treten in den meisten Fällen neben oder unter einer Wortmarke auf. Auf den Waren oder Packungen erscheinen sie selten völlig eigenständig und ohne Zusammenhang zum Markennamen. Submarken können nicht höhere Anforderungen erfüllen als die jeweils zugeordnete Wortmarke. Aus demoskopischer Perspektive wäre beispielsweise eine Farbmarke, die einen Bekanntheitsgrad von 80 Prozent im Verkehrskreis erreicht, eine Zuordnung zu nur einem einzigen Hersteller in Höhe von 65 Prozent und sogar noch eine namentliche Identifikation des Herstellers von 60 Prozent, als eine recht starke Marke einzustufen. Höchstanforderungen an neue Markenformen können realistisch nicht viel höher geschraubt werden.

### 3. Gewicht der Herstelleridentifikation bei der Wertung

Es besteht Anlass, sich näher mit Funktion und Gewicht der namentlichen Identifikation des Herstellers zu befassen, des dritten Beweiselements im Abfragemodell für Verkehrsdurchsetzung. Die Urteile „Kinder“ und „Telekom“ beziehen sich auf von Hause aus nicht unterscheidungskräftige Zeichen, die eine hohe bzw. „gesteigerte Kennzeichnungskraft“ in Form „einhelliger Verkehrsbekanntheit“<sup>15</sup>, nachzuweisen hatten. Die Urteile ver-

5) Z.B. Ströbele/Hacker, MarkenG, 7. Aufl. (2003); § 8 Rdnrn. 137 ff.; Fezer, MarkenR, 3. Aufl. (2001); § 8 Rdnrn. 432 f.; Kahler, GRUR 2003, 10 (12); Ingerl/Rohnke, MarkenG, 2. Aufl. (2003), § 8 Rdnr. 339; Sack, WRP 2000, 1032.

6) Ströbele/Hacker (o. Fußn. 5), § 8 Rdnr. 496.

7) Über 80%: Kahler, GRUR 2003, 12; Ingerl/Rohnke (o. Fußn. 5), § 8 Rdnr. 339; bei Farben über 70%: Dierksen, Abstrakte Farbmarke, 2001, S. 170; Knaak, GRUR 1985, 109. Vgl. auch BPatG, GRUR 1999, 64 – Aral-Blau I: 69% abstrakte Herstellerzuordnung genügten nicht bei einer Farbkombination.

8) Noelle-Neumann/Schramm, GRUR 1966, 78.

9) Zutreffende Hinweise auf die methodisch bedingten Grenzen für Anforderungen bei Ernst-Moll, GRUR 1993, 10; Schulz/Süchting, in: Gloy, Hdb. d. WettbR, 1. Aufl. (1986), § 62 Rdnr. 15; Hodik, ÖBl 1983, 3; Sauberschwartz, WRP 1970, 47.

10) Ergebnis einer Recherche im Archiv des Instituts für Demoskopie Allensbach. Zwischen 1985 und Februar 2004 wurden 46 Gutachten zum Beweisthema Verkehrsgeltung bzw. -durchsetzung erstellt.

11) Zum „Ceiling-Effekt“ in der Skalentheorie vgl. z. B. Bortz, Lehrbuch der empirischen Forschung für Sozialwissenschaftler, 1984, S. 126.

12) BGH, GRUR 1974, 339 – Stonsdorfer.

13) Noelle-Neumann/Schramm, GRUR 1966, 81 f.

14) Vgl. Ingerl/Rohnke (o. Fußn. 5), § 4 Rdnr. 21; v. Schultz, MarkenR, 2003, § 4 Rdnr. 12.

15) BGH, GRUR 2003, 1044 – Kinder; GRUR 2004, 514 (516) – Telekom.

schärfen das bisher geltende Kriterium für einen Umfragenachweis bei diesem Markentyp ganz erheblich, allerdings in einer methodisch nicht tragfähigen Weise. Statt wie bisher auf die abstrakte Herstellerzuordnung, also auf Ebene 2 im Abfragedreischritt abzuheben, wird neu als entscheidendes Kriterium für die Beurteilung bei diesem Markentyp allein auf den Grad korrekter namentlicher Identifikation abgehoben. Die dazu vorgelegten Prozentsätze wurden in beiden Fällen für ungenügend erklärt. Befragte, die fälschlich ein anderes Unternehmen als den tatsächlichen Inhaber nennen, dürften nach – offensichtlich in dieser Frage demoskopisch nicht beratener – Ansicht des BGH zur Ermittlung des entscheidenden Prozentsatzes nicht einbezogen werden. Im Fall „Kinder“ hat ein Privatgutachten eine Herstellerzuordnung von 64 Prozent in der gesamten Bevölkerung und eine zutreffende namentliche Identifikation von 49 Prozent ergeben. Der mit Sicherheit im Rahmen der Befragung erhobene Bekanntheitsgrad wird im Urteil gar nicht mitgeteilt<sup>16</sup>. Im Fall „Telekom“ wird eine Herstellerzuordnung von 75 Prozent in der Gesamtbevölkerung zitiert, eine zutreffende namentliche Identifikation von 60 Prozent. Auch hier wird der Bekanntheitsgrad nicht mitgeteilt<sup>17</sup>. Die Urteile befassen sich nicht eigens mit der Frage, wie die Personen zu werten sind, die im dritten Abfrageschritt unentschieden, ohne konkrete Angabe geblieben sind, die also weder einen richtigen noch einen falschen Herstellernamen genannt hatten. Da aber nur die positiven, richtigen Angaben gewertet wurden, ist davon auszugehen, dass der BGH meint, auch fehlende Angaben zu Ungunsten des Markeninhabers anrechnen zu können.

Diese Neudefinition und Verschärfung des entscheidenden Kriteriums im Dreischritt fußt offenbar auf der Vorstellung, im Allgemeinen könne der Verbraucher bei Marken den Hersteller namentlich zutreffend angeben, und bei starken Marken könnte man doch dieses Wissen erwarten. Es ist allerdings seit langem bekannt und vielfach empirisch nachgewiesen, dass der Name des Herstellers für die Verbraucher bei den meisten Waren für die Orientierung im Markt von stark untergeordneter Bedeutung ist. Zentral für die Orientierung der Verbraucher ist der Markenname. Die Fähigkeit der Verbraucher, Herstelleramen selbst sehr bekannter und markenrechtlich problemloser Marken richtig zu benennen, ist nachweislich generell gering. An dieser Stelle sei die von Anette Kur und Frauke Henning-Bodewig 1984 veranlasste Grundlagenforschung in Erinnerung gerufen: Sie ließen das Institut *Infratest* 15 sehr bekannte Marken in Westdeutschland testen. Die Bekanntheit dieser Marken lag durchschnittlich bei 96 Prozent, die richtige namentliche Identifikation demgegenüber bei durchschnittlich nur 22 Prozent. Die Autorinnen kamen zu dem Schluss: „Bei der Überprüfung der Kenntnisse der Verbraucher über die Hersteller von bekannten Markenwaren zeigte sich, dass nur ein sehr geringer Teil der Verbraucher den Namen des Herstellers angeben kann. (...) Es kann somit davon ausgegangen werden, dass die Verbraucher in aller Regel bei fehlender Identität von Marke und Firmendominanz des Herstellers nicht in der Lage sind, ein Produkt dem richtigen Hersteller zuzuordnen, und zwar auch dann nicht, wenn sich der Name des Herstellers zusätzlich zur Marke auf dem Produkt befindet“<sup>18</sup>.

Im Sommer 2004 hat das *Institut für Demoskopie Allensbach* im Rahmen der Grundlagenforschung aktuelle Werte eingeholt, um zu prüfen, ob das, was für 1984 festgestellt worden ist, für die aktuelle Situation noch zutrifft<sup>19</sup>. Die Ergebnisse decken sich im Wesentlichen mit denen von 1984, bei einer Tendenz zu einem insgesamt

noch etwas *niedrigerem* Wissensniveau. Die Vielzahl der getesteten Marken erreichte im Durchschnitt 94 Prozent Bekanntheit, der Anteil zutreffender Herstelleridentifikationen betrug hingegen nur durchschnittlich 18 Prozent.

Man darf folglich keineswegs davon ausgehen, dass die Herstelleridentifikation umso höher ausfällt, je bekannter eine Marke ist. Die Fähigkeit der Bevölkerung, zu bekannten Markennamen den Hersteller zutreffend zu benennen, korreliert wenig mit der Bekanntheit der Marke. Auch sehr bekannte Markennamen können mit einer niedrigen namentlichen Identifikation einhergehen, zum Beispiel die Wortmarke „Milka“. Sie ist zwar 99 Prozent der Bevölkerung bekannt, nur 1 Prozent kennt sie nicht. Aber nur 4 Prozent der Verbraucher können den Hersteller *Kraft Foods* richtig nennen. Immerhin 16 Prozent nennen eine der früheren Bezeichnungen des Markeninhabers, *Jacobs und/oder Suchard*. Rechnet man letzteres hinzu, kommt man auf höchstens 20 Prozent richtige Identifikation. Die Marke „Knorr“ ist 97 Prozent der Bevölkerung bekannt, nur 3 Prozent beherrschen den Hersteller *Unilever*. 34 Prozent vereinfachen hingegen: „*Knorr kommt von Knorr*“. „Whiskas“ ist 96 Prozent der Bevölkerung bekannt, nur 1 Prozent kennt den aktuellen Inhaber *Masterfoods* und weitere 1 Prozent den früheren Inhaber *Effem*. „Odol“ kennen 94 Prozent der Gesamtbevölkerung, lediglich 1 Prozent kann den aktuellen Markeninhaber *Glaxo Smith Kline* richtig angeben. Den früheren Inhaber *Lingner & Fischer* kennen weitere 1 Prozent, zusammen 2 Prozent.

Viel stärker als die Bekanntheit der jeweiligen Marke erklärt die Bildung der Befragten diese Ergebnisse. Die Bildung ist eine bedeutende intervenierende Variable. Zum Beispiel kennen 49 Prozent der Personen mit Abitur und/oder Studium den Hersteller von „Nivea“, aber nur 27 Prozent der Personen mit höchstens Volks- oder Hauptschulabschluss. Diesen Faktor, die Bildung der Befragten, kann der Markeninhaber nicht durch Werbemaßnahmen beeinflussen. Würde die Namensnennung als entscheidendes Kriterium für Durchsetzung bei diesem Markentyp eingeführt, wären bestimmte Arten von Marken systematisch im Vorteil, andere benachteiligt:

Bei einem Wechsel des Kriteriums wären Wortmarken bevorteilt, bei denen Markenname und Hersteller identisch sind wie z. B. „Shell“, „Avon“. Ebenso im Vorteil wären Marktführer, weil deren Angabe als Rate-Antwort nahe liegt, ohne dass wirkliches Wissen vorhanden sein muss (z. B. „Persil“ bei Waschmitteln, „Nokia“ bei Mobiltelefonen, „Milka“ bei Schokolade). Solche so genannten „Top of Mind Responses“ sind ein bekanntes Phänomen in der Marktforschung. Bevorteilt wären Bildzeichen, 3D-Gestaltungen, Formen und Farben, weil bei diesen außer dem Hersteller auch die Markennamen zur eindeutigen Identifikation herangezogen werden können und Verbrauchern die Angabe eines zu einem konkreten Produkt zugehörigen Markennamens wesentlich leichter fällt als die Angabe des Herstellers. Bevorteilt wären Marken, die eher gebildete Kreise ansprechen wie Software, Titel von Büchern, Zeitungen, Zeitschriften sowie alle Marken, die ein enges und besonders interessiertes Spezial- bzw. Fachpublikum ansprechen.

Benachteiligt wären bei einem Wechsel des Kriteriums alle Marken, die breite Bevölkerungskreise jeder Bildungsebene gleichermaßen ansprechen, wie Süßwaren, Getränk-

16) BGH, GRUR 2003, 1043 – Kinder.

17) BGH, GRUR 2004, 514 (516) – Telekom.

18) *Kur/Henning-Bodewig*, Marke und Verbraucher, Bd. 1, 1988, S. 101 f.

19) Allensbacher Archiv, IfD-Umfrage 7060, (August/September) 2004; Basis: Bundesrepublik Deutschland, Bevölkerung ab 16 Jahre.

ke, Telekommunikation, außerdem alle weniger dominanten Marken, die sich Fehlangaben eines klaren Marktführers, der als Rateantwort nahe liegt, anrechnen lassen müssten, also die Nummern 2, 3, 4 usw. in einem Markt. Marken, die einmal oder mehrfach den Besitzer gewechselt haben, wären benachteiligt, beispielsweise Unternehmen, die vor kurzem eine Traditionsmarke aufgekauft oder sich bei Fusionen oder Übernahmen umbenannt haben. Verbraucher brauchen jahrelang, bis sie einen Herstellernamen verinnerlichen und korrekt beherrschen. Wer weiß denn beispielsweise, wer aktuell die Schokolade „Sarotti“ herstellt? Der Name „Sarotti“ wurde 1852 erstmals genutzt, 1929 ging er an *Nestlé*, 1998 an *Stollwerck* und ist heute bei *Barry Callebaut* mit Sitz in der Schweiz. Benachteiligt sind schließlich Hersteller mit einem komplexen und/oder fremdsprachigen Namen wie *Procter & Gamble*, *Glaxo Smith Kline*, *Masterfoods*, *Bestfoods*, *Ferrero* gegenüber Marken, deren Hersteller einen einfachen Namen tragen.

Ein geeignetes, neutrales Maß zur Durchsetzung sollte die Wahrnehmung der Verbraucher unabhängig von ihrer formalen Bildung abbilden. Ein geeignetes Maß sollte auch bestimmte Markenformen nicht systematisch bevorzugen. Die Namensnennung, der dritte Schritt im Dreischritt, ist als Kriterium für gesteigerte Kennzeichnungskraft einer originär unterscheidungsschwachen Marke daher keinesfalls geeignet, methodisch unhaltbar. Wenn gesteigerte Anforderungen gestellt werden, dann sollte man dies über die Intensität der Zuordnung zu einem einzigen, nicht unbedingt namentlich bekannten Hersteller regeln. Dieser bewährte Maßstab ist gerecht und auch methodisch einwandfrei.

Wie sind nun Herstelleridentifikationen, die man im Dreischritt regelmäßig ermittelt, methodisch einwandfrei zu bewerten? Die Ermittlung ist ja nicht ohne Grund als Element im Dreischritt enthalten. Man darf diese Angaben nicht zum Kern des Modells erklären, den man prozentmäßig in den Mittelpunkt der Wertung stellen könnte. Ebenso wenig darf man aus Fehlzuordnungen in normalem Maß (und auch aus ganz fehlenden Zuordnungen) irgendwelche Prozentabzüge bei der Herstellerzuordnung begründen. Aber die Information, wie viele Befragte in der Lage sind, das Testobjekt bzw. den Hersteller spontan richtig namentlich zu benennen, kann zusätzlich beitragen im Sinne von *Pluspunkten* für den Markeninhaber bei der Kennzeichnungskraft, wenn in überdurchschnittlichem Maße richtige Identifikationen nachzuweisen sind und die Werte sich auffallend nah am Grad der abstrakten Herstellerzuordnung bewegen.

Zu relativieren wäre eine abstrakte Herstellerzuordnung nur dann, wenn Fehlangaben in *massivem* Ausmaß gefunden werden, etwa wenn ausgesprochen viele Namensnennungen erfolgen, von denen jedoch mehr als ein Drittel falsch sind. Es kommt dann weiter darauf an, ob die Fehlangaben zersplittert sind und sich auf viele verschiedene Unternehmen verteilen oder – was ernst zu nehmen wäre – sich auf einen bestimmten anderen Wettbewerber vereinigen. Nur in diesem eher seltenen Fall besteht genug Anlass, an der Alleinstellung des Inhabers zu zweifeln und daran, dass sich der erzielte Wert zur abstrakten Herstellerzuordnung tatsächlich auf den tatsächlichen Inhaber bezieht<sup>20</sup>.

Die zutreffende Namensidentifikation des Herstellers sollte als zentrales Kriterium für gesteigerte Kennzeichnungskraft oder Verkehrsdurchsetzung verworfen werden. Sie ist nicht nur umfragemethodisch äußerst problematisch und führt eine ungerechte Bewertung ein<sup>21</sup>, sondern erscheint auch als ein Rückfall in die Zeit, als die Marke

noch vom Geschäftsbetrieb abhängig war. Das Kriterium Intensität der abstrakten Herstellerzuordnung (Schritt 2 im Abfragedreischritt) ist nach wie vor das geeignete Maß, gerade bei gesteigerten Anforderungen – sofern die im Messinstrument inhärenten Beschränkungen anerkannt und realistische Erwartungen formuliert werden.

#### 4. Weiterentwicklung des Fragemodells

Auf die neuen Markenformen passt das bisherige Dreischritt-Abfragemodell nicht optimal. Um die neuen, nicht konventionellen Markenarten wie Farben, Formen usw. adäquat zu messen, sollte es weiterentwickelt werden. Bei Farben und Formen ist denkbar, dass das Zeichen von vornherein gar keinen Herstellerhinweis aussendet, weil es als bloße Verzierung, als allgemein übliche Gestaltung, Warnfarbe, Signalhinweis oder ähnlich allgemein verstanden wird. Keine gute Lösung wäre, die bewährte Kernfrage zum zweiten Schritt, zur Herstellerzuordnung, durch entsprechende Kategorienvermehrung auszudehnen. Die Verständlichkeit dieser Testfrage ist schon jetzt mit den Alternativen „von einer oder von mehreren Unternehmen?“ problematisch, ein Grenzfall. Die Verständlichkeit der Testfragen muss im Interesse der Validität der Messung jedoch oberstes Gebot sein. Außerdem sind die Antwortkategorien „von mehreren Herstellern“ und die Vorstellung, es könnte eine allgemein übliche Gestaltung sein, nicht trennscharf, sondern überschneiden sich. Wenn etwas als allgemein üblich empfunden wird, dann wird es wahrscheinlich auch von mehreren oder sogar vielen Herstellern benutzt. Demoskopische Testfragen müssen trennscharfe Kategorien aufweisen, sonst ist die Messung unbrauchbar.

Die methodisch beste Lösung ist, in solchen Fällen auf Ebene 2 (Herstellerzuordnung) einen weiteren, zusätzlichen Schritt einzuführen. Als ein Beispiel für einen Fragewortlaut bei nicht-konventionellen Markenformen sei ein Fragewortlaut zur Gestaltung von Sportsocken vorgestellt. Die erste Testfrage zur Ermittlung der Bekanntheit wäre (unter Vorlage eines neutralisierten Fotos der Socken): „*Haben Sie Socken in dieser Gestaltung schon einmal irgendwo gesehen oder noch nicht?*“ Dann als neues Element die grundsätzliche Klärung, ob überhaupt ein Herstellerhinweis bei dieser Art von Marke in Frage kommt (2 a): „*Gehört diese Gestaltung für Sie zu irgendwelchen ganz bestimmten Marken oder Herstellerfirmen, oder ist das für Sie eine bei Socken allgemein übliche Gestaltung, die zu keiner bestimmten Marke oder Firma gehört?*“ Im Fall der Antwort „Ja, gehört zu bestimmten Firmen oder Herstellern“ erfolgt die Frage (2 b): „*Kennen Sie Socken mit diesem Aussehen nur von einer einzigen Marke bzw. einem einzigen Hersteller oder von mehreren verschiedenen Marken bzw. Herstellern?*“ Die Nachfragen zur Herstelleridentifikation müssen schließlich auch noch absolviert werden.

Sobald das Abfragemodell zur Durchsetzung um einen vierten Schritt erweitert wird, können sich auf Grund der Struktur des Messinstruments natürlich keine Prozentsätze mehr in der gewohnten Höhe ergeben, sondern niedrigere. Das muss bei der Wertung berücksichtigt werden. Erwartungen an die Prozentwerte dürfen dann nicht in der gleichen Höhe verharren wie beim traditionellen Dreischritt. Man erhält insgesamt einen guten Kompromiss zwischen Verständlichkeit der Abfrage und dem Bedürfnis der Rechtsprechung, genau das geklärt zu sehen, was Beweisaufgabe ist.

20) So bereits *Noelle-Neumann/Schramm*, GRUR 1966, 70 (73).

21) Diese Auffassung teilt *Pflüger*, GRUR 2004, 656.

## 5. Wichtigkeit von Vergleichswerten

Gemäß der Rechtsprechung des *EuGH* darf zur Feststellung von Verkehrsdurchsetzung nicht allein auf einen einzelnen demoskopischen Prozentwert abgestellt werden, sondern es sind die gesamten Umstände des Einzelfalls und sonstige Indizien aus anderen Quellen – Werbeaufwand, Umsatz und so weiter – zu berücksichtigen<sup>22</sup>. Es gibt keine starren Prozentgrenzen für das Vorliegen von Durchsetzung. Kein einzelner demoskopisch ermittelter Prozentsatz kann die rechtliche Beurteilung determinieren, ob Durchsetzung vorliegt. Ob Durchsetzung zu bejahen ist, ergibt sich vielmehr aus dem Gesamtbild aller Beweismittel im Einzelfall. Die Umstände des Einzelfalls sind allerdings bei der Wertung noch weitaus ernster zu nehmen als bisher. Bisher werden einzelne Prozentwerte isoliert betrachtet, und dabei orientiert sich die Rechtsprechung an eingespielten, rein psychologischen Prozentgrenzen, die umfragemethodisch nicht im geringsten abgesichert sind. Eine solche eingespielte Grenze ist zum Beispiel, ob über oder unter 50 Prozent des Verkehrs exklusiv nur einem Hersteller zuzuordnen.

Die Betrachtung einzelner Werte gewissermaßen nach „Bauernregeln“ wird dem Einzelfall gerade nicht gerecht. Eine adäquate Betrachtungsweise, die auch dem demoskopischen Instrument angemessen ist, geht von der Marktlage des jeweiligen Produkts aus und wertet im Lichte der Leistungen der unmittelbaren Konkurrenten der Marke. Dies entspräche gleichzeitig dem wichtigsten Erkenntnisprinzip der empirischen Sozialforschung: Erkenntnis aus dem Vergleich<sup>23</sup>.

Was bedeutet ein Messergebnis, nach dem eine Marke 60 Prozent des Verkehrskreises bekannt ist? Für die Stärke der Marke kann es völlig Unterschiedliches bedeuten – je nachdem, in welchem Konkurrenzumfeld sie sich bewegt. Märkte unterscheiden sich erheblich danach, welche Bekanntheitsgrade durchschnittlich in ihnen überhaupt erreichbar sind, auch danach, wie viele Marken überhaupt in Konkurrenz zueinander stehen, wie intensiv der Konkurrenzdruck ist. Zum Beispiel hat eine Befragung des *Instituts für Demoskopie Allensbach* ergeben, dass die Wortmarke „Citibank“ eine Bekanntheit von 66 Prozent erreicht hat (Ebene 1 im Abfragedreischritt). Was die 66 Prozent wert sind, erkennt man erst, wenn man diesen Befund in andere Ergebnisse, Vergleichswerte aus dem gleichen Markt einordnet: 66 Prozent Bekanntheit werden von einer Vielzahl anderer Bankennamen übertroffen; in diesem Markt sind Bekanntheitswerte ab 80 Prozent und weit höher die Regel. Die 66 Prozent sind folglich keineswegs erstklassig, sondern Mittelmaß. Für eine Online-Bank wiederum wäre ein solcher Bekanntheitsgrad mehr als ordentlich, denn diese sind weit entfernt von den Bekanntheitsgraden klassischer Banken.

Ein anderes Beispiel. Eine andere *Allensbacher* Befragung ergab, dass die Marke „Intel“ einen Bekanntheitswert von 55 Prozent erreicht hat. Wenn man den Befund einordnet in das Abschneiden der wichtigsten Wettbewerber im Computer- und Online-Markt, wird sichtbar: Die 55 Prozent bewegen sich im oberen Drittel. Die Mehrzahl der Wettbewerber erzielt Bekanntheitswerte von deutlich unter 50 Prozent.

Bei gerichtlich veranlassten Befragungen zum Beweisthema Durchsetzung oder auch zur bekannten Marke sollte es selbstverständlich werden, gleichzeitig einige Vergleichswerte aus dem Markt erheben zu lassen und bei der Wertung zu berücksichtigen. Das sollte gegebenenfalls möglichst schon im Beweisbeschluss verankert werden. Sobald ein Privatgutachten Vergleichswerte enthält, ist das nicht als unerheblich zu betrachten, sondern im Ge-

genteil zentral bei der Wertung zu berücksichtigen. Vergleichswerte erleichtern dem Richter, dem Entscheider, die Arbeit wesentlich<sup>24</sup>.

## II. Durchsetzungsnachweis beim HABM

Die Umfrageinstitute verzeichnen steigende Nachfrage nach – vielfach länderübergreifenden – Untersuchungen, die zur Vorlage in Alicante bestimmt sind. Im HABM treffen derzeit Entscheider aus Ländern, die den Umfragebeweis schon jahrzehntelang anwenden, auf Entscheider aus Ländern, die den Umfragebeweis bisher nicht kannten (vorwiegend aus Südeuropa, auch Frankreich). Sie müssen zu einer gemeinsamen Praxis finden. In wie vielen Fällen wurde der Umfragebeweis in den dortigen Entscheidungen überhaupt schon diskutiert, angewandt, problematisiert? Eine systematische, quantitative Inhaltsanalyse des *Instituts für Demoskopie Allensbach* verschafft einen Überblick. Einbezogen wurden sämtliche Entscheidungen, die auf der HABM-Homepage abrufbar sind, also ab Ebene Widerspruch. Die ersten Entscheidungen, in denen auf Umfragen als Beweismittel Bezug genommen wurde, stammen von 1998. Die Fallzahlen sind seitdem gestiegen. In den Jahren 2000/01 finden sich 74 Entscheidungen, 2002/03 schon 87, und allein 2004 92.

Die Intensität der Auseinandersetzung mit dem Beweismittel nimmt im Zeitverlauf deutlich zu. Vorgelegte Umfragen werden von den Entscheidern im HABM kaum mehr ignoriert. Fälle, in denen zwar in der Entscheidung erwähnt wurde, dass eine Umfrage eingereicht worden war, diese aber nicht weiter angesprochen wurde, haben sich von 30 auf 16 Prozent halbiert<sup>25</sup>. Gleichzeitig hat sich eine angemessen ausführliche Diskussion der Ergebnisse von 20 auf 38 Prozent erhöht.

In 38 Prozent der Fälle, in denen dem HABM Umfragen vorlagen, war die Argumentation mit den Befunden bisher erfolgreich, in 15 Prozent der Fälle war der Beweiswert unklar, in den meisten Fällen allerdings (47 Prozent) scheiterte die Argumentation. Das ist eine unbefriedigende Bilanz. Die Gründe zeigt eine Analyse der in den Entscheidungen ausgeführten Beanstandungen. In erster Linie waren dafür vermeidbare *Formfehler* bei der Einreichung verantwortlich, etwa keinerlei Information zum Befragtenkreis oder zur Stichprobenbildung, oder keine Information, wer die Untersuchung erstellt hat, über das ausführende Institut, den Gutachter. Häufig ist der wörtliche Fragenwortlaut nicht beigefügt und/oder die Fragenabfolge nicht offen gelegt. Hinzu kommen fehlende Vorlagen, also zum Beispiel die Bildvorlagen, oder die nur bruchstückhafte Vorlage des Untersuchungsberichts, in Einzelblättern, ohne autorisierte Übersetzung in die Verfahrenssprache. Dies alles ruft zu Recht Skepsis beim Amt hervor und sind leicht vermeidbare Fehler.

In 16 Prozent der Fälle wurden Befunde zum falschen Zeichen – oft zu einer Hauptmarke statt zur eigentlich begehrten Submarke – eingereicht. Vor allem bei neuen Markenformen wurde versucht, mit Umfragen zu argumentieren, die sich auf die Hauptmarke beziehen, zum

22) *EuGH*, GRUR Int 1998, 795 – Gut Springenheide; GRUR Int 1999, 734 – Lloyd; GRUR Int 1999, 727 – Chiemsee; GRUR Int 2000, 73 – Chevy; GRUR Int 2000, 354 – Lifting Creme; GRUR 2002, 804 – Philips; *BGH*, GRUR 2004, 1044 – Kinder, m. w. Nachw.

23) Zum Prinzip der Erkenntnis aus dem Vergleich vgl. *Niedermann/Noelle*, in: Festschr. f. Tilmann, 2004, S. 857.

24) Zur Wichtigkeit von Vergleichswerten bei der Wertung vgl. *Niedermann/Schneider*, sic! 2002, 827; in *BGH*, GRUR 1991, 866 f. – Avon, wird klargestellt, dass relevante Vergleichswerte aus demselben engeren Marktsegment wie die eigentlich zu prüfende Marke stammen müssen.

25) Der Untersuchungszeitraum umfasst 1998 bis Ende 2004.

Beispiel bei einer Farbkombination auf die zugehörige Wortmarke. Eine Umfrage muss sich jedoch selbstverständlich genau auf das angemeldete Zeichen beziehen. Dem HABM ist in der konsequenten Ablehnung von Befragungen, die das Beweisthema verfehlen, zuzustimmen.

### 1. Zeitpunkt der Umfrage

Die Erstellung der Befragung *nach* dem relevanten rechtlichen Zeitpunkt war ein häufiger Ablehnungsgrad. Dies wurde vom HABM bislang zu formell gehandhabt. Das Amt ist inzwischen Umfragenachweisen, die nicht mehr vor dem rechtlich erheblichen Kollisions- oder den Anmeldezeitpunkt erhoben wurden, sondern kurz danach – bei Vorliegen ganz bestimmter Voraussetzungen – etwas zugänglicher<sup>26</sup>.

### 2. Repräsentativität

Weiterhin sind Fehlaufassungen zum angemessenen Stichprobenumfang beim HABM nicht selten. Manche Beschlüsse offenbaren eklatante Missverständnisse: So musste man beispielsweise in einer Entscheidung einer Widerspruchskammer über Befunde aus einer bevölkerungsrepräsentativen Mehr-Themen-Umfrage zur Bekanntheit eines Markennamens lesen: „*Another problem in relation to the survey is that ... 2017 respondents is a statistically small sample considering the German population of around 80 million inhabitants ...*“<sup>27</sup>. Das Prinzip der Repräsentativität wurde von diesem juristischen Entscheider nicht ansatzweise verstanden. Er beurteilt die Güte der Befragung nach dem Motto: je mehr Befragte, desto besser. Bei einer Stichprobe, die nach anerkannten wissenschaftlichen Grundsätzen gezogen worden ist und idealerweise eine Zufallsauswahl darstellt, genügen selbstverständlich bei der breiten Bevölkerung 1000 Fälle für den Schluss vom Befragungsergebnis auf die Verhältnisse in der Grundgesamtheit, dem interessierenden Verkehrskreis, und zwar mit einem für rechtliche Entscheidungen voll ausreichenden Sicherheitsgrad.

### 3. Offene versus geschlossene Fragen

In der guten Absicht, eine neutrale, nicht suggestive Frageform zu sichern, bevorzugen die Entscheider beim HABM laienhaft so genannte „offene Fragen“, die freie Antworten abrufen ohne jedes Angebot an Antwortkategorien, und diskriminieren die so genannten „geschlossenen Frage“, die mit Antwortkategorien zur Auswahl versehen sind. Folge: In 7 Prozent der Fälle scheiterte die Argumentation mit Umfragebefunden u. a. daran, dass das HABM das angewandte Fragemodell, insbesondere den Fragetyp „geschlossene Frage“ verwarf. Das betrifft natürlich den bewährten Dreischritt zum Nachweis von Verkehrsdurchsetzung, der im Kern aus geschlossenen Fragen besteht. Ein Beispiel für eine offene Abfrage wäre: „*Nennen Sie mir alle Jeansmarken, die Ihnen einfallen!*“ Der Frage hingegen „*Kennen Sie diese Bezeichnung in Zusammenhang mit Jeans, oder ist sie Ihnen in diesem Zusammenhang ganz unbekannt?*“ – Antwortmöglichkeiten: JA/NEIN/UNENTSCHEIDEN/KEINE ANGABE würde geringer Beweiswert zugemessen. Das HABM setzt bisher offene Fragen mit neutralen Fragen und geschlossene Fragen mit Suggestivfragen gleich – eine umfragemethodisch unhaltbare Auffassung, die zu revidieren sein wird.

In mehreren Entscheidungen wurde sogar grundsatzartig formuliert, allein offene Abfragen seien beweiskräftig<sup>28</sup>. Glücklicherweise ist das bisher nur in abgeschwächter Form in den Entwurf für die „Opposition Guidelines“ eingegangen, die sonst alles in allem vernünftige Vorschriften dazu aufstellen, wie beweiskräftige Umfragen

auszusehen haben, wie sie dokumentiert sein müssen<sup>29</sup>. Die ideale Abfrage für das HABM zum Beweisthema Durchsetzung einer Wortmarke lautet tatsächlich in etwa: „*Wenn Sie einmal an Jeans denken, welche Marken fallen Ihnen da ein?*“ Dies ist zwar ein in der Marktforschung verbreitetes Abfragemodell für die so genannte „Markenbekanntheit“ und hat dort auch seinen Nutzen und seine Berechtigung. Aber zur Klärung von markenrechtlicher Verkehrsdurchsetzung ist das Modell ungeeignet, geradezu ein Rückschritt.

Warum ist das in Deutschland und anderen Ländern etablierte Dreischritt-Abfragemodell der marktforscherischen Frageart überlegen?

Der Dreischritt aus geschlossenen Testfragen ist fair, weil diese das bei den Befragten vorhandene Wissen vollständig abbilden. Offene Fragen wie in der marktforscherischen Bekanntheitsabfrage können das vorhandene Wissen über Marken nicht vollständig abbilden. In den Antworten sind die Auffassungen von weniger gebildeten, weniger formulierungsgewandten Befragten unterrepräsentiert. Der BGH hat dies in seinen Urteilen, in denen er sich mit richtigem Fragen bei Irreführung beschäftigt hat, längst erkannt und festgestellt, dass offene Fragen nur einer Seite gerecht werden, nämlich der, die bestimmte Antworten *nicht* wünscht<sup>30</sup>. Das bedeutet: Für die meisten Markeninhaber fiele die Möglichkeit weg, das im Publikum vorhandene Wissen auch vollständig abzubilden – es sei denn, sie sind der jeweilige Marktführer. Für den, der die Liste anführt – beispielsweise mit 98 Prozent Bekanntheit bei Banken oder im Computerbereich – sind offene Abfragen kein Nachteil, wohl aber für Marken, die nicht dominieren. Im Endeffekt wirken sich offene Abfragen zum Nachteil der meisten Markeninhaber aus.

Der Dreischritt mit dem Ausgangspunkt, dass die konkrete Marke, um die es geht, zu Beginn der Befragung vorgestellt wird, ist fair, wenn es sich um weniger bekannte Marken oder andere Marken als den Marktführer handelt. Die marktforscherische Art, Bekanntheit anhand einer offenen Frage abzurufen, bevorteilt starke, große Marken, die im Vordergrund des Bewusstseins leicht abrufbar sind.

Das Dreischritt-Modell nimmt nicht den Markencharakter vorweg, den es erst erforschen soll. Der Ansatz „*Nennen Sie mir alle Marken, die Ihnen in den Sinn*“

26) HABM, Guidelines Concerning Proceedings before the Office for Harmonization in the Internal Market (Trade Marks and Designs), Part C: Opposition Guidelines, Draft (Status May 2003), Part 5: Trade marks with reputation (Art. 8 [5] CTMR), Chapter III, 1.2.5 Relevant point in time; vgl. HABM, Opposition Divisions, 1272/2002 (30. 4. 2002) – Tod's; 2101/2001 (21. 9. 2000) – Pyloritek; 531/2001 (27. 2. 2001) – Teletoon.

27) HABM, Trade Mark Department, 2149/2003 (15. 10. 2003) – Digital Focus, ebenso in 2625/2003 (26. 11. 2003), 2626/2003 (26. 11. 2003) und 2627/2003 (26. 11. 2003) – T Tosinvest, korrigiert in: R 50/2004-4 – Tosinvest, außerdem in Opposition Divisions, 445/2002 (27. 2. 2002) – Monbébé, korrigiert in R 220/2002-1 (4. 3. 2003) – Monbébé.

28) Z. B. HABM, Trade Mark Department, 2098/2004, 8 (29. 6. 2004) – Hitmaker; R 560/2003-1 (27. 4. 2004) – Vitacoat; R 493/2003-1 (11. 5. 2004) – Georgia Pacific (fig.); Opposition Divisions, 2145/2002 (16. 7. 2002) – Premo Intel Design; 1375/2002 (16. 5. 2002) – ICE 24; 1272/2002 (30. 4. 2002) – Tod's; R 730/2002-2 (11. 12. 2002) – Bio; 974/2002 (10. 4. 2002) – Mediasphere; 445/2002 (27. 2. 2002) – Monbébé; 335/2002 (21. 2. 2002) – Krea.

29) HABM, Opposition Guidelines (o. Fußn. 26), Chapter III, 1.4.4 (iv) Opinion polls and market surveys.

30) Vgl. BGH, GRUR 1991, 680 f. – Porzellanmanufaktur; GRUR 1992, 68 – Königlich Bayerische Weisse; GRUR 1993, 922 – Emilio Adani II; GRUR 1996, 913 – Der meistverkaufte Europas; OLG München, GRUR 1989, 124 – Medaillen-Werbung; vgl. auch Noelle-Neumann, GRUR 1968, 133; Ohde, in: Gloy (o. Fußn. 9), 2. Aufl. (1997), § 18 Rdnr. 22.



*kommen!*“ hingegen unterstellt suggestiv bereits einen Markencharakter. Dies hat z. B. klarsichtig das *Schweizer Bundesgericht* im Entscheid „Levi’s“ (1984) bemängelt<sup>31</sup>.

Die marktforscherische Bekanntheitsabfrage gibt im Fall der Durchsetzung nichts her zur eigentlichen Beweisfrage, ob es sich um eine Marke im markenrechtlichen Sinn handelt. Der rechtlich wichtigste Punkt, ob ein Herstellerhinweis vorliegt, bleibt ungeklärt. Dies wird vom Dreischritt-Modell jedoch gerade geleistet.

Während sich die marktforscherische Bekanntheitsabfrage nicht für andere Markenformen als Wortmarken eignet, passt der Dreischritt noch am ehesten auf sämtliche Markenformen, gegebenenfalls erweitert auch auf die nicht-konventionellen, neueren Markenformen.

Aus umfragemethodischer Perspektive sind einige bisherige Vorstellungen des HABM vom „richtigen“ Fragen ein Rückschritt. Der seit Jahrzehnten in Deutschland bewährte Dreischritt zur Messung von Durchsetzung, der längst auch in Österreich und der Schweiz und auch in Schweden der Standard ist, darf auf EU-Ebene nicht diskriminiert werden oder gar verloren gehen. Offenbar müssen fast alle bisher zwischen Umfrageforschern und Juristen erarbeiteten und bewährten Grundsätze über methodisch richtiges und neutrales Vorgehen bei Umfragen zu rechtlichen Beweis Zwecken auf EU-Ebene in den kommenden Jahren noch einmal völlig neu vermittelt und durchgefochten werden.

#### 4. Geografische Ausdehnung

Die jüngste EU-Erweiterung stellt Markeninhaber vor eine neue Situation. Dass man Verkehrsdurchsetzung in einen „wesentlichen Teil“ der EU nachweisen muss<sup>32</sup>, ist nun eine größere Hürde. Für EU-weite Stichproben mit ca. 1000 bis 2000 Fällen, die sämtliche Mitgliedsländer gleichmäßig abdecken, besteht derzeit noch keine regelmäßige Marktnachfrage. Sie gibt es nur in Einzelfällen, etwa das *Eurobarometer*<sup>33</sup> der EU-Kommission (eine

aufwändige Studie mit hohem Etat, die regelmäßig vergleichend politische Einstellungen der EU-Bürger erhebt). Entsprechend sind EU-weite Angebote allenfalls vereinzelt im Repertoire der leistungsfähigen Umfrageinstitute zu finden, schon gar nicht als Beteiligungsuntersuchung (so genannte „Omnibus- oder „Mehr-Themen-Befragung“), wie sie auf nationaler Ebene zur Verfügung steht. Nachfrage für grenzüberschreitende Forschung in Form von Mehr-Länder-Studien besteht zwar in der Marktforschung oder Sozialforschung zunehmend, aber in der Regel beschränkt auf und repräsentativ für wenige, ausgewählte Länder, höchstens drei bis vier Länder pro Studie. Dies entspricht in etwa dem, was ein einzelnes Unternehmen realistischweise an Etat aufbringen kann.

Die vernünftigste Lösung wäre, wenn der empirische Nachweis als ordentliche Repräsentativbefragung beispielsweise in drei gemeinsam mit dem Amt ausgewählten Ländern geführt würde. Dann müsste sich das HABM allerdings im Vorfeld grundsätzlich zu dem Umfragevorhaben äußern, was bisher leider – entgegen den guten Erfahrungen in vielen Ländern – prinzipiell verweigert wird.

Die Anforderungen an den Umfragenachweis auf EU-Ebene müssen auf der einen Seite realistisch bleiben, nämlich sich an der Leistungsfähigkeit von Markeninhabern wie Umfrageinstituten orientieren, auf der anderen Seite aber dürfen sie nicht erheblich zurückbleiben hinter dem in Deutschland und vielen anderen Ländern längst erreichten methodischen Standard.

---

31) *Schweiz. BG*, SMI 1984, 120 – Levi’s Jeans; vgl. auch *Niedermann/Schneider*, sic! 2002, 822 f.

32) *EuGH*, GRUR 2000, 73 – Chery.

33) Eurobarometer, Die öffentliche Meinung in der Europäischen Union, durchgeführt im Auftrag der Generaldirektion Presse und Kommunikation der EU-Kommission ([www.europa.eu.int/comm/public\\_opinion/index.en.htm](http://www.europa.eu.int/comm/public_opinion/index.en.htm)).