



Dr. Ralf Mayer de Groot im Handbuch „Controlling der Kommunikation“, hrsg. von den **Professoren Bruhn, Esch und Langner** (aktuell in Druckvorbereitung – *erscheint 2015*)

Ansätze zur Quantifizierung des Kommunikationserfolges und des Return of Investment

Management Zusammenfassung und Implikationen für Werbetreibende und Agenturen

Werbung kann außerordentlich effektiv die Nachfrage steigern. Das ist das Kernziel der Werbung. Aber 67 Prozent der Werbung für etablierte Produkte und 45 Prozent der Werbung für neue Marken erzielen keine Absatzwirkung. Insofern ist es sinnvoll, den Absatzeffekt der Kommunikation vorher zu überprüfen. Das Problem: Die meisten Testverfahren sagen die Werbewirkung nicht oder nur unzuverlässig vorher.

„Marketing is not working! Niemals zuvor war im Marketing in Theorie und Praxis ein so radikales Umdenken erforderlich wie in unserer Zeit!“ (Kotler 2005 S.35). Um die hohen Flopraten zu senken und den Marketing Return on Investment zu steigern, sollten die modernen wissenschaftlichen Erkenntnisse in Marketing und Kommunikation viel stärker integriert werden als bisher. Dabei sollten folgende Erkenntnisse berücksichtigt werden:

1. Alles an einer Marke kommuniziert und kann die Beziehung und Markenwahl der Konsumenten zu Gunsten einer Marke verändern: Bilder, Geräusche, Musik, Symbole, Körpersprache, dargestellte Beziehungen und viele andere Faktoren bis hin zum Distributionsort können das Markenimage stark beeinflussen.
2. Für den Werbeerfolg sind nicht nur die expliziten, sondern auch die impliziten Kommunikationsinhalte und Assoziationen relevant.
3. Werbung kann auch große Wirkung erzielen, die überhaupt nicht bewusst wahrgenommen oder erinnert, sondern ausschließlich unbewusst decodiert und verarbeitet wird.
4. Menschliche Verhaltensentscheidungen werden immer durch Emotionen beeinflusst. Rein rationale Entscheidungen sind nicht möglich.
5. Erfahrungsgemäß können emotionale Unterschiede im Detail große Absatz-Unterschiede auslösen (Kremin et al. 2014; Mayer de Groot 2014 b/c/d/e, 2012a/ b; Keller et al 2013)
6. Gute emotionale Werbung erzielt höhere nachhaltige Image-, Absatz- und Profiteffekte als gute rein rationale Werbeaufträge (Binet/Field 2013 S. 52f.).
7. Zur Analyse der durch Kommunikation ausgelösten Imageveränderungen sind bei etablierten Marken immer zwei Image-Messungen erforderlich (vor und nach der Werbeexposition).
8. Der Return on Investment von Marketing- und Kommunikations-Maßnahmen ist nachweisbar zuverlässig vorhersagbar. Das vorgestellte Sales Effect Test Markenwahl-Kriterium entspricht den aktuellsten wissenschaftlichen Erkenntnissen und hat sich in über 1.000 Marktsimulationen bewährt. Es liegen veröffentlichte Fallstudien und Validierungen mit Kunden vor.
9. Es gilt – leider noch immer – die Aussage von Achenbaum (1985 a/b) auf der AMA Conference 1985: *„Die traditionellen Markt- und Werbetests entsprechen nicht den Anforderungen. Es macht keinen Sinn, Geld für Forschung auszugeben, die unzuverlässig ist. Markenartikler sollten die Anwendung von unzuverlässigen Marktforschungsmethoden mit geringer Prognosequalität vermeiden – auch wenn ihr Einsatz noch so wünschenswert erscheinen mag.“*
10. Auch markenstrategisch ist es deutlich effektiver, einen markenführungsorientierten Pretest zu wählen, der relevante Erkenntnisse über die Kommunikationswirkung liefert. Veraltete Methoden bevorzugen hingegen Werbung, die *„langweilig, austauschbar ist oder sogar missfällt“* (Kastenholz et al. 2004, S. 36) und in der Regel zu schlechteren Absatzeffekten führt (Mayer de Groot 2014a).

Keywords: Werbeerfolgs-Prognose; Marketing Return of Investment; emotionale Werbewirkung; Markenführung; Validierungen

1. Einleitung

Trotz aller Erkenntnisse der Wissenschaft, trotz aller Studien und Pretests wird in Marketing und Kommunikation immer noch viel Geld verbrannt. Entweder sind Produkte nicht sauber positioniert und bieten keinen überzeugenden Nutzen oder die Kommunikation spricht die Zielgruppe nicht richtig an. Die Auswirkungen falscher Positionierungen und irrelevanter Botschaften lassen sich in wenigen, erschreckenden Zahlen zusammenfassen:

- 94 Prozent aller neuen Produkte werden keine großen Erfolge (Mayer de Groot 2010).
- 73 Prozent aller neuen Produkte sind Flops und nach 2 Jahren im Markt nicht mehr erhältlich. (o.V.: LZ 2005; Basis: rund 30.000 Neuheiten im Lebensmittelhandel und Drogeriemärkten im Beobachtungszeitraum Mai 2003 bis April 2004).
- 67 Prozent der Werbung für etablierte Produkte und 45 Prozent der Werbung für neue Marken erzielen keine Absatzwirkung (Mayer de Groot 2012 b/c, 2011a; Basis: 830 Werbe-Posttests und über 1.000 Pretests).

Der Erfolg bleibt aus, obwohl die meisten Marketingmaßnahmen zuvor getestet wurden. Erfolgreich getestete Produkte und Kampagnen scheitern häufig in der Marktrealität und umgekehrt. Die Kritik an der Forschungspraxis groß. So stellte beispielsweise Broadbent (2004 S. 150) auf der weltweiten Admap Advertising Research Conference fest: *„In den letzten 40 Jahren hat sich nichts verändert“* und *„es gibt keine veröffentlichten Beweise, dass Pretests überhaupt zu zuverlässigen Vorhersagen fähig sind“*.

Laut einer Umfrage (Esch/Beyer 2011) können 88 Prozent aller befragten deutschen Unternehmen bisher den Markenwertschöpfungs-Prozess nicht analysieren oder gezielt beeinflussen. In einer internationalen Befragung von 1.200 Vorstandsvorsitzenden der Fournaise Marketing Group (2014) konnten 79 Prozent den Marketing Return of Investment nicht ermitteln. 70 Prozent der Marketingabteilungen erreichten im Jahr 2013 nicht ihre Ziele. Dies bestätigt erneut die Erkenntnis: Wenn man es nicht messen kann, kann man es nicht managen (Drucker 1979). Eine der großen Herausforderungen der Marketing-Forschung ist deshalb die zuverlässige Vorhersage von Flops oder Erfolgen von neuen Produkten und Kommunikations-Maßnahmen. Zudem ist die Ermittlung von konkreten Optimierungshinweisen im Falle von unbefriedigenden Testresultaten wichtig.

In dem Beitrag wird auf Basis neuester wissenschaftlicher Erkenntnisse dargestellt, wie Kommunikation Veränderungen von Markenimages und der Markenwahl auslösen kann und wie diese in Messoperationen umgesetzt werden sollten. Es werden zudem die Ursachen analysiert, warum die konventionellen Werbe-Pretest-Methoden oft zu unzuverlässigen Prognosen kommen. Es wird herausgearbeitet, welchen Mehrwert eine neue Generation an markenführungsorientierten Werbewirkungs-Pretests nachweisbar liefern kann.

2. Kernziele der Werbung und Kommunikationsforschung

In der Literatur werden viele unterschiedliche Werbeziele aufgelistet. Neben Aufmerksamkeit und Markenbekanntheit werden beispielsweise Kriterien wie Produkt- und Markenwissen, Einstellungen, Probier- sowie Wiederkäufe und Marktanteilsgewinne genannt (Pepels 1994). Für Markenartikler ist zudem der Beitrag der Kommunikation zur Markenführung und Positionierung besonders relevant. Für Markenanbieter ist es wichtig, die Attraktivität und den Preisbewilligungsspielraum einer Marke (z. B. gegenüber Wettbewerbs- und Handelsmarken) zu erhöhen oder aufrechtzuerhalten. Eine Untersuchung von 996 Werbekampagnen in der IPA Effectiveness Databank hat ergeben, dass es langfristig deutlich erfolgreicher und profitabler ist, sowohl Volumen- als auch Preisakzeptanz-Ziele zu realisieren (Binet/Field 2013 S.18).

Werbung ist eine Marketing-Investition, die sich nachweisbar rechnen soll. Insofern interessiert vor allem, was die Kommunikation für die beworbene Marke bewirkt.

Die Markenwahl wird immer durch die Wahrnehmung relevanter, markenspezifischer Eigenschaften und / oder emotionaler Vorteile bestimmt. Das Ziel jeder Marketing-Kommunikation ist, das Image einer Marke so zu verändern oder aufzubauen, dass das Wahlverhalten zu Gunsten der beworbenen Marke positiv beeinflusst wird. Das Kernziel der Werbung ist es letztlich, die Absätze zu steigern. Das ist die eigentliche Existenzberechtigung der Werbung. Es mag mehrere Zwischenschritte dahin geben, aber Teilziele wie z.B. Aufmerksamkeit, Attraktivität, Glaubwürdigkeit und Markenpassung können keine Endziele sein. Die Analyse dieser und anderer Teilziele ist aber im Falle von unbefriedigenden Testresultaten wichtig zur Ermittlung von konkreten Optimierungshinweisen.

Bei starken etablierten Marken kann es auch genügen, die Markenerinnerung über die Kommunikation eindeutiger Werbesymbole zu erhöhen und dadurch die Produktverwendung anzuregen („Salience“).

Beispiele hierfür sind oder waren das Becks Schiff mit grünen Segeln, der Krombacher See, der Marlboro Cowboy oder die Milka Kuh. Falls eine Werbung keinen nachweisbar positiven Umsatzeffekt erzielt, ist sie meistens eine Fehlinvestition. Der frühere Werbeleiter Zyman (2002 S. 1) der Coca-Cola Company stellt fest: „Wenn Werbung keine Absatz-Wirkung erzielt ist sie eine kolossale Geldverschwendung und wenn Sie nicht vernünftiger vorgehen, können Sie letztlich sogar Ihre Marke zerstören“.

Entsprechend ist die eigentliche Kernaufgabe eines Werbe-Pretests zuverlässig zu analysieren, ob und wie stark die Werbung – bewusst und unbewusst – das Markenimage und damit die Nachfrage beeinflusst (und wie sie ggf. gezielt zu diesem Zwecke optimiert werden sollte). Die Mehrzahl aller Werbepretests beschäftigt sich allerdings bisher mit der Wahrnehmung der Werbung selbst, aber kaum mit Veränderungen, welche die Kommunikation am Marken-image und der Markenwahl bewirkt. Damit eignen sich Werbetests dieser Art kaum für die Markenführung oder die Vorhersage des wirtschaftlichen Kommunikations-Erfolges. Abbildung 1 verdeutlicht die Unterschiede im Erhebungs- und Analyseumfang.

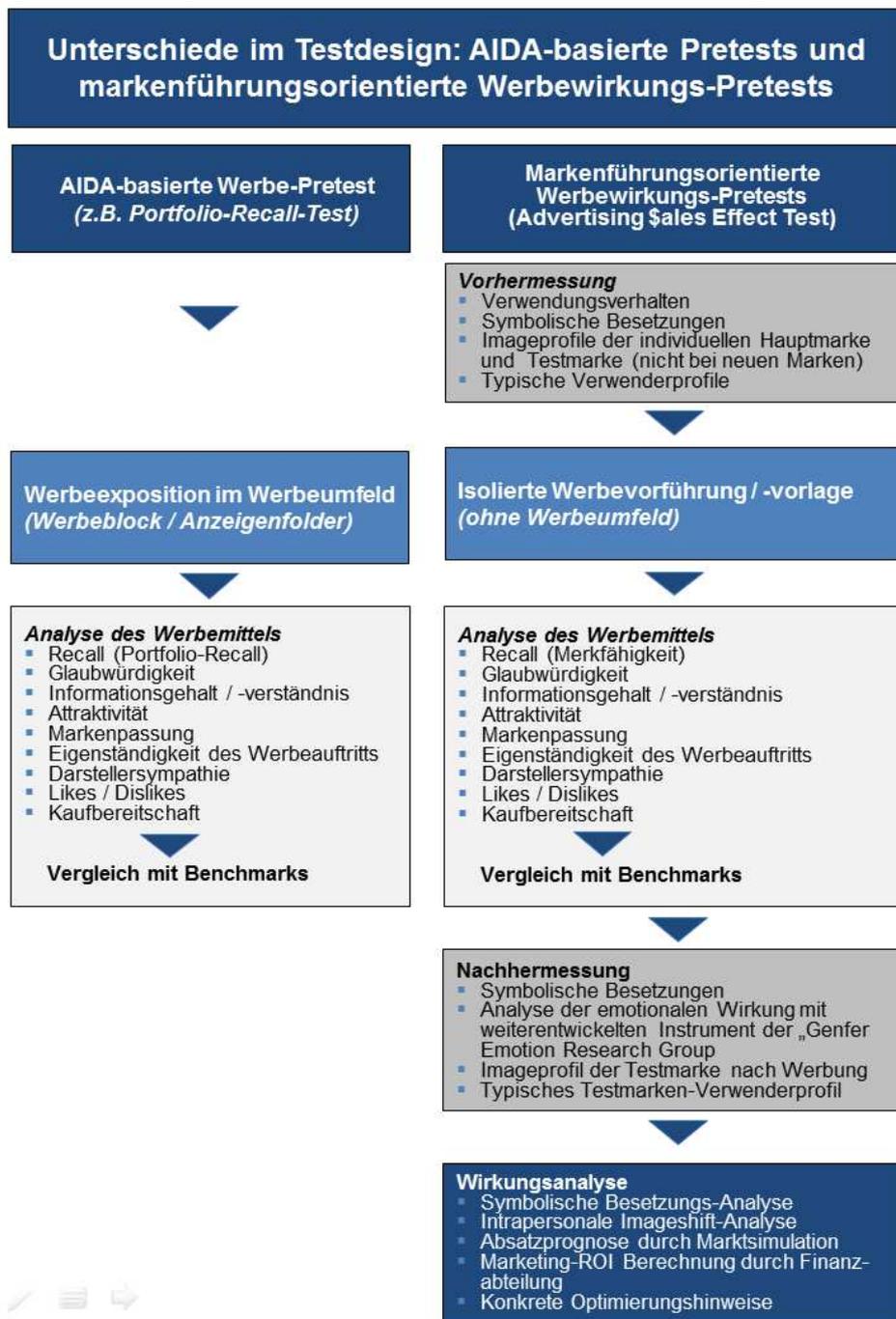


Abbildung 1: Unterschiede im Testdesign: AIDA-basierte Pretests und markenführungsorientierte Werbewirkungs-Pretests

3. Falsche Annahmen führen zu falschen Prognosen

3.1 Die AIDA-Formel ist falsch, beherrscht aber immer noch die Werbeforschung.

Die Mehrzahl der weltweit angewendeten und vor Jahrzehnten standardisierten Werbe-Pretests ignoriert die neuesten wissenschaftlichen Erkenntnisse. Trotz einiger Neuerungen beruhen die meisten Werbe-Pretests immer noch auf Vorstellungen, die vor mehr als 100 Jahren entwickelt wurden. Die sogenannte AIDA-Formel (Attention, Interest, Desire, Action) stammt aus dem Jahre 1898 und war eigentlich als Anleitung für Verkaufsgespräche gedacht. Explizit identifiziert sich aktuell kaum ein Institut mit diesem Modell. Dennoch beherrscht der AIDA-Ansatz (unter den unterschiedlichsten Bezeichnungen) prinzipiell und implizit immer noch die kommerzielle Werbe-forschung. Es gibt verschiedene Gründe, warum sich das AIDA-Modell so lange hält: Es ist ein-fach, vordergründig plausibel, merkfähig und lässt die Werbe-wirkung leicht messbar erscheinen. Zudem fühlen sich Unternehmen und Manager mit (der Fiktion) rational handelnder Verwender wohl. Für Institute ist das AIDA-Modell zudem praktisch und kosten-günstig. Sie können meist einen standardisieren Fragebogen für alle Marken und Produktkategorien verwenden, da nur das Werbemittel an sich analysiert wird – aber nicht die Image- und Absatz-Wirkung auf die jeweils beworbene Test-Marke.

Die Kritik am AIDA-Ansatz ist groß. Um nur einige ausführliche Analysen zu nennen: Heath und Feld-wick (2007) halten das Modell für viel zu einfach und es steht im Widerspruch zu zahlreichen neuen Erkenntnissen der Psychologie und Hirnforschung, indem es die emotionale, unbewusste Werbewah-nehmung ausschließt. Vakratsas und Ambler (1999) fassen die Essenz von über 250 wissenschaftlichen Studien in zwei zentralen Punkten zusammen: Zum einen ignoriert AIDA das individuelle Wissen aus vorangegangenen Erfahrungen und zum anderen arbeitet das Gehirn voll-kommen anders, nicht in der angenommenen Hierarchie und zeitlichen Sequenz. Das AIDA-Modell ist somit wissenschaftlich falsch und ein weiterer Einsatz ist nicht ratsam. (Mayer de Groot 2014a)

**Die rationale Verarbeitungskapazität des Gehirns ist begrenzt:
Die richtigen Emotionen entscheiden.**

In jeder Sekunde...

... erhält das
Gehirn circa

... kann das Gehirn
maximal rational
verarbeiten



**Über 95% aller menschlichen Verhaltens-„Entscheidungen“
werden unbewusst oder emotional getroffen - nicht rational.**

Abbildung 2: Die rationale Verarbeitungskapazität des Gehirns ist limitiert

3.2 Der Verwender entscheidet überwiegend unbewusst und emotional – nicht rational

Wissenschaftlich ist der „Homo oeconomicus“ schon lange tot. Nur nicht im Marketing. Nahezu alle Marketing- und Marktforschungs-Techniken gehen implizit immer noch vom rational handelnden Mensch aus. Aber der Konsument benimmt sich völlig anders. Nach den Erkenntnissen der modernen Hirnforschung verhalten sich Menschen keineswegs bewusst und rational. Über 95% aller menschlichen „Entscheidungen“ werden anhand unbewusster, emotionaler Bewertungskriterien getroffen – nicht rational.

In jeder Sekunde nehmen Menschen über die Sinnesorgane circa 11 Millionen Bits an Informationen auf. Rational verarbeiten kann das Gehirn in einer Sekunde aber nur maximal 50 Bits, durchschnittlich 40 Bits (wie Abbildung 2 zeigt). Das entspricht einer kurzen Telefonnummer oder einem dreisilbigen Wort wie „Portugal“. Von jeder Markenbotschaft nehmen wir Menschen also nur einen winzigen Bruchteil bewusst wahr, über 99,9 Prozent werden im „limbischen System“ unbewusst verarbeitet und dekodiert. (Kahneman 2011, Ariely 2011, Lehrer 2009, Gigerenzer 2007, Kast 2007, Mayer de Groot und Salomon 2007, Zaltman 2003, Damasio 2003, 1994)

Die meisten Werbe-Pretests gehen immer noch davon aus, dass nur der Anteil der Werbung wirken kann, der bewusst erinnert wird. Dies ist aber falsch. So haben beispielsweise Shapiro et al. (1997) nachgewiesen, dass Werbung unbewusst verarbeitet werden und wirken kann, auch wenn sie bewusst keine Aufmerksamkeit erzielt.

Marketing braucht daher Werbe-Pretests, die auch die unbewussten Kommunikations-Effekte direkt und indirekt zuverlässig messen, sonst wird nur ein Bruchteil der möglichen ausgelösten Wirkung erfasst. Denn Kommunikation wirkt hauptsächlich unbewusst.

3. Zur Bedeutung und Messung von Emotionen in der Werbung

Der Erfolg von Marken-Positionierungen und deren Kommunikation stellt immer ein Grenzproblem dar: Nur wenn eine Marke ein relevantes emotionales (und faktisches) Bedürfnis sehr präzise trifft, besteht eine echte Chance auf durchschlagenden Erfolg. Erfahrungsgemäß können emotionale Unterschiede im Detail große Absatz-Unterschiede auslösen. Um nur einige Werbe-Beispiele zu nennen: Axe (Mayer de Groot 2003), Du darfst (Mayer de Groot 2014 c), Dymo LabelWriter (Lübbe et al 2005), Iglo (Mayer de Groot/Plüm 2012), Kinder Pingui (Mayer de Groot 2014 c), Leitz (Lübbe et al 2004), Lefax (Kleij 1996), Mey (Mayer de Groot 2012a), Perrier (Mayer de Groot 2014 c) sowie die WGV-Versicherungen (Keller et al. 2013). Für die Werbewirkung sind nicht nur die expliziten, sondern auch die impliziten Kommunikationsinhalte und Assoziationen relevant. Alles an einer Marke kommuniziert und kann die Beziehung und Markenwahl der Konsumenten zu Gunsten einer Marke verändern: Bilder, Geräusche, Musik, Symbole, Körpersprache, dargestellte Beziehungen und viele andere Faktoren können die Werbewirkung stark beeinflussen.

Eine Untersuchung von 996 Werbekampagnen zeigt, dass gute emotionale Kommunikation deutlich erfolgreicher und profitabler ist als gute rationale Werbung (Binet/Field 2013, S. 52 f.), wie Abbildung 3 anhand von Schlüsselkriterien verdeutlicht. Da es sich hierbei um Einreichungen für einen Werbungs-Effektivitäts-Preis handelt, sind die bewirkten Effekte nicht repräsentativ. Allerdings bestätigen auch unsere Pretest-Erfahrungen, dass emotionale Werbeauftritte deutlich besser wirken können als „rein“ rationale.

Eine besondere Beachtung finden daher Emotionen bei der Entwicklung neuer Ansätze zur Werbewirkungs-Messung. Aber nicht jede Art von Emotion ist verhaltensrelevant und wirkt absatzsteigernd. Werbung versucht oft direkte Emotionen auszulösen. Dies gelingt ihr häufig auch, denn attraktive Bilder oder Musik setzen direkt am Belohnungszentrum im Gehirn an und lösen so Emotionen aus. Diese direkten Emotionen führen jedoch zunächst nur dazu, dass die Werbung gern gesehen oder gehört wird. Einige Werbe-Pretests erfassen diese Art von Emotionen.

Kaufimpulse lösen jedoch meist nur erwartete Emotionen aus, die mit den individuellen Konsequenzen der Markenverwendung zu tun haben. Es muss der Kommunikation nachvollziehbar gelingen, eine Vorstellung oder Ahnung zu vermitteln, wie es sich anfühlt, wenn ich diese Marke verwende. Auch bei der Dekodierung von Werbebotschaften sucht unser Unterbewusstsein nach Antworten: Was habe ich davon, wenn ich diese Marke kaufe? Hat das positive, keine oder gar negative Konsequenzen für mich? Wie werde ich mich bei und nach der Verwendung fühlen? Wie fühlt es sich an, wenn andere Personen sehen oder erfahren, dass ich diese Marke verwende?

Emotionale Kampagnen erzielen größere und nachhaltigere Wirkung als „rein rationale“.

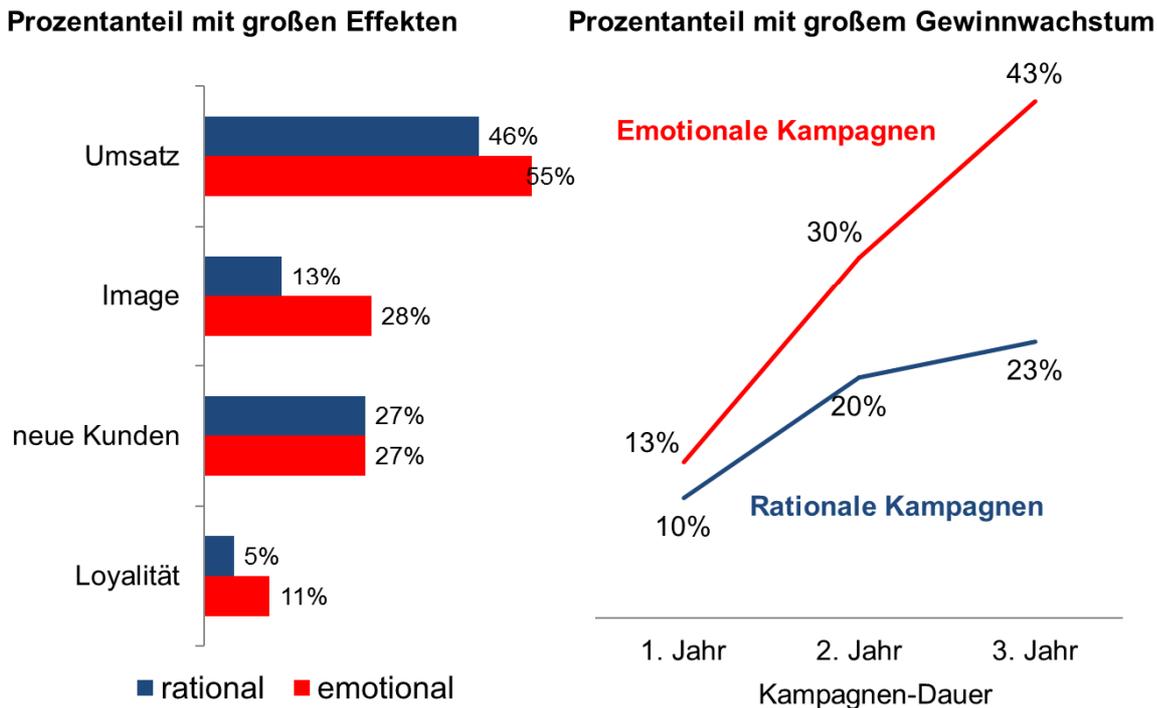


Abbildung 3: emotionale Werbeauftritte wirken besser als „rein“ rationale

Quelle: In Anlehnung an Binet/Field 2013 S.52f.

Wirksame Werbung muss nicht zwingend Emotionen zeigen, aber immer einen Belohnungseffekt im Gehirn auslösen. Viele zunächst hoch gelobte „emotionale Kampagnen“, die starke Gefühle (glückliche Menschen, Liebe, Spaß bis hin zu traurigen Anlässen und Schock-Werbung) kommunizieren, erzielten trotzdem kaum Absatz-Effekte oder wirkten sogar negativ. Denn die Empfänger können zwar die direkten Gefühlssignale in der Werbung nachfühlen, aber die Botschaft motiviert individuell zu wenig das Belohnungszentrum im Gehirn, um eine Verhaltensänderung zu erreichen. Wirksame Werbung löst handlungsorientierte Emotionen aus, wenn der Stimulus von der Zielgruppe meist unbewusst so interpretiert wird, dass die Verwendung der beworbenen Marke positive Konsequenzen für das jeweilige individuelle Wohlbefinden hat. Insofern ist die Unterscheidung zwischen direkt werblich ausgelöst und erwarteten Emotionen bei einer späteren Marken-Verwendung wichtig.

Soweit bekannt, ist der „Ad Sales Effect“-Werbewirkungstest bislang weltweit der erste validierte Test, der die neuen Erkenntnisse der Verhaltens-Ökonomie (Behavioral Economics), der Psychologie der Emotionen, der Hirnforschung sowie zahlreicher weiterer Wissenschaften aufgreift und in konkrete Messoperationen umsetzt. Er definiert eine neue Generation an Werbe-Pretests. Die Messung der mit der Markenverwendung antizipierten Emotionen (nicht der zum Werbemittel!) erfolgt mehrdimensional:

1. durch eine Analyse der bewirkten bewussten und unbewussten Imageveränderungen im intrapersonalen Vergleich der beworbenen Marke (individuelles Markenimage im Vorher- / Nachher-Vergleich)
2. durch die Messung mit einem auf Basis des von der „Genfer Emotion Research Group“ weiter entwickelten Instrumentariums (Sacharin et al. 2012)
3. durch die Analyse des assoziierten typischen Verwender-Leitbildes im Vergleich zu dem zuvor erhobenen Verwender-Profil der individuellen Hauptmarke
4. darüber hinaus werden oft auch quantitativ spontane, emotionale symbolische Besetzungen genutzt. Dies setzt allerdings eine projektive Vorstudie voraus, in der zuvor qualitativ Kategorie- und Markenspezifisch relevante Bilder des „Londoner Limbique Emotional Explorer Symbolic Sets“ herausgearbeitet wurden.

In der Folge werden die Analyse der bewirkten Imageveränderungen und die symbolischen Besetzungen vertiefend diskutiert, um relevante und innovative Erkenntnisse sowie zu berücksichtigende Aspekte stärker herauszuarbeiten. Zudem sind die oben genannten Techniken in der Literatur gut dokumentiert (2 und 3).

4.1. Analyse der bewirkten Imageveränderungen und zur Testanlage

Eigentlich sollte die Messung der Kommunikationsleistung in Form der durch die getestete Werbung ausgelösten Imageveränderungen zum Pflichtenheft jedes Werbe-Pretests gehören.

Aber die meisten konventionellen Werbe-Pretests nehmen entweder keine oder nur eine Imagemessung vor. Die Ursache ist, dass bei den AIDA-typischen „Portfolio-Recall-Tests“ sowie weiteren Test-Ansätzen eine Imagemessung vor der Werbe-Exposition unvereinbar ist mit der klassischen Recall-Messung (auf deren fehlende Zuverlässigkeit wird in Abschnitt 6.21 noch näher eingegangen).

Nur eine Imagemessung vorzunehmen, bezeichnete bereits Rehorn (1983) in seinem Grundsatzwerk über Pretests ausdrücklich als Beispiel für einen häufigen Verfahrensfehler. Bei der Überprüfung der Kommunikationsleistung für eine bereits bestehende Marke interessiert insbesondere, ob und wie die Werbung das Image und die Abverkäufe der Testmarke beeinflussen wird. In vielen Tests wird hierfür nur eine Imagemessung durchgeführt. Selbst wenn dabei unterschiedliche Fragestellungen und -techniken angewendet werden, ist das Ergebnis trotzdem invalide. Denn das Testergebnis ist ein Konglomerat von dem bereits bestehenden Image vor der Werbe-Exposition und der durch die Testwerbung ausgelösten Imagebeeinflussung. Es kann dabei nicht getrennt werden, welcher Faktor das Ergebnis in welchem Ausmaß beeinflusste. Um die Wirkung der Werbung (also die bewirkte Veränderung!) zu ermitteln, muss man das Ausgangsniveau bei jeder Testperson kennen! Also eine Vorhermessung vornehmen. Wie will man sonst feststellen, ob die Werbung ein Markenimage verbessert – und nicht eventuell sogar verschlechtert hat?

Bei der Analyse, wie die Werbung das Markenimage (bewusst und unterbewusst) beeinflusst, muss man daher zwei Messungen im sogenannten experimentellen Testdesign durchführen:

- eine Nullmessung vor Konfrontation mit der Werbung zur Feststellung der Ausgangssituation (Ermittlung des bereits bestehenden Images) und
- eine Zweitmessung nach Konfrontation mit der Werbung zur Überprüfung der Werbewirkung (Ermittlung der Imageveränderung oder des „Imageshifts“).

In der experimentellen Forschung außerhalb der Werbeforschung wird ein Prinzip angewendet, das nach Kerlinger (1973) „MaxMinCon“ genannt wird. Dieser Autor forderte, dass man die „systematische Varianz“ maximieren und Störvariablen möglichst ausschließen muss, um in einem Experiment die Wirkung eindeutig auf den Stimulus zurückführen zu können. Das bedeutet für die Anlage von Werbe-Pretests, dass man die untersuchte Werbung isoliert zeigen sollte, so dass sie in ihrer Wirkung nicht überlagert, verwechselt oder vermischt werden kann. Die Vorführung oder Vorlage zusammen mit Botschaften anderer Werbung, im Programm- oder Zeitschriftenumfeld können erfahrungsgemäß die Wirkung der Testwerbung als Störvariablen erheblich beeinflussen (Clancy/Kweskin 1971; Burke 1975, 1976, 1980; Murphy et al. 1979; Soldow/Principe 1981; Park/McClung 1986; Lord/Burnkrant 1988; Mayer de Groot 1992, Habermeier 1997; Mayer de Groot/Fritzen 2008).

Der übliche Ablauf eines Werbewirkungs-Pretests im experimentellen Design wird in der Folge kurz geschildert und wurde in Abbildung 1 bereits rechts gezeigt. Der Status der individuellen Hauptmarke und der beworbenen Marke werden in der Vorhermessung ermittelt und später als relevante „Messlatte“ oder Benchmark für die Bewertung der Werbemaßnahme herangezogen. Im zweiten Schritt wird die Werbung gezeigt. Anschließend an eine Werbemittel-Beurteilung wird eine Image-Nachhermessung anhand der gleichen Kriterien vorgenommen. Durch intrapersonale Vergleiche wird die durch die Werbung ausgelöste emotionale und faktische Imageveränderung (im Vorher- / Nachher-Vergleich analysiert).

Bei den Imagemessungen empfiehlt es sich erfahrungsgemäß, aus Validierungs- und Kostengründen überwiegend mit Verbalskalen zu arbeiten. Denn Verbalskalen haben sich gegenüber den Versuchen, Emotionen mit generell anwendbaren Bilderskalen (z.B. Bildsätze mit Mimik-Porträts, Gefühlsskalen bis zu Cartoon-Abbildungen) zu messen, als eindeutig überlegen erwiesen. Dies bestätigten Untersuchungen von Bosch et al. (2006) zudem sprechen richtige Erfolgsprognosen und Validierungen für die Zuverlässigkeit unseres Pretest-Verfahrens. Zu den Ursachen: Verbale Beschreibungen bringen Markt- oder Kategorie-übergreifend die individuellen Erfahrungen oder Gefühle besser zum Ausdruck als dies generell anwendbare Bilderskalen können (Bosch et al. 2006). Die Erfahrung zeigt, dass beispielsweise negative Gefühls-Effekte (wie „bedrohlich“, „eklig“, „stimmt mich traurig“, „will mich für dumm verkaufen“) besser durch Verbalskalen gemessen werden.

4.2 Zur Vorgehensweise bei der symbolischen Besetzung

Alle Menschen haben relevante Gedanken, die sie oft nicht verbal oder nicht präzise genug ausdrücken können, weil ihnen die richtigen Worte fehlen oder Sprache nicht das geeignete Instrument hierfür ist. Sie brauchen deshalb adäquate Unterstützungen und Hilfestellungen um diese mitzuteilen (Ariely 2008, Weiser 1993)

Symbolische Besetzungen sind ein Schlüsselmechanismus um Gedanken und Gefühle zu visualisieren. Sie repräsentieren diese in anderer Form. Metaphern sind fundamental für menschliche Gedanken – ohne diese kann Lernen, Denken und Fühlen nicht stattfinden. (Zaltman/Zaltman 2008, Zaltman/Coulter 1995, Lakeoff 1987) Gedanken und Gefühle werden im Gehirn („limbischen System“) oft als nonverbale Images verarbeitet. Insofern erscheint es effektiv, relevante Emotionen zusätzlich direkt durch Bilder (Symbole) zu erfassen. Diese Vorgehensweise reduziert das Risiko, wichtige Informationen im mentalen Transformationsprozess zu verlieren. Symbolische Besetzungen werden als Hilfestellung oder Weg eingesetzt, um unterbewusste Gedanken und Gefühle an die Oberfläche zu transportieren. Andere gut dokumentierte Marktforschungsansätze mit Bildern sind z.B. „Metaphor Based Communication Research“ und die „Metaphor Elicitation Technique“ von Zaltman (Zaltman/Coulter 1995). Hierbei werden allerdings Fotos aus dem privaten Bereich der jeweils Befragten eingesetzt: Diese Vorgehensweise erlaubt daher keine spätere Quantifizierung. Dies ist beim hier vorgestellten Ansatz anders.

Beim Advertising \$ales Effect Test wird mit einem ausgewählten Subset des „Londoner Limbique Emotional Explorer Symbolic Set“ gearbeitet. Dieser Symbol Satz wurde ursprünglich in einem Multi-Science Ansatz entwickelt. In einem aufwändigen Prozess wurden ursprünglich über 150.000 Bilder (www.pixelio.de) zu circa 200 aussagekräftigen Symbolen verdichtet. „Denn ein Bild bedeutet nichts, wenn es keine Geschichte erzählt.“ (Schank 1990) – Zu tiefenpsychologisch relevanten quantitativen Erkenntnissen im emotionalen Bereich kommen wir, indem wir symbolischen Besetzungen sowie Assoziations-Techniken auch bei der Quantifizierung einsetzen. Optimaler Weise werden in einer dem Werbetest vorausgehenden Stufe, in Limbique Emotional Explorern, die Bildvorlagen anzahlmäßig auf die kategorie- und markenspezifisch relevantesten Symbole reduziert.

Letztlich geht es uns darum, die bewusste und unterbewusste Bedeutung (meaning) von Marken oder Produkten herauszuarbeiten sowie die Erkenntnis, ob und wie diese durch Werbung verändert wird, unabhängig davon ob diese mental bildhaft oder verbal codiert werden. Deshalb werden die Probanden gebeten die Gründe für die Auswahl eines bestimmten Bildes zu erläutern. Zudem werden die unterbewussten Gründe für die Bildauswahl psychologisch durch ausgebildete Spezialisten interpretiert (z.B. Handelt es sich bei der dominierenden Farbe um ein „kaltes“ oder „warmes“ Blau? Welche versteckte Botschaft kommuniziert das ausgewählte Bild? etc.).

Mehrere Validierungsstudien mit insgesamt über 3.000 Befragten haben die Zuverlässigkeit der symbolischen Besetzungen bestätigt und zu relevanten zusätzlichen Erkenntnissen geführt. Es zeigten sich interessante Zusammenhänge: Je stärker Schlüsselsymbole für zentrale Markttreiber oder Kaufmotive gewählt wurden, desto höher war der durch die Werbung ausgelöste Absatzeffekt. Auch in Marktsegmentationen (z.B. von Ärzten) erbrachte die symbolische Besetzung wichtige zusätzliche Erkenntnisse.

5. Zur Ermittlung der durch die Werbung ausgelösten Absatzwirkung

Zunächst werden kurz alternative und weit verbreitete Ansätze zur Messung der durch Werbung ausgelösten Absatzwirkung geschildert und im Hinblick auf ihre Validität überprüft.

5.1 Gewinnkörbe als Ansätze zur Ermittlung der Absatzwirkung von Werbung

Bei dieser Vorgehensweise wird das Interesse an einem Gewinnkorb vor der Werbeexposition gemessen. Nach der Werbung wird erneut das Interesse an einem veränderten Gewinnkorb skaliert ermittelt, bei dem eine (oder mehrere) der ursprünglich enthaltenen Marken durch eine (oder mehrere) der beworbenen Marken der jeweils gleichen Produktkategorie ersetzt werden. Aus der Interessensdifferenz versucht man, den Absatzeffekt der Werbung zu ermitteln.

Die Problematik dieser Vorgehensweise ist, dass zusätzlich zu der eigentlichen Werbewirkung der Austausch der Marken im Gewinnkorb eine erhebliche Rolle spielt. Es ist leicht nachvollziehbar, dass es bereits allein durch den Produktaustausch zu einem stärkeren Interesse kommt, wenn z.B. eine Handelsmarke oder eine Marke mit niedrigerer Marktbedeutung durch einen Marktführer ersetzt wird. Der gleiche Effekt ist zu erwarten, falls preiswertere Marken durch Premiumprodukte (z.B. Milka-Schokolade durch Lindt) ausgetauscht werden, ein Nischenprodukt durch eine Massenmarke, etc. Man misst somit eine Mixtur aus beiden Effekten. Die eigentliche Absatzwirkung der Werbung kann nicht isoliert und gegebenenfalls sogar durch den Produktaustausch im Gewinnkorb überlagert werden.

Darüber hinaus ist anzumerken, dass die Bereitschaft ein Markenprodukt zu gewinnen, keinesfalls mit einer konkreten Kaufabsicht für dieses Produkt gleichgesetzt werden sollte. Es ist leicht nachvollziehbar, dass Befragte teurere Markenartikel, die sie sich selbst nicht leisten können, gerne gewinnen aber selbst nicht kaufen würden. Insofern sind Gewinnkörbe nicht zur zuverlässigen Prognose der Verkaufswirkung der Werbung geeignet.

5.2 Kaufbereitschaftsfragen als Ansätze zur Ermittlung der Absatzwirkung von Werbung

Zahlreiche Studien haben bewiesen, dass einfache Kaufintentions-Fragen das Kaufverhalten nur unzuverlässig vorhersagen können. So stellt auch Morwitz (2001 S. 33) zusammenfassend fest: *„Umfangreiche Untersuchungen haben etabliert, dass direkte Messungen (der Kaufbereitschaft) zu unzuverlässigen Vorhersagen führen. [...] Intentionen sind keine perfekten Verhaltensmaßstäbe wegen der zufälligen und systematischen Fehler in dieser Messgröße.“* Dies ist seit langem bekannt und in seriösen Marktforschungskreisen unstrittig (Mayer de Groot 2011a und die dort angegebene Literatur).

Institute, die noch mit Kaufbereitschaftsfragen zu Prognosezwecken arbeiten, verwenden daher meist unterschiedliche Gewichtung-Schemata. Morwitz (2001) nennt allein 6 alternative, teilweise komplexe Gewichtungsmodelle unterschiedlicher Institute, von denen sich keines als zuverlässiger erwies. Grundsätzlich gilt, gravierende und systematische Messfehler können durch eine Gewichtung zwar reduziert, aber nicht beseitigt werden. Einfache Kaufbereitschaftsfragen sind und bleiben somit unzuverlässig. Sie sind für eine zuverlässige Prognose der Absatzwirkung nicht geeignet.

5.3 „introspektiver Persuasion Score“ als Ansatz zur Ermittlung der Absatzwirkung von Werbung

Das Institut Millward Brown ermittelt einen sogenannten *„introspektiven Überzeugungswert“*. Es wird (auf einer Fünfer-Skala) gefragt, ob die Testperson glaube, dass die Marke sich nach Einsatz der Werbung besser verkaufe: *„It would make me more likely to buy the brand.“* (*„Diese Werbung veranlasst mich die Marke eher zu kaufen.“*)

Nach Angaben des Instituts „misst diese einfache Frage genauso gut, wie die sehr viel komplexeren Pre-Post-Messungen“ (Farr 1993). Allerdings widerspricht diese Behauptung allen wissenschaftlichen Erkenntnissen und Erfahrungen:

1. Generell gilt, dass indirekte Messungen immer höhere Messfehler verursachen als direkte Messungen.
2. „Es kann zu stark irreführenden Ergebnissen kommen, wenn man sich nur auf eine Messung verlässt.“ (Hollis/Twose 2007 S. 188)
3. Die Frage entspricht zudem qualitativ weitgehend einer einfachen, indirekten Kaufbereitschaftsfrage. Deren Unzuverlässigkeit wurde im vorangegangenen Abschnitt nachgewiesen.

Das Institut behauptet hingegen, dass dieser Messwert (möglichst) bei gleichzeitig hohem Recall eine hohe Prognosezuverlässigkeit erzielt (Farr/Gardiner 2001), eine kreative Exekution wird zudem als wichtig erachtet. Allerdings ist und bleibt dieser (einseitig rationale) Ansatz sehr umstritten (Hollis/Twose 2007 S. 188) – Mehrere Markenartikler und Kunden (wie z.B. Kellogg und Unilever) haben zudem in zahlreichen Anwendungen keine oder kaum eine Korrelation dieser Art der Werbeabsatz-Vorhersagen mit den eigenen Markenentwicklungen in der Realität festgestellt.

Die Unzuverlässigkeit des „introspektiven Persuasion Scores“ lässt sich auch anders nachweisen: Internationale (Post-Test-) Untersuchungen von 830 Werbekampagnen haben – wie bereits angesprochen – gezeigt, dass circa Zweidrittel der Werbung für bestehende Produkte keine Auswirkung auf den Abverkauf

hatten. Demgegenüber kommuniziert das Institut in seiner Werbung für den „LINK“-Test, (der auch den „introspektiven Persuasion Score“ verwendet), dass etwa ein Drittel der Werbung seine Ziele nicht erreiche. Die Chancen, im „LINK“-Test gut abzuschneiden, sind somit deutlich besser als in der Realität.

5.4 Relevant Set als Ansatz zur Ermittlung der Absatzwirkung von Werbung

Die Annahme einer Marken-Rangordnung im Kopf des Kunden ist in der Marketingtheorie als auch in der Marktforschung weit verbreitet. Die Methode wird oft als „Constant Sum-method“, „Chip Game“ oder „Relevant Set“ bezeichnet. Oft wird als Werbeziel vorgegeben, den Rangplatz der beworbenen Marke in spezifischen Zielgruppen zu verbessern.

Die Annahme einer wahrgenommenen Rangordnung im Kopf des Konsumenten ist aber nach den neuesten Erkenntnissen der Hirnforschung nicht haltbar. Es gibt nur zwei Plätze im limbischen System: erster Platz oder dahinter. Denn in neurowissenschaftlichen Hirncanner-Studien wird deutlich, dass die sogenannte „kortikale Entlastung“ im Kopf nur bei der individuellen Lieblingsmarke auftritt. Positive Urteile und Erfahrungen ermöglichen bei der persönlichen Hauptmarke Markenwahlentscheidungen ohne bewusstes Nachdenken (Hubert/ Kenning 2008; Kenning et al. 2007, 2005, 2002; Deppe et al. 2005). Alle anderen Marken bewirken keine emotionale Reaktion. Es spielt also keine Rolle, ob eine Marke beispielsweise an 2., 3. oder an 4. Stelle steht. Dieser „Winner-take-all“ Mechanismus findet sich an vielen Stellen des Gehirns. Der Grund ist eine Effizienz-Steigerung des Gehirns. Angesichts von circa 50.000 beworbenen Marken wäre es viel zu komplex und zeitaufwändig bei jeder Markenentscheidung über Rangreihen aller Marken nachzudenken, die man kennt. Insofern widersprechen die modernen neurowissenschaftlichen Erkenntnisse einer der Schlüsselannahmen des Relevant Set Modells.

Hinzu kommt, dass die Präferenz-Messungen unter Anwendung des so genannten „Chip Game“ meist ohne Preisinformation durchgeführt werden. Dies dürfte die Vorhersagezuverlässigkeit mindern. Umfangreiche Erfahrungen in Feldstudien haben zudem gezeigt, dass viele Testpersonen die Verteilung von Präferenzpunkten überfordert – insbesondere dann, wenn der individuelle Relevant Set groß ist, da diese Präferenzpunkt-Verteilung für alle möglichen Zweierkombinationen an Marken im Relevant Set vorgenommen wird.

Darüber hinaus sind die dem Relevant Set Modell zugrundeliegenden Annahmen äußerst restriktiv. Diese Prämissen werden bei der Anwendung in der täglichen Praxis jedoch viel zu wenig überprüft. Das Modell setzt implizit etablierte und gesättigte Märkte voraus sowie eine oligopolistische Marktstruktur, wie es um 1980 häufiger der Fall war. Heutzutage sind die Märkte aber oft stark fragmentiert, daher werden diese Bedingungen nur relativ selten erfüllt. Falls eine oder gar mehrere der grundlegenden Anwendungs-Voraussetzungen des Modells nicht zutreffen, dann sind Fehlprognosen gerade zu vorprogrammiert. Hierauf haben die Methoden-Schöpfer Silk und Urban (1978) selbst hingewiesen. Vor dem Hintergrund einiger falscher Vorhersagen bei Nichterfüllung der Prämissen haben Urban und Katz (1983) erneut die zwingende Beachtung der Modellannahmen betont.

5.4 Möglichkeiten der Absatzprognose mittels Neuromarketing

Neuromarketing ist als relativ junge Wissenschaft derzeit in aller Munde. Viele verbinden mit dem Schlagwort „Neuromarketing“ die Anwendung apparativer Verfahren der Hirnforschung auf Marketingfragestellungen. Bildgebende Verfahren, wie die funktionelle Magnetresonanztomografie (fMRT) oder die Positronen-Emissions-Tomografie (PET), haben in der Grundlagenforschung wichtige neue Erkenntnisse geliefert und wesentlich zu einem Umdenken beigetragen, wie das Gehirn funktioniert.

Diese apparativen Neuromarketing-Verfahren sind aber aus zahlreichen Gründen für die tägliche Marketingforschungspraxis wenig geeignet. Um nur einige zu nennen: Die Untersuchungssituation ist wenig biotisch, die Testdesigns sind komplex und die Geräte teuer (ein Scanner kostet ein bis zwei Millionen Euro (Huesing et al. 2006)). Dies wiederum führt zu geringen (finanzierbaren) Fallzahlen und entsprechend oft zu statischen Toleranzen, welche die eigentlichen Messwerte übersteigen. Die sehr zeitaufwändige Interpretation der bildgebenden Verfahren ist auch aus vielen anderen Gründen schwierig und erschließt sich nur Experten. Und selbst diese können bei einer Re-Analyse der gleichen Daten zu anderen Ergebnissen kommen (Grosenick et al. 2008). Professor Kenning nannte beim 1. G.E.M Forum am 18. November 2008 als Beispiel zwei Publikationen des weltweit anerkannten Neuromarketingforschers Knutson et al. (2006/2007).

In der ersten Ergebnis-Publikation wurde ein Regressions-Koeffizient von neuronalen Aktivitäten zum Kaufverhalten von nur 0,11 errechnet. Nach intensiver Kritik anderer Hirnforscher und mit dem Hinweis auf Auswertungsfehler stieg dieser Wert in seiner zweiten Veröffentlichung auf 0,85. Zudem sind die statischen Toleranzen bei geringen Fallzahlen (bei Knutson: 26 Testpersonen) sehr groß.

5.5 Zuverlässige Absatz-Prognose mit einem gut fundierten Markenwahl-Kriterium

Wie bereits angesprochen, empfangen Menschen in jeder Sekunde über die Sinnesorgane circa 11 Millionen Bits an Informationen (davon kann das Gehirn aber nur maximal 50 Bits rational verarbeiten). Wie aber bewältigt das so genannte limbische System diese Informationsflut in so kurzer Zeit? Es verwendet mehrere „limbische Abkürzungen“ gegenüber einer vollständigen Informations-Verarbeitung.

1. Menschen bewerten unterschiedliche Angebote eben nicht mit absoluten Maßstäben, sondern fokussieren sich auf relative Vorteile von einer Offerte gegenüber anderen – so genannte Anker- oder Referenzpunkt-Theorie (Ariely 2011). Zudem reduzieren sie die Komplexität der Entscheidung deutlich, indem sie die gegenwärtig hauptsächlich gekaufte Marke als Vergleichsmaßstab verwenden. Denn die aktuelle Hauptmarke bietet bei häufig gekauften Produkten individuell die „beste Problemlösung“ unter allen bereits bekannten Alternativen (Wirtschafts-Nobelpreisträger Samuelson / Zeckenhauer 1988). Sie definiert, welche erlebte Kombination an emotionalen Benefits und faktischen Eigenschaften auf welchem jeweiligen spezifischen Niveau zurzeit optimal die individuelle Bedürfnisstruktur unter allen bekannten Angeboten befriedigt.
2. Das Unterbewusstsein geht beim Vergleich von alternativen Angeboten hierarchisch vor: Es betrachtet zunächst nur das wichtigste Kaufkriterium für eine Kategorie, dann das zweitwichtigste Kriterium usw. Sobald eine der (beiden) Offerten erlebnismäßig gewinnt, wird diese gekauft (Gigerenzer 2007). Deshalb sind Marken-Positionierungen oft besonders erfolgreich, die den wichtigsten Kategorie-Benefit besetzen (Mayer de Groot 2011b; 2011d, 2008).
3. Menschen haben die Neigung sich mehr darauf zu fokussieren, was sie verlieren als was sie gewinnen könnten. „Menschen hassen Verluste [...] Vereinfacht ausgedrückt, wenn man etwas verliert, wird dies doppelt so stark negativ bewertet, als wenn man das Gleiche erhält.“ Anders formuliert: Menschen sind „Verlust Vermeider“ (Thaler/Sunstein 2009; Wirtschafts-Nobelpreisträger Kahneman (2012); ders. / Tversky (1979).

Hardie et al. (1993) haben empirisch belegt, dass negative Abweichungen von derzeit verwendeten Produkten auf starke Ablehnung stoßen. Die Analyse bestätigt eindrucksvoll das Konzept der „Loss Aversion“ bei der Markenwahl des Wirtschafts-Nobelpreisträgers Kahneman. Ergänzend geht die Regret Theory von der Befürchtung aus fehlerhafte Entscheidungen zu treffen und dem Versuch Wahlmöglichkeiten mit negativen Konsequenzen zu vermeiden, da diese zu Bedauern (= Regret) führen (Landman 1987)

Die aktuellen Erkenntnisse der Behavioral Economics, der Psychologie der Emotionen und der Hirnforschung haben somit bewiesen, dass neue Angebote vorteilhaft und frei von wahrgenommenen relevanten Nachteilen erlebt werden müssen, um nachhaltig gekauft zu werden. Dies wird in den meisten konventionellen Marktsimulations-Ansätzen unterschätzt und führt auch zum Scheitern der Trade off-Modelle.

Insofern lautet unser bewährtes Markenwahl-Kriterium: Kunden wechseln ihre aktuelle Hauptmarke langfristig nur dann, wenn die Alternative zumindest besser in einem kaufrelevanten Kriterium erlebt und zumindest gleich gut in allen anderen kaufrelevanten Dimensionen wahrgenommen wird. Falls es nicht gelingt, den Empfänger von dieser Vorteilhaftigkeit zu überzeugen, dann verspürt dieser individuelle Verwender keine Motivation und sieht keine Notwendigkeit seine bisherige Hauptmarke, mit der er in der Vergangenheit gute Erfahrung gesammelt hat, nachhaltig zu ersetzen. Dieses Markenwahl-Kriterium deckt sich zudem mit der Erkenntnis großer Untersuchungen, welche die Ursachen für Erfolge und Misserfolge beleuchtet haben, wie z.B. Davidson (1997), PIMS (Buzzell/Gale 1987), Peckham (1981) und die Lebensmittel Zeitung (1985).

Weitere Validierungen des bewährten Markenwahl-Kriteriums liefert das Endowment-Modell. Es beschreibt das Phänomen, dass Personen für ihr Eigentum einen höheren Preis verlangen als sie für ein identisches Objekt zu zahlen bereit sind. Dies wurde von Thaler (1980) empirisch validiert.

Abbildung 4 stellt das wissenschaftliche Fundament des bewährten Markenwahl-Kriteriums zusammenfassend dar.

Bei Tests von Innovationen, die eventuell neue Märkte definieren, mag die Hypothese nicht zutreffen, dass ein neues Angebot eine individuelle Hauptmarke (in einem etablierten Markt) verdrängen muss. In diesen Fällen verwenden wir ein ähnlich „hartes“ Simulationskriterium: Es zählen nur die Personen mit einer hohen angegebenen Kaufbereitschaft, welche die Testmarke jeweilig durchgängig positiv beurteilen zum Potential (sogenannte „*multivariate Konsistenzüberprüfung der Hochkaufbereiten*“) und keine relevanten negativen Bewertungen aufweisen. Den anderen glauben wir die hohe angegebene Erwerbsbereitschaft nicht. Dieses alternative Markenwahl-Kriterium führt erfahrungsgemäß zu gleich guten Erfolgsprognosen wie das zuvor genannte. Dies beweisen exemplarisch die Fallstudien des innovativen Etikettendruckgeräts Dymo LabelWriter (Lübbe et al. 2005) oder des PerfectDraft-Bierzapfsystem (Lenatz, Mayer de Groot 2007).

Sales Effect Test Marktsimulation: Methoden sind nur so gut, wie ihr wissenschaftliches Fundament und ihre nachweisbare Prognosezuverlässigkeit im Markt.

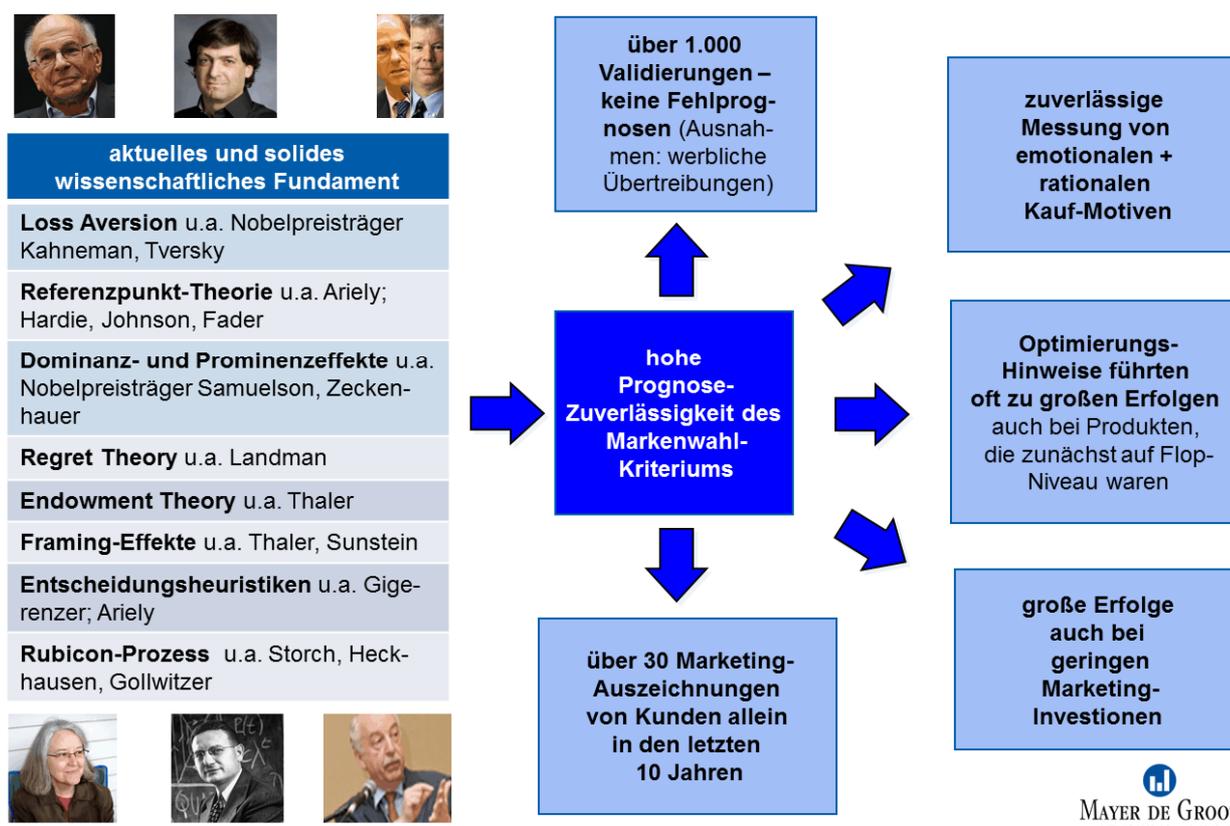


Abbildung 4: Eine Methode ist nur so gut, wie ihr wissenschaftliches Fundament und ihre nachweisbare Prognosezuverlässigkeit im Markt

5.5.1 Zur Notwendigkeit die Prognose individuell für jede Testperson durchzuführen

Die Prognose der Markenwahl wird für jedes Individuum einzeln vorgenommen – so genannter *“segment of one approach”* – so dass der residuale „error term“ praktisch entfällt. Die Analyse überprüft bei jeder Testperson, ob das Markenwahl-Kriterium erfüllt wird oder nicht. Ein Markenwechsel ist – aufgrund der Loss Aversion – nur wahrscheinlich, wenn die neue Alternative einen Vorteil bietet, aber keinen Nachteil gegenüber dem bisher hauptsächlich verwendeten Produkt aufweist. Mit anderen Worten: Das neue Angebot muss in mindestens einem kaufrelevanten Kriterium als besser und ansonsten als gleich gut erlebt werden.

Benchmarks in der bisher genutzten Form entsprechen der Loss Aversion des Wirtschafts-Nobelpreisträger Kahneman nicht und führen daher oft zu gravierenden Fehlprognosen. Benchmarks werden in der Marktforschung ermittelt, indem die Durchschnittswerte verschiedener Test-Ergebnisse errechnet werden. Dabei werden in der Regel alle gemessenen Dimensionen ohne Gewichtung berücksichtigt. Es wird implizit die klassische Annahme der Statistik über die Linearität der Erwartungswerte unterstellt: Nachteile kön-

nen durch eine größere Anzahl von Vorteilen kompensiert werden (Mayer de Groot 2013). Diese entspricht aber nur begrenzt der Realität, wie aktuelle wissenschaftliche Erkenntnisse beweisen. Die Abbildung 5 verdeutlicht, warum Benchmark orientierte Ansätze oft zu anderen Prognosen kommen als Sales Effect Markt-simulationen (eine ausführlichere Kritik an Benchmarks im Marketing findet sich bei Mayer de Groot 2012d; Haimerl/Mayer de Groot 2000.).

Konventionelle Benchmark im Vergleich mit dem Sales Effect Markenwahl-Kriterium, das der Loss Aversion entspricht.

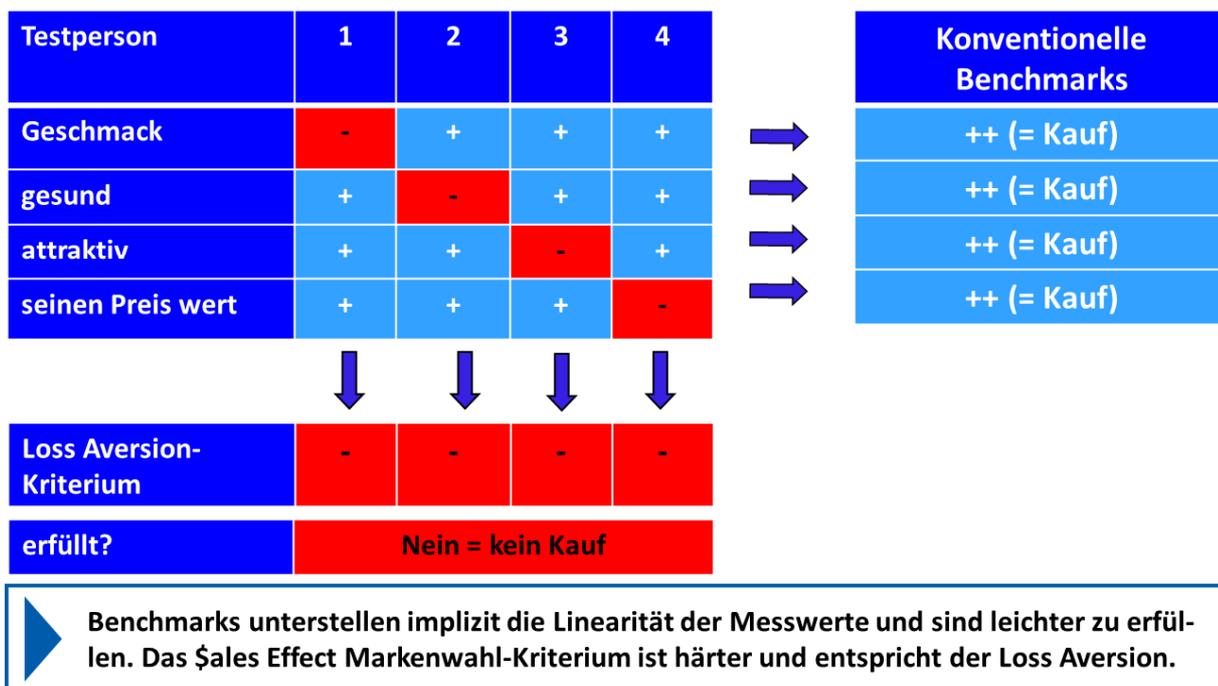


Abbildung 5: Konventionelle Benchmarks im Vergleich mit dem Sales Effect Markenwahl-Kriterium

5.5.2 Zur Prognosezuverlässigkeit Marktsimulation

In zahlreichen Anwendungen in verschiedensten Produktbereichen im In- und Ausland hat sich das wissenschaftliche gut fundierte Markenwahlkriterium in Marktsimulationen bewährt. Die Bandbreite der Erfahrungen reicht von Lebensmitteln und Getränken über Kosmetikprodukte sowie freiverkäuflichen Medikamente (OTC), bis hin zu Medien sowie langlebigen Gebrauchs- und Investitionsgütern. In über eintausend Fällen, in denen die Werbung (nahezu) unverändert geschaltet wurde, lag die Prognose in der Regel innerhalb von 0,5 Marktanteilspunkten der aktuell erzielten Werte. In den bereits publizierten Fallstudien gab es nur eine Prognose-Differenz von 0,1 % bei Beck's Gold (Shaw et al. 2004 a/b), 0,1% bei Milka Tender (Mayer de Groot 2011d), 0,1 % bei Iglo del Mar und 0,2 % bei Iglo Spinat (Schäffken 2008; Mayer de Groot/Plüm 2012), 0,2 % bei NIVEA Soft in Italien (von Dassel et al. 2001 a/b/c), 0,2 % bei Sheba (Mayer de Groot 2010), 0,2 % bei Bärensnack (Mayer de Groot 2014 b/c), 0,4 % bei NIVEA Soft in Deutschland (von Dassel et al. 2001 a/b/c) und 0,5 % bei Dymo LabelWriter (Lübbe et al. 2005). Ähnlich gute Übereinstimmungen wurden in zahlreichen weiteren Studien erzielt, von denen einige publiziert wurden: Efasit (Mayer de Groot 2009), Enzym Lefax N (Kleij 1996), Hasseröder (Lennatz et al. 2007a) ibutop (Reese et al. 2004), Lefax (Kleij 1996), Leitz (Lübbe et al. 2004), Niederegger (Strait et al. 2007), Perfect Draft (Lennatz et al. 2007b); Vorwerk (Weber et al. 2006), WD-40 (Gill/Mayer de Groot 2008). Selbst bei der Vorhersage der Wirkung von Promotions in nur 5 Aktionswochen lassen sich relativ marktnahe Prognosewerte erzielen. So betrug die Abweichung der verkauften Aktionsware in der WD-40 Volks-Profispray Bildaktion nur 0,9% (Gill/Mayer de Groot 2011). – Abbildung 6 zeigt exemplarisch einige Validierungsbeispiele.

Einige Validierungs-Beispiele zur Prognosezuverlässigkeit

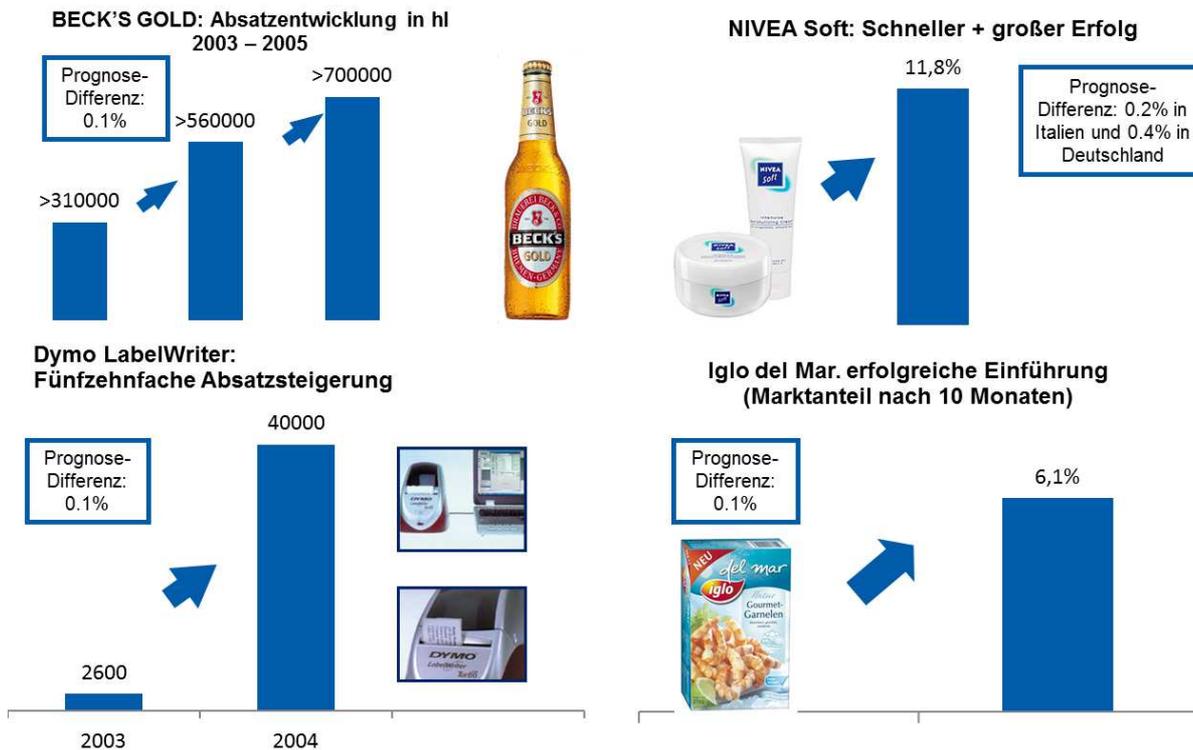


Abbildung 6: Einige Validierungs-Beispiele zur Prognose-Zuverlässigkeit

Die (ehemaligen) Beiersdorf-Manager von Dassel und Wecker (Marktforschungsleiter von 1971-2002) bestätigen, dass „die Prognose-Übereinstimmungen mit den realen Markt-Ergebnissen von NIVEA Soft definitiv mit zu den besten Resultaten gehören, die jemals von Beiersdorf mit unterschiedlichen Markt-Simulationsverfahren erzielt werden konnten.“ (von Dassel et al. 2001a S.33) „Richtig erstaunlich war für mich die Prognose-Zuverlässigkeit. [...] Bei einem zweistelligen Millionen-Betrag im ersten Jahr gab es nur eine Abweichung von 9.000 €.“ (von Dassel 2003 S.4).

In mehr als 1.000 Werbewirkungs-Prognosen wurden größere 'negative' Varianzen zwischen dem vorhergesagten Absatzeffekt und den Marktergebnissen nur in seltenen Fällen beobachtet. Ursache hierfür waren nahezu ausschließlich sogenannte „werbliche Überversprechen“. Wenn der Verwender einen relevanten Unterschied zwischen einem überzogenen Werbeversprechen und der individuellen Bedürfnisbefriedigung der beworbenen Angebote bei der Verwendung erlebt, dann ist der erste Kauf erfahrungsgemäß meistens schon der letzte.

5.5.3 Gezielte Optimierungshinweise Dank hoher Diagnose-Fähigkeit

Die Diagnose-Kraft des Advertising \$ales Effect Test ist hoch. Die Methode gibt klare Hinweise, warum das Markt-Potenzial eine bestimmte Größe hat und nicht größer ist. Neben der Potenzialgruppe, in der das Markenwahlkriterium erfüllt ist, werden in der Analyse weitere Werbereaktionsgruppen ermittelt. Besonders interessant sind die „Optimierungs-potenzialgruppen“, bei denen das Markenwahlkriterium nur auf einer Imagedimension (sowie eventuell daraus resultierend auch im erlebten Preis-Leistungs-Verhältnis) nicht erfüllt wird. Insofern zeigt die Analyse auf, welche Imagedimension gezielt verbessert werden sollte, um zusätzliche Zielgruppen überzeugen zu können und die Absatzwirkung gezielt zu steigern.

Mit Hilfe dieser Optimierungshinweise ist es häufig gelungen, Werbe-Kampagnen und Produkte zu Erfolgen zu entwickeln, die zunächst auf Flop-Niveau waren. Wenig bekannt ist, dass erst gezielte Optimierungen zu großen Markterfolgen führten bei Beck's Gold (Shaw et al. 2004b); Dymo LabelWriter (Lübbe et al. 2005), Iglo 4Sterne Menü (Mayer de Groot et al. 2004), Iglo del Mar (Schäffken 2008, Mayer de Groot/Plüm 2012), Nivea Soft (von Dassel 2003, Mayer de Groot 2003, S. 362; von Dassel et al. 2001a), Sheba (Mayer de Groot 2010b) sowie bei der WD-40 Volks-Profispray Bildaktion (Gill / Mayer de Groot 2011).

6. Werbemittel-Beurteilung

In Grundlagenstudien von Wells et al. (1971) wurden ursprünglich knapp 600 Statements weltweit gesammelt, die zur Werbebeurteilung verwendet wurden. Das Ziel war es, diese Vielzahl auf möglichst wenige Faktoren zu reduzieren. Zu diesem Zweck wurde mit mehreren, unabhängigen Stichproben, verschiedenen Sätzen von Beurteilungskriterien sowie unterschiedlichen Statements gearbeitet. In mehreren Faktorenanalysen ergaben sich praktisch immer dieselben Grunddimensionen. In diesem Zusammenhang ist interessant, darauf hinzuweisen, dass sich 10 Jahre später in einer anderen Studie nahezu die gleichen Faktorenlösungen ergaben (Aaker/Bruzzone 1981). Der Umfang und die Stabilität der Ergebnisse ermöglichen daher eine Standardisierung der Werbemittel-Beurteilungskriterien. Es können alle relevanten Aspekte der Werbemittel-Beurteilung mit folgenden sieben Dimensionen abgedeckt werden:

- Aufmerksamkeits-Wirkung / Recall
- Eigenständigkeit bzw. Austauschbarkeit des Werbeauftritts
- Attraktivität der Exekution
- Informationsgehalt der Werbung
- ausgelöste Irritation
- Brand / Category Fit (Werbung passt zur Marke / zur Kategorie)
- Identifikation mit der Zielgruppe

Diese gestützte Erhebung der Werbemittel-Beurteilungs-Dimensionen durch faktoranalytisch geeichte Statements dienen ebenso wie offene Fragen (Likes / Dislikes sowie Werbebotschafts-Verständnis) in erster Linie zur Analyse und gegebenenfalls zur Verbesserung der Werbemittel-Exekution. Als Parameter für die Qualitätsbeurteilung einer Werbevorbereitung sind sie – im Gegensatz zu den herkömmlichen Pretest-Methoden – eher von nachgelagerter Bedeutung (eine hohe Attraktivität und Eigenständigkeit des Werbeauftritts ist erfahrungsgemäß keine Garantie für die Erzielung hoher Image- und Absatzeffekte). (Mayer de Groot 2002)

6.1 Externe Validierung der Werbemittel-Beurteilungskriterien

Ein Werbe-Pretest ist nur so gut, wie seine Fähigkeit, die späteren Resultate im Markt vorherzusagen. Bei der Validierung ergibt sich allerdings die Schwierigkeit, dass die meisten der sieben angeführten Werbemittel-Beurteilungskriterien in den üblichen Trackingstudien nicht systematisch erfasst werden. Eine Validierung war daher nur in einer höheren Anzahl von Einzelfällen möglich. Eine Ausnahme stellt der Werbe-Recall dar, der in allen Werbe-Trackingstudien üblicherweise gemessen wird.

6.2 Externe Validierung der Recall-Prognosen

Der Recall steht im Pflichtenheft nahezu jeder Werbe-Pretest-Ausschreibung, auch wenn seit längerem bewiesen ist, dass der Recall in keiner direkten Korrelation zur Absatzwirkung steht (Koeppler 1974; Gibson 1983; Percy/Rossiter 1992). Zudem kann Werbung auch unbewusst wirken, wie bereits angesprochen wurde. Der Recall gibt wichtige Hinweise auf die Durchsetzungsfähigkeit einer Kommunikation in der bewussten Werbeerinnerung und auf die erforderliche Intensität in der Schallfrequenz. Insofern überrascht es nicht, dass die meisten Werbe-Pretests behaupten, die Aufmerksamkeitswirkung einer Werbung zuverlässig prognostizieren zu können.

6.2.1 Unzuverlässige Recall-Prognosen konventioneller Pretests

Die Erfahrung zeigt auf der Basis von mehreren hundert Studien, dass die meisten Pretests nicht dazu in der Lage sind, die Werbeerinnerung zuverlässig vorherzusagen. In Tracking-Studien werden meistens völlig andere Recallwerte ermittelt als von nahezu allen Werbe-Pretests vorhergesagt wurden. Dies haben Mayer de Groot et al. (2001) für die weit verbreiteten Werbe-Pretest-Verfahrensgruppen der „Portfolio-Recall-Tests“ (Milward Brown 1991), der Recognition-Tests sowie der Werbe-Pretests mit TV-Programmumfeld nachgewiesen. Dies bestätigt auch eine Studie von Kastenholz et al. (2004) in der 60 TV Spots analysiert wurden. Ein Vergleich der vorhergesagten Werbeerinnerung in verschiedenen Pre-Test-Systemen (IPSOS, Ameritest und Millward Brown) ergab, dass diese zum Teil sehr gering oder sogar negativ mit dem „Day after Recall“ korrelierten.

Vor allem die übliche Portfolio-Recall-Test-Messung ist auch in anderer Hinsicht problematisch: Sie belohnt vornehmlich Kommunikation, bei der die Marke oft erwähnt wird und sehr rationale Produktinformationen kommuniziert werden: „Es wird Werbung belohnt, die langweilig, austauschbar ist oder sogar missfällt, wir halten die Nützlichkeit von Recallwerten daher für stark limitiert.“ (Kastenholz et al. 2004 S. 36).

6.2.2 Zuverlässige Recall-Prognosen durch Merkfähigkeits-Fragen

Werbe-Pretests, die Veränderungen in der Markenwahrnehmung durch Vorher-/Nachher-Vergleiche ermitteln, sind unvereinbar mit direkten klassischen Recall-Messungen. Durch die Erhebung des Markenimage vor der Werbeexposition erfahren die Testpersonen, um welche Marke es geht. Dies würde zwangsläufig zu viel zu hohen klassisch gemessenen Recallwerten führen.

Wells hat bereits (1964 a/b) vorgeschlagen, Werbung auch hinsichtlich ihrer Wirkung auf die Merkfähigkeit durch die Testpersonen bewerten zu lassen. In einer Studie von Zinkhan/ Fornell (1985) hat sich diese Messung sehr gut bewährt. Menschen können offensichtlich gut beurteilen, ob eine getestete Werbung

Recall: hohe Korrelation von Pre-/ Posttest-Ergebnissen

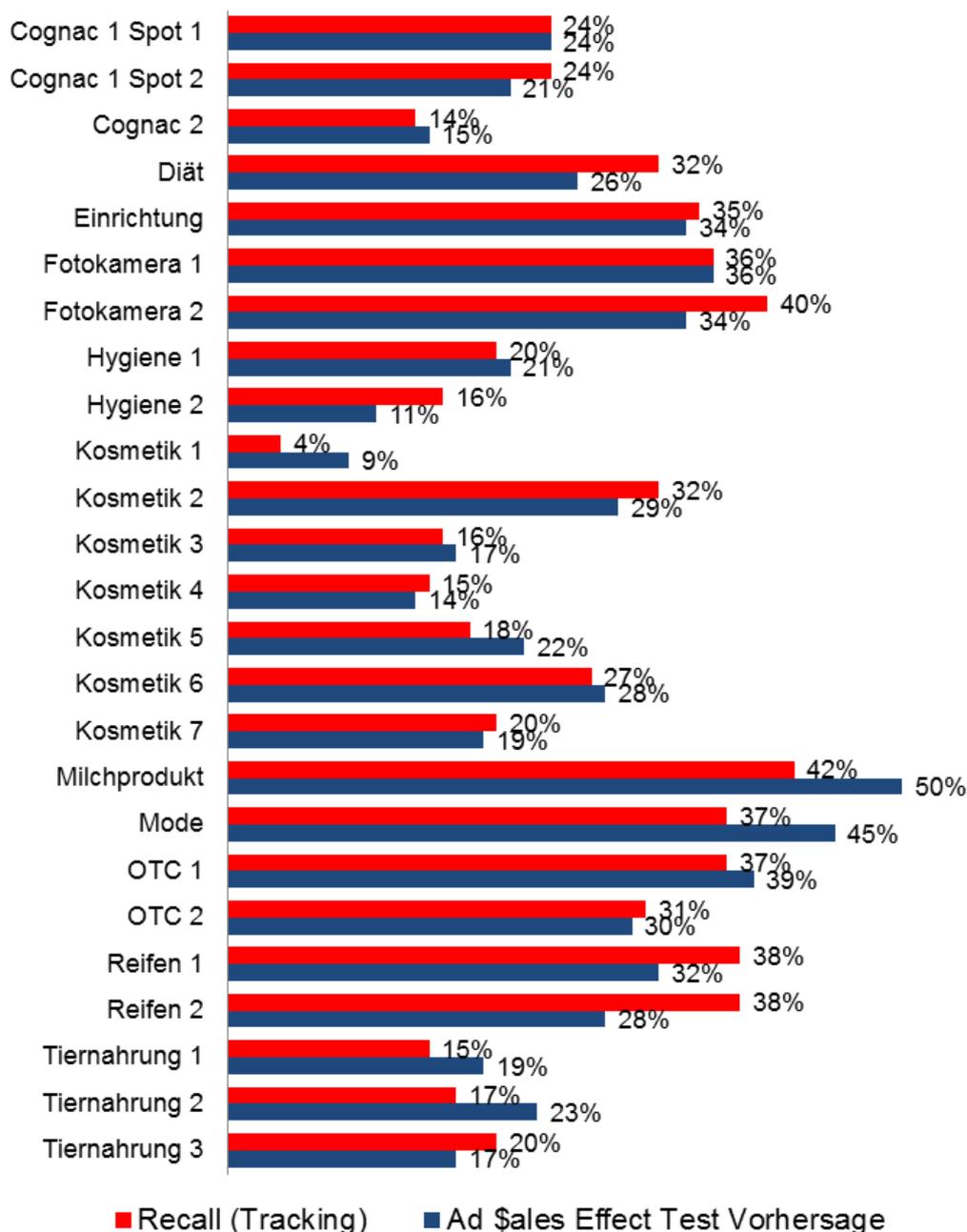


Abbildung 7: Recall: hohe Korrelation von Pre-/Post Test-Ergebnissen

(ursprünglich veröffentlicht: Mayer de Groot et al. 2001)

schnell wieder vergessen wird oder lange im Gedächtnis bleibt. Diese Art der Werbeerinnerungs-Vorhersage wird seit 1981 in der Praxis angewendet (Mayer de Groot et al. 2001). Die hohe Zuverlässigkeit der Recall-Prognose hat sich in mehreren hundert Vergleichen zu Werbeerinnerungswerten aus Trackingstudien bestätigt.

Abbildung 7 beweist anhand von 25 Beispielen exemplarisch, dass sich der Aufmerksamkeitswert in der Recall-Vorhersage in den unterschiedlichsten Produkt- und Dienstleistungskategorien bewährt hat. Dies gilt sowohl für die Prognose von kleinen als auch großen Recall-Werten. Auch bei sehr bekannten Marken und Produktkategorien mit hohem Produkt-Involvement (z.B. Tiernahrung) wird der Recall zuverlässig vorhergesagt. Im Rahmen der statistischen Toleranzen, zeigt sich immer eine Übereinstimmung.

Sicherheitshalber sei angemerkt, dass es in Abhängigkeit von der Höhe des eingesetzten Mediabudgets als auch der verfolgten Mediaschaltstrategie einige Zeit (meist 1 bis 3 Jahre) dauern kann, bis die Ergebnisse validiert werden können.

7. Fazit und Implikationen für Werbetreibende und Agenturen lesen Sie auf Seite 1

Ein ausführliches Literatur- und Quellenverzeichnis erhalten Sie gerne auf Wunsch: mdg@mayerdegroot.com



Dr. Ralf Mayer de Groot führt seit 2007 die MAYER DE GROOT Marketing-Research and Consult sowie ein Netzwerk an ausgewählten Kooperationspartnern. Er war zuvor 7 Jahre Vorstand / COO bei Konzept & Analyse und ist Autor zahlreicher Fachveröffentlichungen und Dozent an Universitäten. Strategische Markenführung und Marktforschung kennt er aus allen Perspektiven: aus über 15 Jahren Marketingforschung (MARS Gruppe, Target Group, Konzept & Analyse), aus 7 Jahren Agenturen (Scholz & Friends, BJP, HMS & Carat), aus 7 Jahren operativer Marketingverantwortung (MARS Gruppe international) und aus über 20 Jahren Marketingstrategie-Beratung.

Tel: 06198 50 29 21 mdg@mayerdegroot.com / www.mayerdegroot.com

Literatur

- Aaker, D. A./Bruzzone, D.E. (1981): Viewer Perception of Prime Time TV-Advertising, in: Journal of Advertising Research, Oct. S. 15-23
- Achenbaum, A. A. (1985a): Can We Tolerate a Double Standard in Marketing Research? in: Journal of Advertising Research, 25 (June-July): RC3-RC7
- Achenbaum, A. A. (1985b): Marketing productivity impeded by malpractices, American Marketing Association Breakfast Conference 07.02.1985
- Ariely, D. (2011): The upside of irrationality, London
- Ariely, D. (2008): Predictably irrational, New York
- Armstrong, J.S. (2001): Principles of forecasting, Boston
- Binet, L./ Field, P. (2013): The Long and the Short of It, hrsg.: IPA Institute of Practitioners in Advertising, London
- Bird, M.; Ehrenberg, A. (1966): Intention to buy and claimed brand usage, Operations Research Quarterly, 17 (1), S. 27-46
- Bosch, C./Schiel, S./ Winder, T. (2006): Emotionen im Marketing: Verstehen, Messen, Nutzen, Wiesbaden
- Broadbent, T. (2004): Why pre-testing is obsolete - Will pre-testing ever predict?, ADMAP issue 454, S. 150-154
- Burke Marketing Research, Inc.: Day after recall television commercial testing, 1975, 1976 & 1980
- Buzzell, R.D./Gale, B.T. (1987): The PIMS (Profit Impact of Market Strategy) Principles, New York

Clancy, K.J., Kweskin (1971): TV commercial recall correlates, *Journal of Advertising Research*, 2, S. 18-20

Damasio, A.R. (2003) *Looking for Spinoza*, London

Damasio, A.R. (1994) *Descartes' Error*. New York, NY

Dassel, C. v. (2003): Vorwort zu den Methoden in: Mayer de Groot, R. (2003): *Marken-Diversifikation und Tragfähigkeit*, Eppstein und Nürnberg, S. 4

Dassel, C. v./Wecker, K.M./Mayer de Groot, R. (2001a): Durch strategische Marktforschung zum Erfolg: NIVEA und NIVEA Soft, in: *planung & analyse* 2, S. 26-33

Dassel, C. v./Wecker, K.M./Mayer de Groot, R. (2001b): "Harte Fakten führen sanfte Creme zur Spitze", in: *Horizont*, 19/2001 S. 25

Dassel, C. v./Wecker, K.M./Mayer de Groot, R. (2001c): NIVEA SOFT – Sanft zur Haut und zur Dachmarke“, in: *Absatzwirtschaft*, 10 S. 60-63

Davidson, H. (1997): *Even more offensive marketing*, London

Deppe, M./Schwindt, W./Kugel, H./Plassmann, H./Kenning, P. (2005): Non-linear Responses within the Medial Prefrontal Cortex Reveal when Specific Implicit Information Influences Economic Decision-Making, in: *Journal of Neuroimaging*, 15, 2, S. 171–183.

Drucker, P. F. (1979): *Managing for results*, 8. Auflage, London

Esch, F. R.; Beyer, S. (2011): Auch Marketing sollte Werte schaffen, *FAZ* 14.06.2011, S. 14

Farr, A./Gardiner, S. (2001): Creative enough for the financial director?, *Admap* 36 S. 25-28

Farr, A. (1993): *Persuasion Shift Testing*. In: *Admap*, Jan.

Fournaise Marketing Group (2014): *FournaiseTrack – Media Releases*, London, 23.01. 2014

Gibson, L.D. (1983): If the answer is copy testing, the answer is ..."Not Recall", *Journal of Advertising Research*, (23) 1 S. 39-46

Gigerenzer, G. (2007): *Bauchentscheidungen*, München

Gill, P.; Mayer de Groot, R. (2011): Große Promotion, kleine Wirkung?, in: *planung und analyse* 2 / 2011, 48-51

Gill, P., Mayer de Groot (2008): WD-40: Mit kleinem Budget zum großen Erfolg, in: *absatzwirtschaft* 9 S. 44-46

Grosenick, L./Greer, S./Knutson, B (2008): Interpretable classifiers for fMRI improve prediction of purchases, *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol. 16, No.6, S. 539-548, Abruf am 28.02.2014

Habermeier, J.: *Umfeldqualitäten systematisch messen – Programmumfeld und Werbewirkung*, *Media Spectrum* 5/1997, S. 8-12

Haimerl, E., Mayer de Groot, R. (2000): Benchmarking in der Marktforschung. Von den Falschen das Falsche lernen, um das Falsche zu tun!, *planung & analyse*, 5 S. 16-20

Hardie, B.; Johnson, E.; Fader, P. (1993): Modelling Loss Aversion and Preference Dependence Effect on Brand Choice, *Marketing Science* 12 4 S. 378-394

Heath, R./Feldwick, P. (2007): 50 Years using the wrong model of TV advertising, University of Bath, School of Management, Working Paper Series 2007.03

Hollis, N./Twose, D. (2007): Refining Market Propositions, in: *ESOMAR Research Handbook*, 5. Aufl. S. 169-198, Chichester, UK

Hubert, M./Kenning, P. (2008): A current overview of consumer neuroscience *Journal of Consumer Behaviour* 7 S. 272–292

Huesing, B./Jäncke, L./Tag, B. (2006): *Impact Assessment of Neuroimaging*, Zürich.

Jones, J. P. (1995): When ads work, *The German Version*, 8

Kahneman, D.: *Thinking, fast and slow*, London 2011

Kahneman, D.; Tversky, A. (1979): Prospect Theory: An Analysis of Decision under Risk, *Econometrica*

Kast, B. (2007): *Wie der Bauch dem Kopf beim Denken hilft*, Frankfurt a.M. 2007

Kastenholz, J.; Young, C.; Kerr, G. (2004): Does day-after Recall Testing Produce Vanilla Advertising?, in: *Admap*, June, Issue 451, 34-36

Keller, W.; Silbermann, L.; Mayer de Groot, R (2013): *WGV-Versicherungen: Welche Werbebotschaft wirkt am besten?* Stuttgart, Eppstein 11 2013

Kenning, P. (2008): 5 Jahre neuroökonomische Markenforschung: Eine Zwischenbilanz, Vortrag auf dem 1. G-E-M Forum „Neuroökonomie und Markenführung“ am 18.11.2008, Universität Münster

Kenning P, Plassmann H, Ahlert D. (2007): Applications of functional magnetic resonance imaging for market research. *Qualitative Market Research: An International Journal* 10 (2):1352–2752.

Kenning, P./Plassmann, H./Deppe, M./Kugel, H./Schwindt, W.(2005): Wie eine starke Marke wirkt, in: *Harvard Business Manager*, März, S. 53–57.

Kenning P./Plassmann H./Deppe M./Kugel H./Schwindt W. (2002): The discovery of cortical relief. *Field of Research: Neuromarketing* 1: 1–26.

Kenning P./Plassmann H./Deppe M./Kugel H./Schwindt W. (2002): Die Entdeckung der kortikalen Entlastung. *Neuroökonomische Forschungsberichte. Teilgebiet Neuromarketing*, Nummer 1. Münster

Kerlinger, F. N. (1973): *Foundations of Behavioral Research*. New York

Kleij, T. (1996): *Entscheidungsjahr für Lefax*, Asche intern Nr. 69

Knutson, B./Scot, R./Wimmer, G.E./ Drazen, P./Loewenstein, G (2007): Neural Predictors of Purchases, *Neuron* 53, January 4, S. 147–156

Koepller, K. (1974): *Werbewirkungen definiert und gemessen*, Velbert

Koschnick, W. J. (2007): *Focus-Jahrbuch 2007, Schwerpunkt: Neuroökonomie, Neuromarketing und Neuromarktforschung*, München

Kotler, P (2005): Marketing is not working!, *marketing journal* 3 S.35-36

Kremin, S., Nemeth, E., Pape, R., Mayer de Groot, R. (2014): simpicol: Neue Farben, frischer Auftritt, *planung & analyse* 6, S. XX-YY

Lakeoff, G. (1987): *Women, fire and dangerous things: What categories reveal about the mind*, Chicago

Landman, J. (1987): Regret following action and inaction, affective response to positive versus negative outcome, *Personality and Social Psychology Bulletin*, 13, S. 524-536

Lebensmittel Zeitung Hrsg. (1985): *Tod im Regal – Eine Analyse der Lebensmittel Zeitung zur me-too Problematik*, Frankfurt

Lehrer, J. (2009): *How we decide*, New York 2009

Lenatz, A.; Mayer de Groot, R. (2007a): Hasseröder: Die neue Markenstrategie greift, Eppstein

Lenatz, A.; Mayer de Groot, R. (2007b): PerfectDraft-Zapfsystem von Philips und Inbev, Eppstein

Lord, K.R., Burnkrant, R.E. (1988): Television program elaboration effects on commercial processing, *Advances in Consumer Research*, 15 S. 213-218

Lübbe, R.; Kappelt, J.; Mayer de Groot, R (2005): Präzises Innovations-Marketing, *planung & analyse* 6 S. 43-48

Lübbe, R.; Mayer de Groot, R; Hoffmann, S. (2004): Werbung mit voller Absatz-Wirkung, *planung & analyse* 5, S. 69-74

Mayer de Groot, R. (2014a): Standardisierte Tests als zentrale Flop-Ursache, *planung & analyse*, 3, S. 56-57

Mayer de Groot, R. (2014b): Zuverlässige Marketing-ROI Prognose statt Russisch-Roulette, Eppstein

Mayer de Groot, R. (2014c): Wie groß sind Ihre Erfolgchancen, wenn Sie über 95% der Kaufmotive nicht präzise kennen?, Eppstein

Mayer de Groot, R. (2014d): Konsumgüteränderungen: Top statt Flop. Wie Details im Unterbewusstsein wirken, *planung & analyse Sonderheft LZ (10) 2014*, S. 37-39

Mayer de Groot, R. (2014e): Kaufentscheidungen vorhersehen, *absatzwirtschaft* 11, S. 36-38

Mayer de Groot, R. (2013): Using New Product Chances to Full Advantage, *p&a international market research* 1 (March), S. 22-25

Mayer de Groot, R. (2012a): Emotionale Kauf-Impulse aus dem Unterbewusstsein: Details entscheiden über Erfolg oder Flop! Eppstein

Mayer de Groot, R. (2012b): Best Practice in Communication around the World, extended version, originally published in: *planung und analyse international* 1

Mayer de Groot, R. (2012c): Effektive internationale Kommunikation – Zehn Best Practice Fallstudien, Eppstein

Mayer de Groot, R. (2012d): Fiese Benchmark-Fälle: Flops über Flops! Eppstein

Mayer de Groot, R. (2011a): Werbe-Abverkaufswirkung pretesten! Auf jeden Fall!, aktualisiert und erweitert 2011, ursprünglich veröffentlicht in: *planung & analyse* 3 S. 38-42

Mayer de Groot, R (2011b): Erfolgreich positionieren funktioniert anders, weil das Gehirn anders arbeitet, Eppstein

Mayer de Groot, R. (2011c): Erfolgreich positionieren, *Markenartikel* 1 / 2 S. 54-56 und 3 S. 60-61

Mayer de Groot, R. (2011d): YES gegen Milka Tender: Kosteneffiziente Abwehr der Konkurrenz, Eppstein

Mayer de Groot, R. (2010a): Welche Werbung wirkt, in: *MARKANT HandelsMagazin*, 10, S. 20 -21

Mayer de Groot, R. (2010b): Tops statt Flops, Eppstein 2010

Mayer de Groot, R. (2009): Die Macht des Unterbewusstseins, *Markenartikel* 6 S. 38-41

Mayer de Groot, R. (2008): Erfolgreiche Positionierung: Warum Marktsegmentationen oft falsch sind, Eppstein

Mayer de Groot, R. (2003): Marken-Diversifikation und Tragfähigkeit, Eppstein, Nürnberg

Mayer de Groot, R. (2002): Werbewirkungstests, in : Pflaum, D. u. a.: Lexikon der Werbung, München, S. 588-592

Mayer (de Groot), R. (1992): Wie man die Effizienz des Mediabudgets durch eigene Mediaforschung um 30% oder mehr steigern kann, *planung und analyse* 4/1992 S. 62-68

Mayer de Groot, R./ Fritzen, T. (2008): Aus weniger ein Mehr herausholen, *Markenartikel* 1 / 2 S. 30-33, dies: Mehr Umsatz aus Media herausholen, erweiterter Sonderdruck

Mayer de Groot, R./ Pallek, T./Haimerl, E. (2001): Lug und Betrug beim Recall?, *planung & analyse*, 3 S. 45-51; auch veröffentlicht in: *marketingjournal*, 1/2002, S. 56-61

Mayer de Groot, R./Plüm, G. (2012): Sichere Prognose, *Research & Results* 6 S. 38-40

Mayer de Groot, R./Salomon, J. (2007): Innovativ, traditionell oder lieber schräg? Wie unser Unterbewusstsein die Werbewirkung beeinflusst, *Markenartikel* 9 S. 117-120

Mayer de Groot, R./Haimerl, E./Zerzer, M. (2004): Von Werbe-Flop auf Marktführer in 3 Monaten *planung & analyse*, 1 S. 61-65

Millward Brown International Plc (1991): How advertising affects the sales of packaged goods brands

Morwitz, V.G. (2001): Methods for forecasting from intention data, in: Armstrong, J.S. (ed): *Principles of forecasting*, Boston S. 33-56

Murphy, J.H., Cunningham, I.C.M., Wilcox, G.B. (1979): The impact of program environment on recall of humorous television commercials, *Journal of Advertising*, 19 (8) S. 17-21

Park, C.W./McClung, G.W. (1986): The effects of TV-program involvement on involvement with commercials, *Advances in Consumer Research*, 13, S. 544-548

Peckham, J.O. Sr. (1983): *The marketing Wheel*, Scardale, USA

Percy, L./Rossister, J.R. (1992): Measuring advertising effectiveness: copy testing (pre-testing) ads and tracking (post-testing) the campaign, working paper, no. 92-020, Kensington, Univ. of New South Wales, Australia

Pepels, W. (1994): *Kommunikationsmanagement: Marketing-Kommunikation vom Briefing bis zur Realisation*, Stuttgart.

Reese, W./Fritzen, T./Mayer de Groot, R. (2004): Media-Effizienz: Werbebudget runter – Absatz rauf, *Absatzwirtschaft*, 7 S. 92-96

Rehorn, J.: *Was leisten Pretests?*, Düsseldorf 1983

Schäffken, J. (2008): Iglo – eine Marke zurück auf Wachstumskurs, Vortrag beim 1. Kongress für mehr Markenerfolg, Ludwigsburg 14.11.2008

Sacharin, V./Schlegel, K./Scherer, K. R. (2012): Geneva Emotion Wheel rating study (Report). Genf: University of Geneva, Swiss Center for Affective Sciences, August.

Samuelson, W.; Zeckenhauer, R. (1988): Status Quo Bias in Decision Making, *Journal of Risk and Uncertainty* 11, S. 1 ff.

Schank, R. (1990): *Tell me your story: A new look at real and artificial memory*, New York

Shapiro, S./MacInnis, D.J./Heckler, S.E. (1997) The effects of incidental ad exposure on the formation of consideration sets. *Journal of Consumer Research* Vol. 24 S. 94-104

Shaw, D. J./ Mayer de Groot, R./Scharf, T. (2004a): Goldrichtig. Fallbeispiel: Mit Präzision stößt Beck´s auf Gold, in: *Markenartikel* 3 S. 38-44

Shaw, D. J./Schipke, A./Mayer de Groot, R. (2004b): Beck´s Gold segelt auf Erfolgskurs, in: *planung & analyse* 2 S. 20-26

Silk, A.J; Urban, G.L. (1978): Pre-Test Market Evaluation of New Products goods: A Model and Measurement methodology, *Journal of marketing research*, 15 S. 171-191

Soldow, G., Principe, V. (1981): Response of commercials as a function of program content, *Journal of Advertising Research*, 2 S. 59 ff.

Strait, H./Arndt, T./Mayer de Groot, R. (2007): Niederegger: Erfolgreicher Imagetransfer auf Tafelschokolade – Vermeidung negativer Rücktransfers, Eppstein

Thaler, R. H., Sunstein, C. R. (2009): *Nudge*, New York

Thaler, R. H. (1980): Toward a positive theory of consumer choice, *Journal of Economic Behavior & Organization*, 1, March , S. 39-60

Urban, G.L./Katz, G.M. (1983): Pretest market models: Validation and Managerial Implications, *Journal of marketing research*, 20 S. 221-234

Vakratsas, D./Ambler, T. (1999) 'How Advertising Works: What Do We Really Know?', *Journal of Marketing* 63(1): 26–42

Weber, T./Mayer de Groot, R./Fritzen, T.: Vorwerk-Teppiche fliegen zu neuen Höhen, in: *marketingjournal* 7/2006, S. 30-33

Weiser, J. (1993): Phototherapy techniques: Exploring the secrets of personality snapshots and family albums, San Francisco

Wells, W.D./Leavitt, C./McConville, M: A Reaction Profile for TV-Commercials, in: Journal of Advertising Research, Dec.1971, S.11-17

Wells, W.D. (1964a): EQ, son of EQ and reaction profile, Journal of Marketing, Okt. S. 45-52

Wells, W. D.(1964b): Recognition, Recall, and Rating Scales, in: Journal of Advertising Research, Vol. 4 (3), 2-8

Zaltman, G./ Zaltman L.H. (2008): Marketing metaphoria, what seven deep metaphors reveal about the mind of consumers, Boston

Zaltman, G. (2003): How customers think: Essential insights into the mind of the market, Boston

Zaltman, G./ Coulter, R. (1995): Seeing the voice of the consumer: Metaphor-based advertising research, Journal of Advertising Research, July/August, S. 35-51

Zinkhan, G. M./Fornell, C. (1985): A Test of Two Consumer Response Scales in Advertising, in: Journal of Marketing Research, 22 (4), 447-452.

Zyman, S. (2002): The End of Advertising as we know it, Hoboken USA