

## **Benachteiligt "Day-after-recall" emotionale Werbung?**

Die Erinnerung an Anzeigen und Fernsehspots 1 Tag nach deren Vorführung wird seit rd. 100 Jahren als ein Indikator von „Werbewirkung“ betrachtet. Beispielhaft seien die Ausführungen in der Methodensammlung von Heinz Alpers aufgeführt (19..). Es wird an der Methode auch Kritik geübt (s. z.B. Arnold & Bird, 1982), aber sie erfreut sich großer und auch heute wohl noch immer einer gewissen Popularität, vermutlich, weil sie technisch so einfach und plausibel erscheint.

Der amerikanische Werbeforscher Hubert Zielske hatte sich vor vielen Jahren mit einer experimentellen Untersuchung einen Namen gemacht, in der er einen grundlegenden Beitrag zur Frage der optimalen Werbefrequenz lieferte ("How often should you advertise?" Zielske 1959). Seiner Neigung zur Grundlagenforschung entsprach eine wiederum experimentell angelegte Untersuchung, mit der er der seinerzeit diskutierten Frage nachging, ob emotionale ("feeling") Werbemittel gegenüber rationalen ("thinking") Werbemitteln von dieser "day-after-recall" - Methode benachteiligt werden. Die Erörterung dieser Frage muss uns hier nicht beschäftigen; man lese hierzu den Originalbericht oder dessen späteren Abdruck (Zielske 1982/94).

Zielskes Bericht hat eine ausserordentlich starke und dauerhafte Aufmerksamkeit gewonnen. William Greene hat die Befunde vor Zielskes Veröffentlichung in einem ARF 'workshop paper' (1981) angesprochen. Im Journal of Advertising Research habe ich 6 Beiträge gefunden, in denen er zitiert wird: Singh & Churchill (1987), Stayman et al. (1989), Stewart et al. (1990), Fenwick & Rice (1991), du Plessis (1994), Laskey et al. (1994). Last but not least drei Bezüge in einem Buch: in Aufsätzen von Coulson, Pechman & Stewart, Thorson & Friestad in Cafferata & Tybout (ed. 1989). In den meisten Fällen wurde sein Befund akzeptiert, dass die "day-after-recall" - Methode emotionale Werbemittel benachteiligt; es gab auch Zweifler, die sich jedoch mit seinem Test nicht vom Standpunkt der Methodologie befassten.

Überdies wurde Zielskes Bericht in die Sammlung wichtiger Beiträge zum Testen von Werbemitteln der Advertising Research Foundation aufgenommen: "Understanding Copy Pretesting" (Dunn et al., 1994).

### **Anlass zur näheren Betrachtung**

Den Anstoß, sich den Zielske-Bericht erneut und genauer anzusehen, gab eine Anmerkung des amerikanischen Kollegen Joel Dubow zu meinem "Recognition"-Aufsatz (1995): Ich hätte dort anscheinend "Recall" als eine reine Erinnerungsmessung behandelt. Die Sache sei aber nicht so simpel: Es handle sich um Erinnerung, aber dazu käme die Fähigkeit, das Erinnernte so zu verbalisieren, dass sich die Auswerter damit zufrieden geben. (Dubow 1995)

"Recall" spielt in meiner angesprochenen Reanalyse der „Recognition“-Methode nur eine sekundäre Rolle, wie man in diesem Band nachlesen kann. Deshalb habe ich mich nicht sehr eingehend mit der Analyse der Recall-Daten und der Frage beschäftigt, was für eine Art von Messung 'Recall' sein mag.

Dubows Anmerkung kam zu einem Zeitpunkt, wo 'Recall' ziemlich heftig von ihm (1995), Gibson (1995) und Ross (1995) diskutiert wurde; auch duPlessis (1994) bezog sich auf die Methode. Ferner fand man in der großen "Validity Study" der ARF heraus, dass alle „Recall“-Messungen nicht besonders gut abschnitten (Haley und Baldinger 1991).

Die berechtigte Ergänzung des Kollegen Dubow erinnerte mich zuerst an den berühmten Aufsatz von Herbert Krugman "Memory without Recall . . ." (1977), den ich gut kannte. Sodann sah ich den Bericht von Zielske erneut durch.

### Zielskes Ansatz

Für seine Untersuchung benutzte Zielske einen Satz von 6 TV-Spots und einen Satz von 6 Anzeigen als Grundlage. Jeder Satz bestand aus 3 Werbemitteln, die als „gut gestaltete rationale“ ('thinking') und 3 „emotionale“ ('feeling') betrachtet wurden. Ein Panel von Werbefachleuten hatten diese aus 54 TV-Spots und 22 Anzeigen ausgewählt.

Das Konzept seiner Analyse beruhte nach Aussage von Zielske auf dem angeblichen Befund: „1972 habe Krugman berichtet, dass Recall die wahre Erinnerung der Werbung unterschätzt.“ Als Konsequenz daraus bezeichnete Zielske die generell niedrigeren "Recall"- Werte im Vergleich zu den ebenfalls ermittelten "Recognition"- Ergebnissen als „Unterschätzung“ (gemeint: der „wahren Erinnerung“).

Er verwendete zwei Messwerte: "Proven Recognition" und "Day-after-recall" (Beschreibungen s. Alpers 19xx ) Deren Meriten sollen uns hier nicht beschäftigen, sondern die Art, wie ihre Ergebnisse analysiert wurden.

### Zielskes Ergebnisse & Berechnungen

In der Anlage sind die Originalwerte mit Veränderungen und Zusätzen wiedergegeben. Zielske hat die "day-after-recall" -Werte von den "proven-recognition"- Werten subtrahiert. Die Resultate hat er sodann als relative Prozente auf der Basis der „recognition“- Werte ausgewiesen, und diese Ergebnisse als „Unterschätzung“ bezeichnet.

Seinen Begriff „Unterschätzung“ durch den neutralen Begriff „Differenz“ habe ich ersetzt; ferner habe ich nicht nur seine „relative Differenz“ ausgewiesen, sondern zusätzlich die absoluten Differenzen als extra Kolonne.

Bei den TV-Spots untermauern die Durchschnitte die Hypothese, dass „emotionale“ Werbung durch die "day-after-recall" - Methode bestraft wird:

	<u>„rational“</u>		<u>„emotional“</u>
absolute Differenz:	6 %	vs.	13 %
relative Differenz:	16 %	vs.	41 %

Bei den Zeitschriften-Anzeigen sind die Unterschiede geringfügig:

absolute Differenz:	25 % vs.	26 %,
relative Differenz:	64 % vs.	60 %.

### Ungeeignetes Analyseverfahren

Man findet in vielen Analysen Berechnungsweisen wie die hier von Zielske verwendete: Es werden Prozente gerechnet, dazu Prozente von Prozenten, manchmal als „Index“ dargeboten, ferner werden Durchschnitte präsentiert.

Das Problem solcher Berechnungen wird deutlich, wenn man eine absolute Differenz von 10 % in extremen Bereichen in relative Differenzen umwandelt:

<b>20% vs. 30%</b> = „absolute“ Differenz: + 10%	„relative“ Differenz: + 50,0 %
<b>80% vs. 90%</b> =	+ 10% + 12,5 %.

Es ist offensichtlich dass die „relativen“ Differenzen von 50% und 12,5% in keiner vernünftigen Weise miteinander in Beziehung gesetzt und interpretiert werden können. Wer nicht glaubt, dass dergleichen Unsinn auch noch von einer Fachzeitschrift veröffentlicht wird, sei auf einen bemerkenswerten Fall hingewiesen, wo die so gewonnenen relativen Differenzen dann noch prozentual in Relation zu den am Anfang der Berechnungen stehenden „Recognition“-Werten gesetzt wurden. Wen wundert es, dass eine Punktwolke nach Art der Milchstraße herauskam, mit einer Korrelation von  $r^2 = 0,01$ . (Walker & von Gonten 1989). Und das wurde publiziert in der Fachzeitschrift „Journal of Advertising Research“ (3/1989).

### Verarbeitung per Regressionsanalyse

Die Idee, die Ergebnisse von „Recognition“- zu „Recall“- Messungen in Beziehung zu setzen, ist richtig, aber nicht neu. Wie in meiner Reanalyse von „Recognition“ dargelegt, hatten Sherwood Dodge und Jerry Greene dergleichen bereits 1956 per Regressionsanalyse praktiziert. Jedoch litt Dodges Bericht (1956) offensichtlich am „file & forget“ - Syndrom.

Wir gehen von Starchs Vorstellung aus, dass eine „Recognition“- Methode Daten zur Beachtung liefert (Starch 19.. und ..). Es liegt auf der Hand, dass die Leser einer Publikation eine Anzeige beachtet haben müssen, und dass Zuschauer einem TV-Spot irgendeine Art von Aufmerksamkeit geschenkt haben müssen (einschl. Beachtung von Marke/Hersteller), um irgendetwas von Relevanz erinnern zu können.

Ferner ist davon auszugehen, dass „Erinnerungswerte“ maximal so hoch sein dürften wie „Beachtungs-/Aufmerksamkeitswerte“, in der Regel jedoch (deutlich) niedriger; alle einschlägigen Berichte bestätigen diese Annahme.

Da Theorie und Praxis ferner aussagen, dass die Beachtung/Aufmerksamkeit der getesteten Anzeigen bzw. TV-Spots (sehr) unterschiedlich ausfällt, ist logischerweise zu erwarten, dass diese Unterschiede die Erinnerungs-/"Recall"-Werte stark beeinflussen. Wenn man nicht weiß oder nicht berücksichtigt, wie gut oder schlecht die Anzeigen / TV-Spots beachtet wurden, dann muss die „Recall“- Analyse zu falschen Ergebnissen führen.

Also ist es notwendig, die Beziehung zwischen der „Recall“- und den „Recognition“-Ergebnissen per Regressionsanalyse festzustellen. Es ist zu erwarten, dass diese Beziehung nichtlinear ist, wie wir in fast allen der vielen hundert von uns durchgeführten Untersuchungen dieser Art gefunden haben.

### Sehr hohe Korrelationen

Bei den TV-Spots liefert eine Kurve die beste Anpassung an die Daten, mit einer sehr hohen erklärten Varianz:  $y = a + bx + cx^2$ . Die Korrelation ist hoch:  $r^2 = 90,1\%$ .

Im Fall der Zeitschriften-Anzeigen liefern drei Formeln praktisch gleichwertige Anpassungen:  $r^2 = 94,1\%$ . Deshalb habe ich die einfachste gewählt:  $y = a + bx$ .

### Unterschiedliche Abweichungen

Auf der Grundlage von so kleinen Stichproben und so hohen Korrelationen können keine hoch signifikanten Abweichungen erwartet werden. Um wenigstens Indikationen für die Sachlage zu gewinnen, habe ich mich für ein Signifikanz-Niveau von 80% entschieden; eine Wahrscheinlichkeit von 4:1 erscheint akzeptabel. Die Ergebnisse sind in den Graphiken „TV“ und „A“ dargestellt. Signifikante Abweichungen von den „Erwartungswerten“ (dargestellt durch die Kurve/Gerade) sind durch Sterne gekennzeichnet.

Bei den TV-Spots befindet sich

- der „rationale“ Spot „E“ oberhalb
- der „emotionale“ Spot „F“ unterhalb der Kurve.

Beide Abweichungen betragen 5%. Das Ergebnis bestätigt Zielskes Hypothese.

Bei den Zeitschriften-Anzeigen sind ebenfalls zwei signifikante Abweichungen zu erkennen, allerdings in entgegengesetzter Richtung:

- die „rationale“ Anzeige „E“ ist unterhalb
- die „emotionale“ Anzeige „F“ oberhalb der Kurve.

Dies stimmt mit Zielskes Hypothese nicht überein.

## Diskussion

Beim Zielske'schen Ansatz beruht bereits die Auflistung der TV-Spots und der Zeitschriften-Anzeigen auf der vorgefassten Meinung, dass die Werbemittel in die zwei Kategorien „emotional“ und „rational“ gehören.

Damit entgeht dem Analytiker als grundlegende Information, dass bereits auf der Messebene der Beachtung/ Aufmerksamkeit („recognition“) sehr große Unterschiede zwischen den einzelnen Werbemitteln auftreten: Bei den TV-Spots erstrecken sich die Beachtungs-Werte von 23% bis zu 56 %, bei den Anzeigen von 27% bis zu 67%.

Zweitens ist erkennbar, dass die Eigenschaften „emotional“ und „rational“ die unterschiedliche Höhe der „Recognition“- und „Recall“-Werte nicht erklären:

- Bei den TV-Spots führt ein „rationaler“ mit großem Abstand sowohl bei „Recognition“ wie bei „Recall“, gefolgt von zwei „emotionalen“.
- Bei den Zeitschriften-Anzeigen führt eine „emotionale“ mit großem Abstand sowohl bei „Recognition“ wie bei „Recall“, gefolgt von zwei „rationalen“.

Die zweite grundlegende Tatsache wird bei der Zielske'schen Art der Auflistung ebenfalls nicht erkennbar: Dass die Reihenfolgen der Werbemittel bei „Recognition“ und bei „Recall“ mit nur einer Ausnahme gleich sind. Das wird von der extremen Höhe der Korrelationen bestätigt.

Also ist die Schlussfolgerung zu ziehen:

- Die unterschiedlich hohen „Recall“- Werte werden primär von den ebenso unterschiedlichen „Recognition“-Werten bestimmt. Also sind die von Zielske präsentierten Ergebnisse und die darauf basierenden Schlussfolgerungen nicht zutreffend.

Damit wird die Aufmerksamkeit auf eine weitere, sehr wichtige Frage gelenkt: Was erklärt die so stark unterschiedlichen „Recognition“-Werte?

Eins ist klar: Es gibt viele Faktoren, die zu unterschiedlichen Beachtungs- bzw. Aufmerksamkeitswerten führen – neben den ziemlich vagen Eigenschaften „emotional“ und „rational“. Leider wird im Zielske'schen Bericht nichts über die getesteten Werbemittel, die beworbenen Marken und die Strukturen der Testteilnehmerinnen mitgeteilt.

Bereits 1960 hatte Lucas in seiner Analyse der großen PARM-Untersuchung der Advertising Research Foundation gezeigt, dass das Personen-Merkmal „gehört zu den prospektiven Käufern/ gehört nicht dazu“ einen großen Einfluss auf die Werte für „Recognition“ und „Recall“ hat. Es ist nichts darüber bekannt, wie groß die Zielgruppen für die in den verschiedenen Werbemitteln beworbenen Marken sind. Sind die Zielgruppen unterschiedlich groß, bewirkt alleine deren Umfang unterschiedlich hohe Beachtungs- und Erinnerungswerte.

Dodge hatte in seiner Analyse (1956) der zitierten ARF-Untersuchung Abweichungen der „Recall“-Werte von den Erwartungswerten gefunden, die z.T. viel grösser sind als jene, die wir hier finden. Und er hat eine Reihe von Merkmalen für die beworbenen Marken und Anzeigen nennen können, die für relativ gute oder schlechte Ergebnisse verantwortlich sind.

Aufgrund dieser, anderer sowie eigener Erfahrungen würde man ein halbes Dutzend Einflussfaktoren erwarten, die sich auf die Werbebotschaft beziehen, ebenso viele, die sich auf die Werbemittelgestaltung beziehen, sowie weitere auf Produkte und Marken bezogene Merkmale.

Nur wenn man dergleichen wesentliche Faktoren neben der „emotionalen“ oder „rationalen“ Gestaltung der Werbemittel in eine kontrollierte Untersuchung einbezieht, kann geprüft werden, ob die Merkmale „emotional“ und „rational“ tatsächlich eine ins Gewicht fallende Bedeutung haben.

### **Krugman zu korrigieren ?**

Wenn man Zielskes Analyse und damit seinen angeblichen Befund infrage stellt, was geschieht dann mit der von Zielske zitierten angeblichen Aussage von Herbert Krugman ?

Eine nähere Betrachtung dessen, was Krugman 1972 geschrieben und 1977 erweitert hat, erweist, dass seine Vorstellungen und Befunde die Hypothese betr. „emotionale“ Gestaltung nicht notwendigerweise unterstützt. Krugman betrachtete „Recognition“ und „Recall“ als zwei Messwerte, die zwei unterschiedliche Aspekte der Erinnerung betreffen. Und dass „Recognition“ nahe an dem ursprünglichen „zu Gesicht bekommen“ (exposure) liegt, wohingegen „Recall“ dies unterschätzt. Man findet in diesen Aufsätzen keine Aussage des Inhalts, dass unterschiedlich geartete Werbemittel durch die „Recall“- Messung unterbewertet werden.

Die am Anfang zitierte Meinung von Dubow bezieht sich auf etwas anderes:

„Was Image-Werbung betrifft, so haben wir festgestellt, dass die Wiedergabe von Werbeaussagen (Anm.: durch die Befragten) praktisch nutzlos ist – sie liefert das Offensichtliche, wohingegen das Subtile unterschätzt wird. Es könnte sein, dass diese Unterschätzung des Subtilen der Grund dafür ist, dass ein so großer Teil der Image-Werbung bei der Werbeaussagen-Wiedergabe unterhalb der Normen liegt – die Konsumenten können Subtiles nur schwer in Worte fassen.“ (Dubow 1986)

### **Ausblick**

Die Debatte betr. „Recall“ wird wohl ewig und im Kreis weiter gehen, sofern wir nicht einen echt wissenschaftlichen Ansatz für die Forschung nutzen:

- 1.)** „Recall“ sollte nicht isoliert untersucht werden, sondern nur im Kontext mit den anderen Messwerten, die relevant für die Erforschung des Werbewirkungsprozesses sind:

- Aufmerksamkeit / Beachtung
  - Beurteilung der Werbemittel
  - Probier- und Kaufneigung.
- 2.) Es ist zu unterscheiden zwischen „Recall“ des Werbemittels und „Recall“, der sich auf die wesentlichen Werbeaussagen bezieht.
  - 3.) Untersuchung der Beziehungen zwischen reinem, spontanem „Recall“ und unterschiedlichen Niveaus des „Aided Recall“, im Zusammenhang mit den anderen Messwerten der Werbewirkungen.
  - 4.) Anwendung der angemessenen analytischen Verfahren.
  - 5.) Gemäß der Erkenntnis „Nichts ist so praktisch wie eine gute Theorie“, ist eine (vorläufige) Theorie zu formulieren: **Was für einen Messwert hinsichtlich Werbewirkung stellt „Recall“ dar?** Die Theorie ist dann aufgrund entsprechender Tests zu korrigieren oder zu erweitern.

### Anlage: Tabellen

Table a (nr.3, original data plus abs. diff.)

Proven recognition vs. Day-after recall

<u>TV commercials</u>	<u>Recogn.</u> %	<u>Difference</u>		<u>rel.</u> %
		<u>Recall</u> %	<u>abs.</u> %	
A	56	49	7	13
D	32	24	8	25
E	24	21	3	13
Average	37	31	6	16

"Feeling"

TV commercials

B	37	21	16	43
C	36	25	11	31
F	23	10	13	57
Average	32	19	13	41

Table b (nr.6, original data plus abs. diff.)

Proven-recognition versus Day-after recall

<u>"Thinking"</u> <u>magazine ads</u>	Difference			
	<u>Recogn.</u> %	<u>Recall</u> %	<u>abs.</u> %	<u>rel.</u> %
V	48	18	30	63
W	42	16	26	62
Y	28	9	19	64
Average	39	14	25	64

"Feeling"  
magazine ads

U	67	32	35	51
X	36	16	20	56
Z	27	4	23	85
Average	43	17	26	60

Table c (regression analysis)

Proven recognition vs Day-after recall

<u>TV commercials</u>	<u>Recogn.</u> %	<u>Recall</u> %	<u>"e"</u> %	<u>Diff.</u> %
A (T)	56	49	49	0
B (F)	37	21	25	- 4
C (F)	36	25	24	+ 1
D (T)	32	24	21	+ 3
E (T)	24	21	16	+ 5*
F (F)	23	10	15	- 5*

F = "Feeling", T = "Thinking" Commercials

"e" = expected values

\* = difference from expected value significant on 80% level



Table d (regression analysis)

Proven-recognition versus Day-after recall

<u>Magazine ads</u>	<u>Recogn.</u>	<u>Recall</u>	<u>"e"</u>	<u>Diff.</u>
	%	%	%	%
U (F)	67	32	32	0
V (T)	48	18	20	- 2
W (T)	42	16	16	0
X (F)	36	16	13	+ 3*
Y (T)	28	9	8	+ 1
Z (F)	27	4	7	- 3*

F = "Feeling", T = "Thinking" Advertisements

"e" = expected values

\* = difference from expected value significant on 80% level

\* \* \*