Raoul V. Kübler

Essays on Corporate Communication

Empirical Applications to Product Recall Communication and Advertising Creativity

Betriebswirtschaftliche Aspekte lose gekoppelter Systeme und Electronic Business

Herausgegeben von
Prof. Dr. Dr. h.c. Sönke Albers,
Prof. Dr. Birgit Friedl,
Prof. Dr. Achim Walter,
Prof. Dr. Dr. h.c. Joachim Wolf,
Institut für Betriebswirtschaftslehre,
Christian-Albrechts-Universität zu Kiel

Prof. Dr. Dr. Christian Henning
Institut für Agrarökonomie,
Christian-Albrechts-Universität zu Kiel

Prof. Dr. Udo Konradt,
Institut für Psychologie,
Christian-Albrechts-Universität zu Kiel

In der Schriftenreihe werden Ergebnisse von Forschungsarbeiten veröffentlicht, die sich in herausragender Weise mit Fragen des Managements lose gekoppelter Systeme, virtueller Unternehmen und elektronischer Geschäftsprozesse beschäftigen. Die Reihe richtet sich an Leser in Wissenschaft und Praxis, die Anregungen für die eigene Arbeit und Problemlösungen suchen. Sie ist nicht auf Veröffentlichungen aus den Instituten der Herausgeber beschränkt.

Essays on

Corporate Communication

Empirical Applications to Product Recall Communication and Advertising Creativity

---

Inaugural-Dissertation zur Erlangung des Grades eines
Doktors der Wirtschafts- und Sozialwissenschaften der
Wirtschafts- und Sozialwissenschaftlichen Fakultät
der Christian-Albrechts-Universität zu Kiel

vorgelegt von
Dipl.-Kfm. Raoul V. Kübler
aus Kiel

Kiel, 2012

Gedruckt mit Genehmigung der
Wirtschafts- und Sozialwissenschaftlichen Fakultät
der Christian-Albrechts-Universität zu Kiel

Dekan: Prof. Dr. Roman Liesenfeld
Erstberichterstattender: Prof. Dr. Dr. h.c. Sönke Albers
Zweitberichterstattender: Prof. Dr. Dr. h.c. Joachim Wolf

Tag der Abgabe der Arbeit: 29.03.2012
Tag der mündlichen Prüfung: 11.05.2012

Vorwort

Die vorliegende Arbeit ist teils im Rahmen des DFG-Graduiertenkollegs „Betriebswirtschaftliche Aspekte lose gekoppelter Systeme und Electronic Business“ und teils im Rahmen des DFG-geförderten Projektes „Optimierung von Rückruf-Verfahren im Gebrauchsgütersegment“ entstanden. Im Fokus der Arbeit steht die empirische Analyse über den Einsatz und die Wirkung verschiedener Formen unternehmerischer Kommunikation mit einem Schwerpunkt auf die Kommunikation von Produkt-Rückrufen.

Oft erinnerte mich mein Dissertationsprojekt an die Segelei. Neben ruhigen Tagen mit Kaiserwetter, Sonnenschein und ordentlich Wind, der einen gut voran kommen lässt, fand sich auch der ein oder andere Tag mit dicken Wolken und Regen. Stürmische Böen sorgten dafür, dass das ganze Projekt spannender, rasanter und abwechslungsreicher wurde. Gleichzeitig aber auch anstrengender und herausfordernder. Ich bin froh, dass ich trotz all dieser Herausforderungen am Ende noch sagen kann, dass mir meine Dissertation immer Freude bereitet hat und ich glücklich auf vier Jahre „Up-and-Down“-Segelei zurückschaue.

Analog zum Segelsport lebt eine Promotion nicht alleine von ihrem Skipper, sondern auch von der Crew, die einen tatkräftig unterstützt und hilft. Bei dieser möchte ich mich im Folgenden herzlich bedanken:

Mein erster Dank gebührt dabei meinem Doktorvater Prof. Dr. Dr. h.c. Sönke Albers, der nicht nur die Idee zu diesem Projekt lieferte, sondern auch bei der Beschaffung von Drittmitteln und der Umsetzung meines Forschungsvorhabens mir immer tatkräftig zur Seite stand. Ohne seine exzellente fachliche Erfahrung und seine methodische Expertise wäre dieses Projekt nicht umsetzbar gewesen. Weiter danken möchte ich Herrn Albers auch für die internationale Ausrichtung meiner Promotion und für die vielen Möglichkeiten, mit anderen Forschern weltweit in Kontakt zu treten und sich auszutauschen. Sicherlich haben diese Internationalisierung und die vielen

dabei entstandenen Kontakte ihren Teil dazu beigetragen, dass ich mich am Ende für eine akademische Laufbahn entschieden habe.

Ebenso möchte ich mich bei meinem Zweitgutachter, Prof. Dr. Dr. h.c. Joachim Wolf für die schnelle Erstellung des Gutachtens, sowie bei Prof. Dr. Achim Walter für die Übernahme des Prüfungsvorsitzes bedanken.

Weiterer Dank gebührt Prof. Dr. Koen Pauwels, der mich für ein halbes Jahr an die Ozyegin University Istanbul einlud und mich dort hervorragend betreute. Die dort gemachten Erfahrungen mit dem internationalen Lehr- und Forschungsbetrieb, als auch die Herzlichkeit und Gastfreundschaft des dortigen Teams, haben mich in meinem Entschluss für eine akademische Laufbahn bekräftigt.

Ein weiterer essentieller Teil der Crew sind die zahlreichen Kollegen und Freunde, die einen bei einem solchen „Törn“ begleiten. Hier gilt zu aller erst mein Dank meinen Kollegen am Institut für Betriebswirtschaftslehre. Ihre Unterstützung, ihr Rat, der großartige Teamgeist und ihre Hilfe haben mich stets zuversichtlich gestimmt, dass dieses Projekt irgendwann sicher einen Hafen finden wird. Egal ob bei Sturm oder Flaute, dank des intensiven Austauschs zu Forschungs- und Lebensfragen, sowie der einzigartigen Arbeitsatmosphäre und Freundschaften, wurde es an Bord nie langweilig oder fade. Ebenso möchte ich mich ausdrücklich bei unserer Sekretärin Frau Hahn-Mieth bedanken, die stets dafür sorgte, dass ich alle administrativen Klippen erfolgreich umschiffen konnte. Der gleiche Dank gilt auch den zahlreichen wissenschaftlichen Hilfskräften, die mich bei den verschiedenen Projekten tatkräftig unterstützt haben.

Gleichfalls bedanken möchte ich mich bei den fachfremden Mitgliedern der Crew. All den Freunden, die mit Verständnis reagierten, wenn lange Freundschaften wieder zu kurz kamen, vereinbarte Treffen abgesagt wurden und es unser kleines lilanes Boot am Wochenende trotz vieler Versprechen nicht zur Startlinie der norddeutschen Regatten schaffte.

Mein besonderer Dank gilt meinem ältesten Freund Flo, der mir oft half, einmal eine andere Perspektive auf die Dinge zu gewinnen. Ebenso danke ich Mark und Edda, die mich das ein oder andere Mal zum Entspannen auf dem Wasser überredet haben. Auch Markus gilt ein besonderer Dank, da er stets dafür sorgte, dass ich neben all der Theorie nicht den Blick für das Praktische verlor.

Mein tiefster Dank gebührt allerdings meiner Familie, ohne die ich dies alles nie erreicht hätte. Ich danke besonders meiner Verlobten Katharina, für all ihre Toleranz, die liebevollen, tröstenden, bewundernden, mahnenden und verständnisvollen Worte, die mir oft geholfen haben, mich über Erfolge noch mehr zu freuen und Probleme schneller und besser zu überwinden. Ebenso danke ich meiner Schwester Claudia, ihrer Familie und meinen Eltern für ihren Glauben in mich, ihre nie endende Unterstützung und die vielen wertvollen Ratschläge. Euch allen widme ich dieses Buch, das ohne euch nie entstanden wäre.

Raoul V. Kübler

Table of Contents

Synopsis

1 Introduction

The number of product recalls has significantly increased during the last decade (Bapuji and Beamish 2007, Felcher 2003). Shorter product development- and product life cycles as well as the international diversification of the supply chain have increased the risk of major product failures and thus the probability that companies have to recall defective products from the market.

The related costs are substantial, both in human and in financial terms. In the US toy sector alone, 2006 saw 22 toy-related deaths and an estimated 220,500 toy-related injuries. In addition, the public costs related to such product crises or recalls are substantial. According to the U.S. Consumer Product Safety Commission, crisis related costs due to product failures only in the United States sum up to a yearly bill of over 700 billion US-dollars (CPSC 2005).

With the increasing number of product recalls national governments enforced the legal regulations for recalling companies to ensure higher levels of consumer protection. Companies are now obliged to communicate any product failure as early as possible to a widest possible audience of affected customers (e.g. CPSC 2005, GPSG 2004). While this ensures higher levels of consumer protection and is aimed to inform and warn as many affected customers as possible, it simultaneously increases the risks for the recalling company. Previous research has identified serious consequences of product recalls for companies such as, e.g., an immediate loss in sales (Mowen, Jolly and Nickell 1981), a permanent loss in image (de Matos and Rossi 2007, Jolly and Mowen 1985) and brand equity (Dawar and Pillutla 2000) and a negative impact on future buying intentions (Cleeren, Dekimpe and Helsen 2008).

In contrast to products that have been sold directly by the company or that have been registered either by the producer or, as in the case of the automobile industry, by a central authority, most consumer products cannot easily be traced due to the intermediate role of the retailer. In such cases, companies are forced to communicate the product defect to a wide audience consisting of affected and non-affected customers, as well of possible future customers. Companies thus face the risk that the product recall may scare existing or future customers or will lead to long lasting losses of image or market share. Taking into account all the above-mentioned possible negative customer reactions to a product recall, marketers should know how to communicate the particular product defect in public to minimize consumers' negative reactions.

The present work investigates four research questions in the context of the impact of product recall communication on consumers' reactions to product recalls while dealing with methodological as well as empirical aspects:

1. What kinds of communication content and forms of product recall execution do companies use to minimize negative effects?

2. Which context factors (situational variables) play an important role in influencing the behavior of affected companies?
3. How do these communication forms of a recall influence consumer's perception in terms of brand image, brand attitude and perceived product quality?

4. How is this communication induced effect moderated by the recall specific situational components such as the type of product defect or the degree and probability of the defect-induced hazard?

2 Research Contributions

Two scholarships, one granted by the postgraduate program 'Business Aspects of Loosely Coupled Systems and Electronic Business' of the German Research Foundation (Deutsche Forschungsgemeinschaft) and one from the DFG Research Project "Optimization of Product Recalls in Case of Recalls without Customer Identification Information", allowed to investigate the aforementioned research questions in a doctoral thesis. This research has been conducted under the supervision of Prof. Dr. Dr. h.c. Sönke Albers during my employment as a research associate at the Department of Innovation, New Media, and Marketing at Christian-Albrechts-University at Kiel. His support allowed excellent research conditions, which have led to the papers shown in Table 1.

Table 1: Overview of Papers

Methodological Article	
A. Best/Worst-Scaling	Raoul V. Kübler *forthcoming in: Albers, Sönke, Daniel Klapper, Udo Konradt, Achim Walter and Joachim Wolf (eds.): Methodik der empirischen Forschung, 4th ed., Gabler, Wiesbaden 2013.*
Empirical Studies	
B. Communication Behavior of Companies in Case of Product Recalls Without Customer Identification Information	Raoul V. Kübler and Sönke Albers *Published in: Marketing - Journal of Research and Management, Vol. 6 (2010), No. 1, pp. 19-30.*
C. The Impact of Product Recall Communication on Brand Image, Brand Attitude, and Perceived Quality	Raoul V. Kübler and Sönke Albers *Submitted to the Journal of Business Ethics*
D. Faking or Convincing: Why Do Some Advertising Campaigns Win Creativity Awards?	Raoul V. Kübler and Dennis Proppe *forthcoming in: BuR – Business Research*

While the first article (A) presents a comprehensive overview of the technique of preference measurement through best/worst choice tasks, the explorative study (B) identifies which instruments companies use to communicate product recalls. In addition, the study analyzes how marketers adapt these communication instruments to situational factors such as the degree of hazard or the related probability. Whereas Study B lays its focus solely on the identification of possible communication instruments and their usage, study (C) investigates how these instruments affect consumer's perception of a product recall. In particular, it analyzes the impact of the instruments - identified in study (B) - on established brand metrics such as e.g. brand image, brand attitude, or perceived product quality. In contrast to the initial two studies, study (D) addresses a more specific communication orientated research question by investigating which strategies help advertising agencies to win creativity awards. In the following, a brief summary of each paper's contribution to its respective research area is given.

2.1 Methodological Article on Best/Worst-Scaling

Getting detailed insight into customer minds and preferences is a key task of marketers, when addressing consumers and markets (Kotler and Keller, 2006, p83). The specific preferences of a customer do affect the way how she becomes aware of a product, evaluates the product, compares the product to its competitors and finally decides to buy, rebuy or to reject a product (McAlister 1979). Identifying and measuring the preferences of the own or future customers is thus one of the most important tasks for marketers when e.g. developing new products (Wittink, Vriens and Burhenne 1994), targeting and clustering customers (Steiner and Baumgartner 2004) or making pricing decisions (Albers, Becker, Clement, Papies and Schneider 2007). Companies deciding to launch a communication campaign to inform their customers about new products, promotions or even a product defect or a product recall need as well a broad knowledge about their clients' specific preferences (Kotler and Keller

2006). Otherwise they risk that their messages get ignored or immediately forgot. Misinterpreting or ignoring its customers' wishes, needs and preferences may in some cases even result in contrary effects such as refusal or long lasting image losses (Kover 1995).

The article *Best/Worst-Scaling* gives a detailed understanding of a novel and innovative method to gather preference data and to measure consumers' preferences first introduced to academia by Finn and Louviere (1992). In contrast to other state of the art preference measurement approaches, like e.g. Choice-Based-Conjoint-Analyses, Best/Worst-Scaling does not only ask for the most preferred option in a choice set but also for the least preferred option. According to Finn and Louviere (1992) the underlying choice model better reflects human choice behavior than traditional forms of Discrete Choice Experiments and thus leads to better estimation results as it reflects the human habit to react more consistently to extremes.

The article addresses the most relevant issues including (i) the classification and comparison of Best/Worst-Scaling to other state-of-the-art forms of preference measurement, (ii) the theoretical background, (iii) the three main approaches to collect data, (iv) the different estimation approaches, and (iv) software solutions. In sum, the article intends to encourage the knowledge and use of Best/Worst-Scaling to measure preferences in practice and research.

The article is going to be published in the upcoming new edition of the book "Methodik der Empirischen Forschung" edited by Albers, Klapper, Konradt, Walter, and Wolf whose readers are young academics and empirically interested practitioners, who will most hopefully act as a multiplier for this topic.

2.2 Explorative Study on the Communication Behavior of Companies during a Product Recall Scenario Without Customer Information

The increasing pressure for innovations and new products (Trommsdorff and Steinhoff 2007) and the related increasing risk for failing products results in an increasing risk for product recalls. Due to the tightening of the legislation and jurisdiction of consumer safety and consumer protection (CPSIA 2008), companies have nowadays to communicate any product failure to a wide audience as soon as they get knowledge about it. Thus marketers should prepare themselves for such product crisis scenarios, by implementing a recall strategy as soon as they prepare to enter into a new market or to launch a new product, as after the initial failure, time to do so will be limited and the risk for miss-communication and further losses will be high (Berman 1999).

Although research has identified the various possible negative outcomes for recalling companies, such as e.g. long lasting image losses (Mowen, Jolly and Nickell 1981), the loss of consumer trust (Standop 2006), the risk of consumer boycott campaigns (Folkes 1984), a permanent loss of market share (Vassilikopoulou, Lepetsos, Siomkos and Chatzipanagiotou 2009), a drop in stock prices (Ting-Heng, Che-Chun and Prather 2005), and the immediate loss of marketing effectiveness (Van Heerde, Helsen and Dekimpe 2007), only little has yet been done on exploring what instruments companies use to communicate a recall hoping to minimize or encounter the above stated negative effects.

The article "*Communication behavior of companies in product recalls without customer identification information*" closes this research gap by content-analyzing 104 actual cases of recall communication. The analysis identifies six general dimensions of recall communication that can be varied and adapted to the recall scenario specific components such as the degree of hazard, its probability, whether the product is branded or if it represents a high-involvement product. In a second step the study shows that companies usually follow a consistent communication behavior pattern across different situations. The degree of hazard and the probability of a product defect

play an important role in expressing social responsibility and showing transparency. Furthermore, we find empirical evidence for the fact that companies producing branded and high-involvement products organize their recalls to be more convenient for their customers.

Beside helping affected marketers to better plan and implement recall communication campaigns the results are intended to provide a base for further research on the impact of product recall communication on consumer's perception and evaluation of recall campaigns, the recalled products and the affected companies.

2.3 Empirical Study on the Impact of Corporate Recall Communication on Different Brand Metrics

The main target of the previous study was to identify suitable communication strategies, on which companies can basically rely when communicating a product recall in public. Due to the explorative character of the study, the results do not allow to provide a deeper insight into how the different identified communication tools and patterns do influence consumer's perception and reaction to the specific product recall. Taking into account all possible negative outcomes of a product recall - as described in the introduction and the previous section -, affected managers should know how consumers are going to react to their recall communication campaigns. Only then, they can design an optimal recall campaign and publicly communicate a product failure without having to fear to risk the future of the company and its stakeholders.

Therefore the article "*The impact of product recall communication on brand image, brand attitude, and perceived quality*" analyzes how the in the previous study identified different patterns of recall communication and their different characteristics do influence consumer's evaluation of the recalling company.

Relying on more than 2.200 answers from 450 different study participants assessing 16 different major European product recall campaigns, the article reveals how consumers react to the before identified attributes of a recall message. Using these

main attributes of recall campaigns (social responsibility, transparency, source of recall and convenience of product return), the study determines with the help of a seemingly unrelated regression model the impact of the different attribute characteristics on the three major brand metrics brand attitude, brand image and perceived product quality. To allow for more generalizable results the study further implements three moderating effects, that help to better understand how situational factors like the degree and the probability of hazard and the source of recall (viz. the media channel in which a consumer first got information about the recall) influence the impact of the different recall communication attributes on the different brand metrics.

The results show that stating social responsibility does not lead to a better evaluation of the recalling company. Surprisingly, higher degrees of stated social responsibility seem even to enforce negative customer reactions. In contrast, high degrees of transparency as well as a high level of product return convenience help companies to secure higher levels of brand attitude, image and perceived product quality. Marketers have to be aware that two of the afore-mentioned effects are moderated by the degree and the probability of hazard. It appears that higher degrees of hazard even increase the negative impact of social responsibility and that a higher hazard probability decreases the positive impact of transparency.

In addition, the study shows for the first time, that the impact of the different elements of the recall communication is significantly moderated by the media-channel in which the message is published.

2.4 Empirical Study on Success Drivers of Advertising Campaigns in Creativity Award Shows

Whereas the previous two empirical studies focused on how to communicate a product recall, the third empirical study focuses on a more specific topic in advertising research and creativity research.

Since the Sarbanes-Oxley Act was passed in 2002, it has become commonplace in the advertising industry to use creativity award show prizes (like e.g. the Cannes Lions or the One Show Award) instead of gross income figures to attract new customers (Butkys and Herpel 1992, Helgesen 1994). Therefore, achieving top placements in creativity rankings and winning creativity awards have become high priorities in the advertising industry (Myers 2004). To showcase their creative skills, some copywriters and art directors started to develop campaigns for "alibi" or fictitious clients. Producing advertising for fictitious clients allows agencies to avoid daily business restrictions or long discussions with clients or product managers, who may want to replace creative concepts with less creative approaches (West 1993). By using these "fake campaigns" agencies try to maximize their outcomes at award shows.

Agencies and marketers have always wondered what elements in the advertising creation process would lead to winning creativity awards. The related debate has been dominated since years by pure speculation about the success of different routines, approaches and strategies in winning creativity awards, while research has focused on finding a fundamental concept of creativity in advertising (Amabile 1993, Ang, Lee and Leong 2007, El-Murad and West 2004) and the impact of creative advertisements on buying behavior (Bell 1992, Bergkvist and Rossiter 2008).

The article *„Faking or Convincing: Why Do Some Advertising Campaigns Win Creativity Awards?"* addresses this shortcoming by examining a data set of 108 German advertising campaigns that won different international creativity awards. The study theoretically derives a novel set of factors that influence the success of an individual campaign in a creativity award show. Additionally, the underlying model measures the influence of several strategic factors and routines on creativity award show success. The results challenge anecdotal industry knowledge in not finding any support for the assertion that fake campaigns are more likely to win creativity awards. Further, the results help copywriters and art directors to win creativity award shows by producing novel approaches in terms of layout generation, conceptualization and

media usage. The study results further deliver some important topics for future research into advertising creativity. Until now, empirical studies have simply operationalized advertising creativity by the fact that these campaigns have won some type of creativity award in various international award shows. Researchers should thereby keep in mind that winning creativity awards primarily reflects the novelty of an approach, whereas the other two aspects of the general construct of creativity, i.e., meaningfulness and connectedness, seem to be ignored by festival juries.

3 Manifestations on Co-Authorships

Three of the four papers have been developed in co-authorship. In all of the three cases the order of authors reflects the contribution to the paper.

References

Albers, Sönke, Jan U. Becker, Michel Clement, Dominik Papies and Holger Schneider (2007): Messung von Zahlungsbereitschaften und ihr Einsatz für die Preisbündelung, *Marketing ZFP*, 29 (1): 7-23.

Amabile, Teresa M. (1993): What Does a Theory of Creativity Require?, *Psychological Inquiry*, 4 (3): 179-182.

Ang, Swee Hoon, Yih Hwai Lee and Siew Meng Leong (2007): The Ad Creativity Cube: Conceptualization and Initial Validation, *Journal of the Academy of Marketing Science*, 35 (2): 220-232.

Bapuji, Hari and Paul W. Beamish (2007): *Toy Import and Recall Levels: Is There a Connection?*, Source: http://www.asiapacific.ca/analysis/pubs/pdfs/commentary/cac45.pdf.

Bell, Jack A. (1992): Creativity, TV Commercial Popularity, and Advertising Expenditures, *International Journal of Advertising*, 11 (2): 165-172.

Bergkvist, Lars and John R. Rossiter (2008): The Role of Ad Likability in Predicting an AD's Campaign Performance, *Journal of Advertising*, 37 (2): 85-97.

Berman, Barry (1999): Planning the Inevitable Product Recall, *Business Horizon*, 42 (3/4): 69-78.

Butkys, Adolph S. and George Herpel (1992): How Advertising Agencies Handle Their Own Advertising Strategy: An Industry-Wide Overview of its Self-Promotion Efforts, *Journal of Advertising Research*, 32 (5): 18-22.

Cleeren, Kathleen, Marnik G. Dekimpe and Kristiaan Helsen (2008): Weathering Product-Harm Crises, *Journal of the Academy of Marketing Science*, 36 (2): 262-270.

CPSC (2005): US Consumer Product Safety Commission 2005 Performance and Accountability Report, Source: http://www.cpsc.gov/cpscpub/pubs/reports/2005rpt.pdf.

CPSIA (2008): Consumer Product Safety Improvement Act, http://www.cpsc.gov/cpsia.pdf.

Dawar, Niraj and Madan M. Pillutla (2000): Impact of Product-Harm Crises on Brand Equity: The Moderating Role of Consumer Expectations, *Journal of Marketing Research*, 37 (2): 215-226.

de Matos, Celso Augusto and Carlos Alberto Vargas Rossi (2007): Consumer reaction to product recalls: factors influencing product judgement and behavioural intentions, *International Journal of Consumer Studies*, 31 (1): 109-116.

El-Murad, John and Douglas C. West (2004): The definition and measurement of creativity: What do we know?, *Journal of Advertising Research*, 44 (2): 188-201.

Felcher, E. Marla (2003): Product Recalls: Gaping Holes in the Nation's Product Safety Net, *Journal of Consumer Affairs*, 37 (1): 170-179.

Finn, Adam and Jordan Louviere (1992): Determining the Appropriate Response to Evidence of Public Concern: The Case of Food Safety, *Journal of Public Policy & Marketing*, 11 (1): 12-25.

Folkes, Valerie S. (1984): Consumer Reactions to Product Failure: An Attributional Approach, *Journal of Consumer Research*, 10 (4): 398-409.

GPSG (2004): *Gesetz über technische Arbeitsmittel und Verbraucherprodukte (Geräte- und Produktsicherheitsgesetz – GPSG)*, Source: http://bundesrecht.juris.de/-bundesrecht/gpsg/gesamt.pdf.

Helgesen, Thorolf (1994): Advertising Awards and Advertising Agency Performance Criteria, *Journal of Advertising Research*, 34 (4): 43-53.

Jolly, David W. and John C. Mowen (1985): Product Recall Communications: The Effects of Source, Media and Social Responsibility Information, *Advances in Consumer Research*, 12 (1): 471-475.

Kotler, Philip and Kevin Keller (2006): *Marketing Management*, Prentice Hall: Upper Saddle River.

Kover, Arthur J. (1995): Copywriters' Implicit Theories of Communication: An Exploration, *Journal of Consumer Research*, 21 (4): 596-611.

McAlister, Leigh (1979): Choosing Multiple Items from a Product Class, *Journal of Consumer Research*, 6 (3): 213-224.

Mowen, John C., David Jolly and Gary S. Nickell (1981): Factors Influencing Consumer Responses to Product Recalls: A Regression Analyses Approach, *Advances in Consumer Research*, 8 (1): 405-406.

Myers, Jack (2004): Advertising Accountability, R-O-I & Sarbanes-Oxley Concerns Spawn Audit Industry, Source: http://www.mediavillage.com/jmr/2004/07/07/jmr-7-7-04/.

Standop, Dirk (2006): Der Verlust von Konsumentenvertrauen gegenüber Anbietern: der Fall von Produktrückrufen, in: Hans Bauer, Marcus Neumann and Anja Schüle (eds.): *Konsumentenvertrauen*, Vahlen, München, 95-104.

Steiner, Winfried J. and Bernhard Baumgartner (2004): Conjoint-Analyse und Marktsegmentierung, *Zeitschrift für Betriebswirtschaft*, 74 (6): 611-635.

Ting-Heng, Chu, Lin Che-Chun and Larry J. Prather (2005): An Extension of Security Price Reactions Around Product Recall Announcements, *Quarterly Journal of Business & Economics*, 44 (3/4): 33-48.

Trommsdorff, Volker and Fee Steinhoff (2007): *Innovationsmarketing*, Vahlen: München.

Van Heerde, Harald, Kristiaan Helsen and Marnik G. Dekimpe (2007): The Impact of a Product-Harm Crisis on Marketing Effectiveness, *Marketing Science*, 26 (2): 230-245.

Vassilikopoulou, Aikaterini, Apostolos Lepetsos, George Siomkos and Kalliopi Chatzipanagiotou (2009): The Importance of Factors Influencing Product-Harm Crisis Management Across Different Crisis Extent Levels: A Conjoint Analysis, *Journal of Targeting, Measurement & Analysis for Marketing*, 17 (1): 65-74.

West, Douglas C. (1993): Cross-National Creative Personalities, Processes, and Agency Philosophies, *Journal of Advertising Research*, 33 (5): 53-62.

Wittink, Dick R., Marco Vriens and Wim Burhenne (1994): Commercial Use of Conjoint Analysis in Europe: Results and Critical Reflections, *International Journal of Research in Marketing*, 11 (1): 41-52

A Best/Worst-Scaling

Raoul V. Kübler

Erscheint in:
Albers, Sönke, Daniel Klapper, Udo Konradt, Achim Walter, and Joachim Wolf (eds.): *Methodik der empirischen Forschung*, 4th ed., Gabler, Wiesbaden.

Online verfügbar unter: http://www.bwl.uni-kiel.de/bwlinstitute/grad-kolleg/new/typo3conf/ext/naw_securedl/secure.php?u=0&file=/fileadmin/publications/pdf/Methodik_der_empirischen_Forschung_-_Best_Worst_Scaling__Raoul_Kuebler.pdf&t=1337257782&hash=526f726aec34cf9d71e6a9595f642171

Web Appendix:

1. Beispiel Count Analyse in MS Excel

1 Grundgedanke des Best/Worst Scaling

Das Best/Worst-Scaling (B/W-Scaling), das in vielen Studien auch als Max/Diff-Scaling bezeichnet wird, ist eine Sonderform der Discrete-Choice-Analyse (DCA). Analog zur Choice-Based-Conjoint-Analyse (CBC) (vgl. Aufsatz Himme in diesem Buch) liegt das Ziel des B/W-Scalings in der Erfassung von Präferenzen hinsichtlich von Attributen und deren Ausprägungen. Hierzu wählen die Probanden mehrmals zwischen verschiedenen speziell erstellten Stimuli in mehreren Choice-Sets aus. Mittels statistischer Verfahren können so Rückschlüsse über die Präferenzen der Studienteilnehmer für die einzelnen Attribute und deren Ausprägungen berechnet werden. Während bei klassischen Discrete-Choice-Experimenten die Probanden

jeweils die beste Alternative in einem Choice-Set angeben müssen, markieren bei B/W-Scaling Experimenten die Probanden zusätzlich zur besten Alternative auch noch die als am schlechtesten betrachtete Alternative (Finn und Louviere 1992).

Entsprechend seiner Verwandtschaft zur CBC-Analyse kann das B/W-Scaling ebenfalls als Methode zur Produkt- oder Preisgestaltung genutzt werden (Louviere und Islam 2008). Auf Basis der gewonnenen Präferenzen können Unternehmen nutzenoptimale Produkte gestalten oder die maximale Zahlungsbereitschaft ihrer Kunden ermitteln. Ebenso kann das B/W-Scaling zur Segmentierung der eigenen Kunden anhand der gemessenen Präferenzen genutzt werden.

Im Gegensatz zur CBC-Analyse, die zwingend das Vorhandensein ganzheitlich komponierter und multiattributer Stimuli voraussetzt (Louviere, Henscher und Swait 2000), erlaubt das B/W-Scaling auch die Wahl zwischen einzelnen Attributen mit nur einer Ausprägung (Potoglou, Burge, Flynn, Netten, Malley, Forder und Brazier 2011). Somit kann, je nach Gestaltung der von den Probanden wählbaren Stimuli und deren Zusammensetzung, beim B/W-Scaling zwischen drei grundsätzlichen Formen (Cases) von Experimenten unterschieden werden:

Bei den so genannten Case1-Experimenten wählen die Probanden lediglich zwischen verschiedenen Attributen aus, die aber im Gegensatz zur den bekannten Formen anderer DC-Experimente jeweils nur eine Ausprägung besitzen. Die Attribute werden hierbei mittels eines experimentellen Designs auf die verschiedenen Choice-Sets verteilt und die Probanden wählen stets das am meisten und das am wenigsten präferierte Attribut des Sets. Aufgrund des Auswahlverhaltens lässt sich die Wichtigkeit der einzelnen Attribute berechnen. Somit kann das B/W-Scaling auch als eine Alternative zur klassischen Messung von Wichtigkeiten oder Einstellungen mittels skalenbasierter Ansätze (wie z.B. der Abfrage durch Likert-Skalen) verwendet werden (Lee, Soutar und Louviere 2007).

So genannte Case2-Experimente erlauben die Nutzung von Attributen mit mehreren Ausprägungen. Allerdings wird den Probanden im Gegensatz zur CBC-Analyse nicht

ein Choice-Set mit mehreren ganzheitlich konstruierten Produkten zur Wahl vorgelegt, sondern jeweils nur ein konstruiertes Produkt. Die Probanden markieren dann die für sie jeweils beste und schlechteste Ausprägung eines Attributes innerhalb dieses Produktes. Analog zu anderen Formen der DC-Analyse wird diese Auswahlentscheidung für verschiedene Produkte wiederholt, bis genügend Beobachtungen vorliegen, um die Wichtigkeit und den Nutzen der Attribute und ihrer einzelnen Ausprägungen mittels statistischer Verfahren zu bestimmen.

Case3-Experimente verlaufen analog zu den klassischen CBC-Analysen. Den Probanden werden verschiedene Choice-Sets mit mehreren Stimuli zur Wahl vorgelegt. Die Stimuli bestehen hierbei aus denselben Attributen, unterscheiden sich aber in deren Ausprägungen. Im Gegensatz zur CBC-Analyse markieren aber die Probanden nicht nur die am meisten präferierte Alternative, sondern auch die am wenigsten präferierte.

Der vorliegende Aufsatz soll einen anwendungsorientierten Einblick in die Anwendung von B/W-Experimenten zur Erfassung von Präferenzen geben. Hierzu wird im nächsten Abschnitt zuerst auf die Stärken des B/W-Scalings im Vergleich zu anderen Formen der Präferenzmessung eingegangen. Danach erfolgt ein abgrenzender Überblick über die drei verschiedenen Formen des B/W-Scalings. Im Anschluss wird detailliert das Vorgehen bei einer B/W-Erhebung erläutert. Der Aufsatz endet mit einem Überblick über die gängigen Softwarelösungen für B/W-Experimente und einer Zusammenfassung.

2 Vorteile des B/W-Scalings im Vergleich zu anderen Formen der DCA

Die Abfrage der besten als auch schlechtesten Alternative bietet im Vergleich zu bisherigen Formen der DCA - wie der CBC-Analyse - für den Wissenschaftler mehrere Vorteile:

So lassen sich durch die Angabe der Präferenzextrema in Form der besten und der schlechtesten Alternative schneller mehr Informationen über die Präferenzstruktur der Probanden gewinnen (Louviere, Street, Burgess, Wasi, Islam und Marley 2008).

Gehen wir beispielsweise davon aus, dass ein Media-Unternehmen ein neues Produkt auf den Markt bringen will. Um zu erfassen, welche Media-Kanäle hierfür bedient werden sollten, ist zu erfassen, welche Quellen Konsumenten zur Informationsbeschaffung am liebsten nutzen. Basierend auf den Ergebnissen soll dann ein spezielles Produkt für diesen Kanal weiter entwickelt werden. Zur Auswahl stehen Tageszeitungen, Fernsehen, Wochenmagazine und das Internet. Mittels einer B/W-Studie werden verschiedene Probanden befragt.

Wir nehmen an, dass einer der befragten Probanden sehr internetaffin ist und die Kosten einer Tageszeitung scheut. Er wird sich daher für das Internet als beste Alternative entscheiden und gleichzeitig die Tageszeitung als schlechteste Alternative markieren. Dadurch lassen sich für den Experimentator aus dieser Information weitere wichtige Informationen ableiten. So wissen wir, dass für den Probanden das Internet besser ist als das Fernsehen, die Wochenzeitung und die Tageszeitung. Gleichzeitig wissen wir aber nun auch, dass die Tageszeitung am wenigsten präferiert wird. Analog lässt sich daher ableiten, dass die Wochenzeitung und das Fernsehen für den Probanden besser sind als eine Tageszeitung. Die einzige Information, die nicht direkt abgeleitet werden kann, ist der direkte Vergleich zwischen Wochenzeitung und Fernsehen, da hierfür keine Informationen erfasst wurden.

Finn und Louviere (1992) zeigen darüber hinaus, dass es Probanden einfacher fällt, gleichzeitig beide Extrempunkte ihrer Präferenz - also das am meisten präferierte und das am wenigsten präferierte Attribut - zu benennen, als sich in schwierigen Auswahlsituationen zwischen ähnlichen Stimuli für einen Besten zu entscheiden, oder gar eine komplette Rangfolge für eine Vielzahl an Stimuli zu bilden. Marley und Louviere (2005) betonen weiter, dass das B/W-Scaling, durch seinen Bezug auf die menschliche Neigung, konsistenter auf Extrema zu reagieren, anderen Formen der DCA überlegen ist.

Ebenso wird von Finn und Louviere (1992) argumentiert, dass die Wahl eines Attributpaares oder zweier Stimuli zur besten und schlechtesten Alternative auch als der maximale Abstand zwischen den gewählten Alternativen auf der persönlichen

Nutzenskala des Probanden verstanden werden kann. Die schlechteste Alternative bildet dabei den Nullpunkt der Skala und die beste Alternative das Maximum der jeweiligen Präferenzskala. Durch die gleichzeitige Benennung des Maximums und des Minimums (bzw. der maximalen Differenz = MaxDiff) erfolgt somit eine Normierung der individuellen Präferenzskala.

3 Überblick über die verschiedenen Formen des B/W-Scaling

3.1 Auswahl zwischen einzelnen Attributen - Case1

Case1 stellt die Urform des B/W-Scalings dar, wie sie 1992 von Finn und Louviere vorgestellt wurde. Case1-Experimente dienen alleine zur Messung der Wichtigkeit einzelner Attribute. Sie können somit als Alternative zur Erfassung von Wichtigkeiten mittels Ratingskalen angesehen werden.

Typischerweise werden bei Case1-Studien verschiedene Attribute, die nur eine Ausprägung haben, gemäß einem experimentellen Designplan auf Choice-Sets verteilt. Die Probanden wählen dann stets das von ihnen am meisten und am wenigsten präferierte Attribut innerhalb eines Sets aus. Dieses Vorgehen wird über mehrere Choice-Sets wiederholt, bis - abhängig von der Anzahl an Attributen - genügend viele Beobachtungen vorliegen. Mittels eines Discrete-Choice-Models (siehe Abschnitte 4.4.2 bis 4.4.4) oder mittels einfacher Auszählung der Häufigkeiten, (siehe Abschnitt 4.4.1) kann dann die Wichtigkeit jedes Attributes bestimmt werden.

Tabelle 1 gibt ein Beispiel für eine Case1-Erhebung im Falle unseres eingangs dargestellten Beispiels zur Messung der Wichtigkeit von verschiedenen Medien-Kanälen. Eine Messung der Wichtigkeit einzelner Attribute anhand von Case1-Experimenten umgeht die üblichen Probleme, die bei skalenbasierten Ansätzen vorherrschen. So unterdrückt der DCA-typische Trade-Off-Mechanismus sozial erwünschtes Antwortverhalten (Auger, Devinney und Louviere 2007) und führt zu einer Diskriminierung zwischen den verschiedenen zur Wahl stehenden Attributen. Dies führt zu dem Effekt, dass - im Gegensatz zur Abfrage mittels Skalen - nicht alle

Attribute gleich hoch bewertet werden können und somit eine finale Unterscheidung oder Gewichtung der Attribute möglich ist (Bacon, Lenk, Seryakova und Veccia 2008).

Tabelle 1: Beispiel eines Case1-Choice-Sets

Beliebtestes Medium zur Informationsbeschaffung		Unbeliebtestes Medium zur Informationsbeschaffung
x	Internet	
	Fernsehen	
	Tageszeitung	x
	Wochenzeitung	

Weiter schließen Case1-Experimente Ipsativität aus (Baumgartner und Steenkamp 2001). Unter Ipsativität wird allgemein der Unterschied im Niveau der Beantwortung verstanden (Meade 2004). Es wurde gezeigt, dass Konsumenten sich aufgrund sozialer oder kultureller Unterschiede in der allgemeinen Höhe ihrer maximalen und minimalen Bewertung unterscheiden (Paulhus 1991). Während manche Probanden dazu neigen, aus Höflichkeit als schlechteste Bewertung auf einer Skala von eins bis sieben stets nur eine vier zu vergeben, neigen sehr kritische Probanden hingegen dazu, als beste Note maximal nur eine fünf zu vergeben. Baumgartner und Steenkamp (2001) zeigen, dass derartige Verhaltensmuster zu einer problematischen Verzerrung der Mittelwerte und Varianzen bei einer Erhebung mit „reinen“ Skalen führen können. Deshalb können keine aussagekräftigen Vergleiche zwischen Individuen oder verschiedenen Erhebungsorten getroffen werden. Dieses Problem wird von Case1-Experimenten umgangen, da die Probanden durch die Angabe der besten und der schlechtesten Alternative eine quasi normierte Spannweite angeben (Cohen und Orme 2004).

Somit kann die Case1-Variante des B/W-Scalings als wirkungsvolle Alternative zu den traditionellen, skalenbasierten Ansätzen zur Messung von Einstellungen verstanden werden (Lee, Soutar und Louviere 2008). Aufgrund dieser Vorteile kann man Anwendungen von Case1-Experimenten sowohl im Bereich der Erfassung

sozialer und ethischer Werte beim Konsum sozial verantwortlicher Produkte (Auger, Devinney und Louviere 2007) als auch bei der Erfassung der generellen Einschätzung der eigenen Lebensqualität (Lee, Soutar und Louviere 2008; Lee, Soutar und Louviere 2007) sowie bei der Bedeutung so genannter weicher Faktoren bei der Wahl der Universität (Tavares, Cardoso und Dias 2010) und der Messung der Mitarbeiterzufriedenheit (Garver, Zachary und LeMay 2010) finden.

Im Gegensatz zur Conjoint-Analyse wird allerdings bei Case1-Erhebungen zwischen den Choice-Sets nicht die Ausprägung der einzelnen Attribute verändert, sondern alleine die Zusammensetzung der einzelnen Choice-Sets.

3.2 Auswahl innerhalb eines Produktes oder einer Situation - Case 2

Case2-Experimente erlauben eine Variation der Ausprägungen einzelner Attribute. Sie erfassen dabei die Präferenzen für jede einzelne Attributausprägung. Wie in Tabelle 2 dargestellt, wird im Gegensatz zu den bekannten Formen der CBC-Analyse bei Case2 allerdings nicht zwischen verschiedenen fertig komponierten, multiattributen Stimuli gewählt, sondern nur zwischen den verschiedenen Ausprägungen eines einzelnen konstruierten Stimuli.

Mittels eines experimentellen Designs werden die verschiedenen Attributausprägungen zu ganzheitlichen Stimuli zusammengefügt. Hierbei bestehen alle Stimuli aus denselben Attributen und unterscheiden sich nur anhand der Kombination ihrer Ausprägungen. Im Gegensatz zur CBC-Analyse besteht ein Choice Set bei Case2-Experimenten allerdings nicht aus mehreren Stimuli. Den Probanden wird lediglich jeweils ein Stimulus vorgelegt. Die Probanden geben nun an, welche Ausprägung sie bei diesem Stimulus als am besten und welche Ausprägung sie als am schlechtesten bewerten. Dies wird gemäß dem experimentellen Design so lange wiederholt, bis genügend Beobachtungen vorliegen, um die Präferenzen schätzen zu können.

Tabelle 2 gibt ein Beispiel für ein Choice-Set eines Case2-Experimentes zur Messung von Präferenzen bei der Informationsbeschaffung. Hierbei besteht das Set aus den

festen Attributen: Lieferung, Medium, Kosten und Mediaform, die sich nicht zwischen den Sets verändern.

Tabelle 2: Beispiel für Case 2 Choice-Set

Beste Komponente	Attribut	Ausprägungsform	Schlechteste Komponente
	Lieferung	Abonnement	x
	Medium	Wochenzeitung	
	Kosten	15 € / Monat	
x	Mediaform	Print und ebook	

Die Probanden geben nun an, welche Attributausprägungen des vorliegenden Produktes für sie die besten und die schlechtesten darstellen. In Abhängigkeit vom Design und der Anzahl an Ausprägungen pro Attribut werden den Probanden nun wiederholt verschiedene Sets vorgelegt. Basierend auf den Auswahlentscheidungen der Probanden kann dann der jeweilige Gesamtnutzen durch die Addition der Teilnutzenwerte für jede einzelne Attributausprägung berechnet werden (Marley und Louviere 2005).

Im Vergleich zu Case1 bietet Case2 die Möglichkeit, verschiedene Ausprägungen der Attribute in die Erhebung einzubeziehen. Gleichzeitig bleiben aber - im Vergleich zu einer Conjoint-Analyse - der Beobachtungsaufwand und die Belastung der Probanden durch die Konfrontation mit nur einem konstruierten Produkt oder einer Situation gering. Dies kann als weiterer Vorteil des B/W-Scaling betrachtet werden. Während in den klassischen Anwendungsfeldern der Präferenzmessung der Vergleich mehrerer Stimuli in Form einer simulierten Kaufentscheidung bzw. der Auswahl zwischen mehreren Produkten real erscheinen mag, stellt sich vielfach bei Präferenzmessungsprojekten mit einem Fokus jenseits einer Kaufentscheidung das Problem, dass diese wenig realistisch ist. Entsprechend führt eine gleichzeitige Wahlmöglichkeit im Falle vorgegebener Situationen zu einer Verringerung des Realitätscharakters der Studie, was wiederum zu Verzerrungen bei den gemessenen Präferenzen und einem Anstieg der Mess-Fehler führen kann.

Diese Problematik wird durch die Case2-Variante umgangen, da die Probanden hier immer nur mit einem durch ein experimentelles Design erstelltes Szenario konfrontiert werden, indem die Ausprägungen der festen situationsspezifischen Attribute entsprechend variiert werden und nur markiert werden soll, was in der gegebenen Situation als der beste und der schlechteste Aspekt betrachtet wird. Beachtet werden muss hierbei allerdings, dass Case2-Erhebungen durch die spezielle Gestaltung nur schwer die Integration von Interaktionseffekten zulassen (Louviere und Islam 2008).

Bisher wurden nur wenige Case2-Studien publiziert. Anwendungen fanden dabei zur Messung von Kundenzufriedenheit (Chrzan und Golovashkina 2006), von Patienten-Präferenzen im medizinischen Bereich (Flynn, Louviere, Peters und Coast 2008) und von Kunden-Präferenzen beim Kauf von Wein (Cohen 2009) statt.

3.3 Auswahl zwischen verschiedenen Stimuli - Case 3

Während die Gestaltung von Case1- und Case2-Experimenten keine realistische Simulation einer Kaufsituation zulässt, können die Probanden bei Case3-Experimenten zwischen mehreren unterschiedlich gestalteten Produktalternativen innerhalb eines Choice-Sets auswählen. Typischerweise werden diese Produkte analog zur CBC-Analyse mittels eines experimentellen Designs erstellt. Hierbei unterscheiden sich - wie beispielhaft in Tabelle 4 dargestellt - die einzelnen Alternativen eines Choice-Sets nicht anhand ihrer Attribute, sondern anhand von deren Ausprägungen. Die Probanden wählen im Falle eines Case3-Experimentes analog zu Case1 und Case2 die von ihnen am meisten präferierte und die von ihnen am wenigsten präferierte Alternative. Kritisch anzumerken ist, dass die zeitliche Belastung der Probanden im Vergleich zur CBC oder gar zur Adaptive Conjoint-Analyse höher ist, da Probanden pro Choice-Set zwei Auswahlentscheidungen treffen müssen (Cohen und Orme 2004). Entsprechend ist zu erwarten, dass Ermüdungseffekte früher auftreten, was allerdings noch nicht empirisch gezeigt werden konnte. Bisher finden sich nur wenige empirische Arbeiten, die einen Case3 Ansatz zur Messung von Präferenzen heranziehen, was sich vor allem durch die geringe Verbreitung an Software-Lösungen zur passenden Erstellung von Fragebögen und der Schätzung der einzelnen Präferenzfunktionen erklären lässt.

Tabelle 4: Beispiel für Case3-Choice-Set

	Alternative1	*Alternative2*	*Alternative3*	*Alternative4*
Abo-Laufzeit	12 Monate	3 Monate	6 Monate	24 Monate
Mediaform	Print	ePaper	ePaper & Print	Print
Medium	Tageszeitung	Online Newsgroup	Wochenzeitung	Tageszeitung
Preis / Monat	16 Euro	5 Euro	18 Euro	13 Euro
Am meisten bevorzugte Alternative	x			
Am wenigsten bevorzugte Alternative		x		

Louviere, Street, Burgess, Wasi, Islam und Marley (2008) untersuchen mittels verschiedener Case3-Experimente Kundenpräferenzen hinsichtlich Langstreckenflügen und Tiefkühlnahrungsmittel und vergleichen die gewonnen Daten mit den Ergebnissen von analog durchgeführten Discrete-Choice-Experimenten. Ihre Ergebnisse zeigen, dass Case3-Erhebungen unabhängig vom Design und der Stichprobengröße ebenso reliable Ergebnisse liefern wie CBC- und Adaptive-CBC-Ansätze.

Mit einem ähnlich gelagerten Kontext vergleichen Louviere und Islam (2008) Case3-Experimente mit verschiedenen anderen Formen der Präferenzmessung, wie der CBC-Analyse und der direkten Messung der „Willingness-to-Pay“. Mittels Simulationen und mehrerer empirischer Beispiele, mit dem Hintergrund, Präferenzen verschiedener Lebensmittel-Lieferdienste und Fruchtsäfte zu messen, zeigen die beiden Autoren, dass mittels Case3-Experimenten Präferenzen aufgrund der höheren Informationsmenge bei gleicher Anzahl an Choice-Tasks besser und genauer gemessen werden können.

4 Ablauf eines B/W-Scaling Experimentes

4.1 Theoretisches Wahl-Modell

Wie eingangs beschrieben, kann das B/W-Scaling generell den Discrete-Choice-Experimenten zugeordnet werden. So stellt das theoretische Grundmodell eine Erweiterung des Mulitnomialen-Logit-Modells (MNL) auf Basis der Arbeiten von Thurstone (1927) und McFadden (1974) dar. Für eine generelle Einführung in die Modellbildung der Discrete Choice Theory wird auf den Aufsatz von Temme in diesem Buch verwiesen.

Da das MNL-Modell nur die Wahrscheinlichkeit für die Wahl einer Alternative abbildet, muss es für die Berechnung der Wahrscheinlichkeit, dass sowohl eine beste als auch eine schlechteste Alternative gewählt werden, um die letztere Auswahlwahrscheinlichkeit erweitert werden. Marley und Louviere (2005) zeigen hierfür, dass sich für jedes Choice-Set mit mehr als zwei Alternativen die B/W-Wahrscheinlichkeit, dass die Stimuli j und k eines Choice-Sets zusammen als beste (j) und schlechteste (k) Alternative ausgewählt werden, als das Produkt aus der Best-Wahrscheinlichkeit B_j und der Worst-Wahrscheinlichkeit W_k abbilden lässt.

Die Wahrscheinlichkeit B_j, dass der Stimulus j als beste Alternative gewählt wird, kann hierbei unter Rückgriff auf das klassische MNL-Modell gemäß McFadden (1974) dargestellt werden. Unter der Annahme, dass die nicht beobachtbaren Nutzenelemente ε_j und ε_r des individuellen Nutzens eines Entscheiders einer Gumbel-Verteilung folgen, sowie der gebräuchlichen Annahme, dass die Auswahl der besten Alternative B_j auf den identischen intervallskalierten Werten der Nutzenkomponenten der betrachteten Alternativen beruht, folgt, dass für alle j gilt:

$$(1) \qquad B = \frac{\exp(V_j)}{\sum_j \exp(V_j)}$$

Die Annahme, dass die Störterme einer Gumbel-Verteilung folgen, gilt allerdings nicht für die Wahlwahrscheinlichkeit der schlechtesten Alternative. Entsprechend kann für deren Bestimmung nicht auf das klassische MNL-Modell zurückgegriffen werden.

Gemäß Wirth (2010) kann allerdings für die Bestimmung der Wahrscheinlichkeit Wk, dass k aus der verbleibenden Menge an Alternativen im Choice Set C (ohne j) als die schlechteste Alternative ausgewählt wird, auf die folgende Identität zurückgegriffen werden:

(2) $$W_k = \sum_{\eta \in R(C\backslash \mathrm{j,k})} B_{C\backslash(\mathrm{k})}(\eta_2)...B_{\eta(J-1,k)}(\eta_{J-1})$$

Hierbei ist R(C\(j,k)) die Menge aller möglichen weiteren Rangfolgen von noch zur Wahl stehenden Attributen in einem Choice-Set C\(j,k), das die beiden gewählten Alternativen j (beste) und k (schlechteste) nicht mehr beinhaltet. $\eta = \eta_2 \eta_3 ... \eta_{J-1}$ ist dann als eine beliebige Rangfolge mit den Elementen η_2 bis η_{J-1} zu verstehen. Hierbei wird die Position einer Alternative in einer Rangfolge als das Produkt der Best-Wahlwahrscheinlichkeiten innerhalb dieser Rangfolge ausgedrückt.

Zur besseren Veranschaulichung der Berechnung der Worst-Wahrscheinlichkeit W_k soll auf die von Wirth (2010) hergeleitete Darstellung unter Rückgriff auf das eingangs genannte Beispiel zur Messung der Wichtigkeit verschiedener Medienkanäle bei der Informationsbeschaffung zurückgegriffen werden. Die Probanden konnten zwischen einer Tageszeitung (TZ), einer Wochenzeitung (WZ), dem Internet (IT) und dem Fernsehen (TV) wählen. Wie beschrieben wird IT als beste Alternative j und TZ als schlechteste Alternative k ausgewählt. Somit umfasst die Menge R(C\(IT,TV)) aller möglichen weiteren Rangfolgen im Choice Set C die Rangfolgen η=WZ,TV und η =TV,WZ. Die Wahrscheinlichkeit W(TZ), dass TZ als schlechteste Alternative gewählt wird, kann dann wie folgt bestimmt werden:

(3) $$W_{(TZ)} = B(TV) * B_{(WZ,TZ)}(WZ) + B(WZ) * B_{(TV,WZ)}(TV)$$

Hierbei drückt $B_{(WZ,TZ)}(WZ)$ (bzw. $B_{(TV,WZ)}(TV)$) die Wahrscheinlichkeit aus, dass bei einer Wahl zwischen WZ und TZ, WZ präferiert wird (bzw. bei einer Wahl zwischen TV und WZ, TV präferiert wird). Alle hierbei verwendeten Wahrscheinlichkeiten können somit als eine Best-Wahlwahrscheinlichkeit betrachtet werden. Diese erfüllen somit auch die Anforderungen eines normalen MNL-Modells und können folglich ebenso gemäß Gleichung 1 durch ein klassisches MNL-Modell bestimmt werden.

Multipliziert man nun die Wahrscheinlichkeit zur Wahl der besten Alternative (Gleichung 1) mit der Wahrscheinlichkeit zur Wahl der schlechtesten Alternative (Gleichung 2), so erhält man für die Wahl der B/W-Kombination die folgende Wahrscheinlichkeit:

$$(4) \qquad BW_c(j) = B_c(j) * \sum_{\eta \in R(C\setminus(\mathrm{j,k}))} B_{C\setminus(\mathrm{k})}(\eta_2)...B_{\eta(J-1,k)}(\eta_{J-1})$$

Gleichung 4 wird als das konsistente B/W-Extremwert-Zufallsnutzenmodell bezeichnet, das unter zu Hilfenahme des klassischen MNL-Modells (bzw. durch dessen Einsetzen) geschätzt werden kann.

Marley und Louviere (2005) leiten noch eine Reihe weiterer Wahrscheinlichkeits-Modelle für die Wahl von Best-, Worst- und B/W-Entscheidungen her. Hierbei stützen sich die beiden Autoren allerdings nicht mehr auf die Zufallsnutzentheorie, sondern leiten ihre Modelle anhand eigener, spezifisch entwickelter Theorien zum individuellen Auswahlverhalten her. Aus Platzgründen, und da diese Modelle in der wissenschaftlichen Praxis bisher wenig Anwendung gefunden haben, wird auf eine detaillierte Darstellung im Rahmen dieser Arbeit verzichtet. Interessierte Leser seien auf den grundlegenden Artikel von Marley und Louviere (2005) und auf die Arbeit von Wirth (2010) verwiesen, die beide detailliert die entsprechenden Vorgehensweisen darstellen.

4.2 Design

Bei der Gestaltung der einzelnen Stimuli und deren Verteilung auf die Choice-Sets wird beim B/W-Scaling analog zu den anderen Formen einer DCA vorgegangen. Hierbei ist zuerst zu entscheiden, ob ein s.g. „Full-Factorial"-Design verwendet werden soll oder ein „Fractional Design" (vgl. Aufsatz von Himme in diesem Buch).

„Fractional"-Designs stellen einen Auszug aus der Menge aller möglichen Stimuli dar. Sie ermöglichen es, dass im Gegensatz zu „Full-Factorial-Designs" nicht alle Stimuli miteinander verglichen werden müssen, und verhindern so eine Überforderung der Probanden, die zu einem negativen Einfluss auf die Reliabilität der gewonnenen Ergebnisse führen würde (Johnson und Orme 1996). Bei der Konstruktion fraktioneller

Pläne ist vor allem auf deren Effizienz zu achten. Unter Effizienz wird hierbei die Minimierung der Varianz und Ko-Varianz der Nutzenparameter verstanden (Huber und Zwerina 1996). Zur Erstellung fraktionierter Designpläne finden sich verschiedene Ansätze. Himme (vgl. Aufsatz in diesem Buch) liefert eine ausführliche Übersicht über die verschiedenen manuellen Methoden zur Erstellung fraktionierter Designpläne, so dass hier auf eine weitere Erläuterung verzichtet werden kann.

Neben der manuellen Erstellung von Designs finden sich auch verschiedene Softwaregestützte Verfahren, die die automatische Erstellung von Experimentierplänen ermöglichen. So bietet SPSS mittels Orthoplan eine sehr rudimentäre Umsetzung, die die Erstellung von sehr einfachen Haupteffektplänen ermöglicht. Am weitesten verbreitet ist die Software Sawtooth (Sawtooth 2012), die die Erstellung auch komplexerer Designs mit möglichen Interaktionen zwischen den Variablen erlaubt. Aizaki und Nishimura (2008) präsentieren eine Methode, wie effiziente, fraktionierte Designs mittels der Software R erstellt werden können. Es sei darauf verwiesen, dass - unabhängig von jeder Software - die individuellen Design-Ergebnisse vor Einsatz jeweils auf ihre Effizienz gemäß den Vorgaben von Huber und Zwerina (1996) geprüft werden sollten. Burgess (2007) bietet dafür eine onlinegestützte freie Software. Diese ermittelt, wie sich das gewählte Design auf die Varianz und Kovarianz der zu schätzenden Nutzenparameter auswirkt. Damit wird es möglich, das Design vor der Erhebung auf mögliche Probleme zu überprüfen.

4.3 Stichprobe, Set-Anzahl und Set-Größe

Die Bewertung der Stimuli erfolgt durch die Auswahl der besten und der schlechtesten Alternative. Zur Erzielung reliabler Ergebnisse muss festgelegt werden, wie viele Probanden insgesamt befragt werden, wie viele Choice-Sets jedem Probanden vorgelegt und wie viele Stimuli in einem Choice-Set zusammengefasst werden sollen.

Generell muss bei der Bestimmung der optimalen Stichprobengröße ein Tradeoff zwischen der Minimierung des Messfehlers und der dadurch entstehenden Kosten (wie Zeit, Aufwand oder monetäre Incentivierung) berücksichtigt werden (Orme 2009). Die

Stärke des Messfehlers bei DC Experimenten ist von einer Vielzahl an Faktoren abhängig, wie beispielsweise den Eigenschaften der Grundpopulation, deren Heterogenität oder ob die Auswahl der Probanden zufällig oder willkürlich vorgenommen wird (Johnson und Orme 1996). Ebenso beeinflusst die Frage, ob am Ende der Studie generell nur die Präferenzen aller Teilnehmer gemessen oder ob verschiedene Gruppen anhand ihrer Präferenzen verglichen werden sollen, die Festlegung der finalen Anzahl an zu befragenden Subjekten. Hierbei gilt zu beachten, dass höhere Grade an Heterogenität bzw. eine größere Anzahl an zu vergleichenden Gruppen auch größere Stichproben verlangen. Johnson und Orme (2003) haben für DC-Experimente eine Faustregel erstellt, mit der sich die Mindestanzahl an Teilnehmern in Abhängigkeit von der Anzahl an Attributen und deren Ausprägungen bestimmen lässt. Hierbei gilt:

$$\frac{n * t * a}{c} \geq 500 \quad (5)$$

n gibt die Anzahl an Probanden an, t steht für die Gesamtanzahl an Choice-Sets und a für die Anzahl an Alternativen in einem Choice-Set (ohne die No-Choice-Option). Der Parameter c misst die Komplexität der Stimuli. Wenn nur eine Schätzung der Haupteffekte angestrebt wird, ist c gleich der Anzahl der Ausprägungen des größten Attributs zu setzen. Im Falle einer Schätzung von Haupteffekten und Interaktionseffekten ist c gleich dem Produkt aus der Anzahl an Ausprägungen der zwei größten Attribute. Je nach Kontext kann somit anhand der Gleichung 5 die Mindestanzahl an nötigen Studienteilnehmern bei gegebener Anzahl an Attributen und Ausprägungen bestimmt werden. Ebenso kann - ausgehend von der zur Verfügung stehenden Sample-Größe - bestimmt werden, wie hoch die maximale Anzahl an Attributen oder Ausprägungen sein darf bzw. wie viele Choice-Sets pro Probanden mindestens gezeigt werden müssen.

4.4 Schätzung der Nutzen

Zur Schätzung der Nutzenfunktion von B/W-Scaling Experimenten können zwei unterschiedliche Verfahren verwendet werden. So lassen sich die orthogonal erstellten

B/W-Experimente durch einfaches Auszählen der Wahlhäufigkeiten, der sogenannten Count-Analyse, auswerten (Finn und Louviere 1992). Alternativ können B/W-Experimente aber auch analog zum klassischen Vorgehen bei anderen Formen der DC-Analyse anhand Multinomialer-Logit-Modelle (wie in Abschnitt 4.1 dargestellt) mit Hilfe des ML-Verfahren geschätzt werden.

4.4.1 Count-Analyse

Bei der Count-Analyse wird gezählt, wie oft ein Attribut (Case1) oder seine verschiedenen Ausprägungen (Case2 und Case3) als beste Alternative und als schlechteste Alternative ausgewählt werden. Dies kann sowohl auf aggregiertem Level über alle Studienteilnehmer hinweg als auch auf individueller Ebene pro Probanden geschehen (Finn und Louviere 1992; Cohen 2009). Orme (2009b) betont allerdings, dass jede Ausprägung gleich oft dem Probanden präsentiert werden und mindestens viermal zur Wahl gestanden haben muss, damit reliable Ergebnisse auf individueller Ebene garantiert werden können. Tabelle 5 gibt einen Überblick über die aggregierten Ergebnisse der Case3 Erhebung aus dem Eingangsbeispiel. Die Spalten „best“ und „worst“ geben die Anzahl an, wie oft eine Attributausprägung als beste bzw. schlechteste benannt wurde. Für die Aufbereitung und Berechnung der Scores reicht ein einfaches Tabellenkalkulationsprogramm. Der B/W-Score wird durch Subtraktion der beiden Individual-Scores berechnet. Anhand der B/W-Scores lässt sich schon die Wichtigkeit der einzelnen Attributausprägungen bestimmen und sofort untereinander vergleichen. Um weiter eine Vergleichbarkeit der Ergebnisse zwischen verschiedenen Erhebungen mit unterschiedlichen Anzahlen an Teilnehmern oder unterschiedlichen Designs zu erreichen, müssen die Scores standardisiert werden (Cohen 2009). Hierzu werden die einzelnen B/W-Scores durch das Produkt aus der Anzahl an Probanden und der Häufigkeit des Auftretens der Attributausprägungen geteilt. Die dadurch berechneten „Average Best-Worst-Scores“ können nun unabhängig von Stichprobengröße oder Design zwischen verschiedenen B/W-Experimenten verglichen werden. Eine andere Form, die Wichtigkeit der verschiedenen Attributausprägungen

zu bestimmen, ist die Berechnung einer Verhältnisskala (Auger, Devinney und Louviere 2007).

Tabelle 5: Ergebnisse Menü B/W-Experiment

Attributausprägung	Best	Worst	Best-Worst	Averages B/W	$\sqrt{\frac{B}{W}}$	RI (in Prozent)
Abo-Laufzeit						
3 Monate	146	456	-310	-0,115	0,566	12 %
6 Monate	354	221	133	0,049	1,266	26 %
12 Monate	321	125	196	0,073	1,602	33 %
24 Monate	432	18	414	0,153	4,899	100 %
Media-Form						
Print	80	329	-249	-0,092	0,493	10 %
ePaper	276	213	63	0,023	1,138	23 %
ePaper und Print	252	243	9	0,003	1,018	21 %
Mobile App	376	76	300	0,111	2,224	45 %
Medium						
Tageszeitung	476	45	431	0,160	3,252	66 %
Wochenzeitung	321	187	134	0,050	1,310	27 %
Magazin	221	117	104	0,039	1,374	28 %
Special Interest	376	87	289	0,107	2,079	42 %
Preis / Monat						
5,00 Euro	245	98	147	0,054	1,581	32 %
13,00 Euro	423	101	322	0,119	2,046	42 %
16,00 Euro	331	182	149	0,055	1,349	28 %
18,00 Euro	121	243	-122	-0,045	0,706	14 %
450 Teilnehmer, jede Ausprägung wurde 6 mal gezeigt						

Hierbei wird die Wurzel aus dem Quotienten der Anzahl an Best-Nennungen durch die Anzahl der Worst-Nennungen berechnet. Der Ausprägung mit dem höchsten Wert wird nun der Wert 100 zugeordnet und die Wichtigkeiten der restlichen Attributausprägungen werden nun im Verhältnis zu dieser Ausprägung ausgedrückt. So hat die Ausprägung „12 Monate“ Laufzeit einen Wert von 33%, was sich aus dem Quotienten aus 1,602 und 4,899 ergibt. Weiter kann die Wichtigkeit einzelner

Attribute (hier zum Beispiel Laufzeit im Vergleich zu Preis) analog bestimmt werden, indem die Summen über die verschiedenen Scores berechnet und dann in Relation zueinander gesetzt werden. Da dieser Ansatz auch auf individueller Ebene möglich ist, können die B/W-Scores analog zu den durch eine CBC gewonnenen individuellen Nutzenparametern genommen werden, um unter zu Hilfename einer geeigneten „Choice Rule“ (Skiera und Gensler 2002), eine Marktsimulation zu berechnen.

Im Online-Anhang zu diesem Aufsatz finden sich zwei Beispiel-Files, die zeigen, wie mittels der Tabellenkalkulation Excel und unter zu Hilfenahme von VBA-Makros eine B/W-Studie mit einer beliebigen Anzahl an Probanden durchgeführt werden kann. Die Files erlauben im Falle eines 3x3x3 fraktionierten, orthogonalen Designs die individuellen und aggregierten Teilnutzenwerte mittels des oben gezeigten Berechnungsverfahrens zu ermitteln. Weiter kann dann mittels dieser Werte eine Marktsimulation in Excel erstellt werden. Die Ergebnisse der „Count-Analyse“ weisen eine hohe Korrelation mit den Ergebnissen sowohl einer klassischen MNL-Analyse, den Ergebnissen einer Latent-Class-Analyse, als auch den Ergebnissen einer HB-basierten Schätzung auf und können daher im Falle orthogonaler Designs als ein weniger aufwendiger und schnellerer Ersatz für eine Schätzung mittels MNL betrachtet werden (Marley und Louviere 2005; Orme 2009a). Orme (2009a) zeigt weiter, dass die Prognosegüte anhand einer Holdout-Validierung auf individueller Ebene sich nur gering zwischen den verschiedenen getesteten Schätzverfahren unterscheidet. So sind die Ergebnisse der Count-Analyse (63,0%) annährend deckungsgleich mit den Schätzungen via MNL (63,3%), während sich die Ergebnisse einer Schätzung durch ein hierarchisches Bayes-Modell eine geringfügig bessere Prognosegüte aufweisen (64,2%).

Somit bietet sich die Count-Analyse vor allem an, wenn es darum geht, schnell und kostengünstig eine Schätzung der Präferenzen zu erreichen, ohne dabei auf kostspielige und aufwendige Softwarelösungen zurückgreifen zu müssen.

4.4.2 Conditional Logit

Alternativ zur „Count-Analyse“ können B/W-Experimente auch mittels Multinomial-Logit-Modellen ausgewertet werden. Aufgrund seiner weiten Verbreitung in den gängigen Softwarepaketen ist hierbei das Conditional-Logit-Verfahren (CLM) am gebräuchlichsten. Hierbei kann die eigentliche Schätzung sowohl durch einen „gewichteten kleinste Quadrate (WLS)-Ansatz“ erfolgen als auch klassisch durch einen ML-Ansatz (Flynn, Louviere, Peters und Coast 2007). Louviere und Woodworth (1983) zeigen, dass der WLS-Ansatz im Falle einer Aggregation der Ergebnisse über alle Teilnehmer hinweg bessere Schätzungen liefert, während der ML-Ansatz im Falle einer Untersuchung von Unterschieden zwischen den einzelnen Studienteilnehmern bessere Ergebnisse erbringt (Flynn, Louviere, Peters und Coast 2008). Bei der Datenaufbereitung wird jede bestehende Kombinationsmöglichkeit aus zwei Stimuli in einem Choice-Set bestimmt und via Effekt-Kodierung notiert. Bei der Effekt-Kodierung wird jeder möglichen Attributausprägung eine Spalte zugewiesen. In den Zeilen werden die verschiedenen Kombinationen bzw. Stimuli aufgeführt. Ist eine Ausprägung bei einer Kombination von Stimuli vorhanden, wird ihr eine 1 zugeordnet. Andernfalls erhält sie den Wert 0. Die abhängige Variable des Modells nimmt den Wert 1 an, wenn diese Kombination als B/W-Paar gewählt wurde, und den Wert 0, wenn das Paar nicht gewählt wurde.

Eine genaue Anleitung zur Modellierung und Datenaufbereitung findet sich in Flynn, Louviere, Peters und Coast (2008). Zu beachten bei der Anwendung von CLM ist allerdings, dass die Schätzung individueller Präferenzen für einzelne Studienteilnehmer nicht möglich ist.

4.4.3 Latent Class

Bei LC-Schätzungen werden die Studienteilnehmer anhand ihres Wahlverhaltens in homogene (s.g. latente) Klassen unterteilt. Während bei einem HB-Ansatz (siehe Abschnitt 4.4.4) die Schätzung auf individueller Ebene und erst in einem nachgelagerten Prozess eine Clusterung stattfindet, erfolgt im Falle einer LC-

Schätzung die Ermittlung der segmentspezifischen Präferenzen anhand einer einzigen, integrierten Schätzung. Die Ermittlung der individuellen Teilnutzen erfolgt dann durch die Multiplikation der segmentspezifischen Schätzer mit den Wahrscheinlichkeiten, dass ein Proband einer Klasse zugeordnet wird (Wedel und Kamakura 2000). Zur Schätzung von B/W-Modellen wird das Grund-LC-Modell (vgl. Beitrag Temme in diesem Buch) an die Form eines sequentiellen Wahlprozesses angepasst (Vermunt und Magidson 2005). Die Auswahl der besten Alternative wird dabei klassisch modelliert. Die Wahl der schlechtesten Alternative wird als die Auswahl aus den restlich verbleibenden Alternativen im Choice-Set verstanden, wobei die Verbindung zwischen der Auswahlwahrscheinlichkeit und dem Nutzen der ausgewählten Alternative negativ ist. Dies wird durch die spezielle Einführung eines Gewichtungsfaktors, der den Wert -1 trägt, erreicht. Die Schätzung selbst erfolgt wie üblich über einen ML-Ansatz. Ein Vorteil eines LC-Modells ist, dass im Vergleich zu einer HB-Schätzung im Vorfeld weniger Annahmen getroffen werden müssen. Allerdings ist zu beachten, dass im Falle der LC-Schätzung mehr Beobachtungen vorliegen müssen, was zu einem höheren Bedarf an Studienteilnehmern führt.

4.4.4 Hierarchical Bayes

Die Schätzung von Discrete Choice Modellen mittels HB-Modellen wird aktuell als State-of-the-Art-Methode betrachtet. HB-Modelle bieten den Vorteil, dass sie vor allem bei geringer Antwortzahl pro Probanden immer noch zuverlässige Schätzergebnisse liefern (vgl. hierzu den Beitrag von Temme in diesem Buch). Wirth (2010) als auch Orme (2009a) geben einen guten Überblick über die notwendigen Anpassungen an den in Abschnitt 2 dargestellten Modellen für eine HB-Schätzung und die notwendigen Spezifizierungen der individuellen Likelihood-Funktionen. Das Vorgehen hierbei ähnelt stark dem Vorgehen bei der HB-Schätzung einer CBC. Der einzige, aber essentielle Unterschied liegt in der unterschiedlichen Spezifikation der Likelihood-Funktion, die im Falle des B/W-Experimentes um die zusätzliche Wahl der schlechtesten Alternative erweitert wird. Basierend auf einer Markov-Chain-Monte-Carlo-Simulation erfolgt dann - analog zum Vorgehen bei einer CBC - durch

wechselseitiges Ziehen aus den bedingten Verteilungen die Bestimmung der notwendigen Posteriori-Verteilung (Wirth 2010).

5 Software-Unterstützung

Software zur Erstellung, Erhebung, Aufbereitung und Schätzung von B/W-Experimenten gibt es bisher nur wenige. Während eine Vielzahl an verschiedenen Software-Produkten Lösungen für die einzelnen Schritte einer Erhebung - wie Designerstellung, Erhebung, Schätzung oder eine anschließende Marktsimulation - bieten, liefert aktuell nur Sawtooth einen integrierten Ansatz. Die anderen Produkte unterscheiden sich sowohl in der Vielfalt der Leistungen als auch bezüglich Benutzerfreundlichkeit und Preisniveau (vgl. Tabelle 6). Die Tabellenkalkulation Excel bietet ausreichende Leistung für die Auswertung von B/W-Experimenten mittels der Count-Methode und kann mittels VBA-Makros sogar in Teilen für eine rudimentäre Form der Datenerhebung und Datenaufbereitung verwendet werden. Ebenso kann mittels Excel anhand der Schätzergebnisse recht einfach eine Marktsimulation programmiert werden. Auch wenn man mit Excel mittels Add-Ins generell eine MNL- und LC-Schätzung durchführen kann, sollte beachtet werden, dass die den Add-Ins zugrunde liegend2en Modelle nicht an die Besonderheiten des B/W-Scalings angepasst werden können. Statistikprogramme wie SPSS, Stata, SAS und Limdep ermöglichen eine Schätzung via Conditional Logit-Ansatz und bieten den Vorteil, dass sie im Gegensatz zu Excel und anderen Tabellenkalkulationsprogrammen einen effizienteren Umgang mit größeren Datenmengen garantieren. Hierbei ist zu beachten, dass SPSS und Stata durch die übersichtlicheren Programmoberflächen dem unerfahreneren Anwender einen schnelleren Zugang und eine einfachere Bearbeitung ermöglichen. SPSS bietet dazu dem Anwender durch die Orthoplan-Funktion die Möglichkeit, - wenn auch nur sehr rudimentäre - Designpläne zu erstellen.

Tabelle 6: Softwareprogramme zur Durchführung von B/W-Experimenten

				Schätzverfahren			
Software	*Design*	*Erhebung*	*Aufbe-reitung*	*Count-Analyse*	*HB*	*LC*	*CL*
Excel	---	--+	+++	+++	---	--+	---
SPSS	--+	---	-++	-++	---	---	+++
SAS	---	---	--+	-++	---	---	+++
Stata	---	---	-++	-++	---	---	+++
Limdep	---	---	--+	--+	+++	---	+++
Latent Gold Choice	---	---	---	---	---	+++	---
Matlab	--+	---	---	--+	+++	---	+++
R	-++	--+	--+	-++	+++	--+	+++
Sawtooth	+++	+++	--+	--+	-++	--+	-++
DISE	---	+++	-++	--+	---	---	---
Globalpark	--+	+++	--+	---	---	---	---
Lime Survey	---	+++	--+	--+	---	---	---
--- = nicht unterstützt, --+ = bedingt möglich -++ = möglich +++ = komfortable Implementierung							

Zu beachten bleibt allerdings, dass keines der Programme die Erhebung der Daten in Form eines Fragebogens oder die Erstellung einer anschließenden Marktsimulation ermöglicht.

Spezialsoftware wie Latent Gold Choice und Sawtooth bieten die Möglichkeit, mittels bequemer „Click-and-Go“-Menüs Modelle zu spezifizieren und zu schätzen. Beide Produkte bieten darüber hinaus umfangreiche Online-Tutorials mit Beispielfiles speziell für B/W-Experimente an, die dem Anwender den Einstieg und die Datenaufbereitung erleichtern. Während Latent Gold Choice alleine die Schätzung von LC-Modellen erlaubt, bietet Sawtooth die Möglichkeit, verschiedene Modelle zur Schätzung zu verwenden. Ebenso bietet Sawtooth einen Designgenerator und ein Tool zur Programmierung von Fragebogen und Marktsimulation an. Matrixbasierte Softwarelösungen wie Matlab oder R ermöglichen die Nutzung aller hier vorgestellten Schätzverfahren. Allerdings gilt zu beachten, dass für beide Programme eine hohe

Erfahrung vorausgesetzt werden muss oder eine lange Einarbeitungszeit von Nöten ist. R bietet darüber hinaus erfahrenen Nutzern die Möglichkeit, Designpläne zu erstellen sowie einen Online-Fragebogen zu erstellen, der eine Live-Auswertung der Experimente ermöglicht. Beide Tools können auch mit den gängigen Umfragetools wie Globalpark und Lime Survey gekoppelt werden. Besonders zu beachten ist die an der Universität Frankfurt entwickelte Plattform „DISE“ (Schlereth 2011), die eine spezielle Basis zur Programmierung von B/W-Experimenten beinhaltet und den direkten Export der Daten in Matlab unterstützt.

6 Literatur

Aizaki, Hideo and Kazushi Nishimura (2008): Design and Analysis of Choice Experiments Using R: A Brief Introdction, *Agricultural Information Research*, 17 (2): 86-94.

Auger, Pat, Timothy Devinney and Jordan Louviere (2007): Using Best–Worst Scaling Methodology to Investigate Consumer Ethical Beliefs Across Countries, *Journal of Business Ethics*, 70 (3): 299-326.

Bacon, Lynd, Peter Lenk, Katya Seryakova and Ellen Veccia (2008): Comparing Apples to Oranges, *Marketing Research*, 18 (1): 29-34.

Baumgartner, Hans and Jan-Benedict E. M. Steenkamp (2001): Response Styles in Marketing Research: A Cross-National Investigation, *Journal of Marketing Research*, 38 (2): 143-156.

Burgess, Leonie (2007): *Discrete Choice Experiments*, Source: http://crsu.science.uts.edu.au/choice/.

Chrzan, Keith and Natalia Golovashkina (2006): An Empirical Test of Six Stated Importance Measures, *International Journal of Market Research*, 48 (6): 717-740.

Cohen, Eli (2009): Applying Best-Worst Scaling to Wine Marketing, *International Journal of Wine Business Research*, 21 (1): 8-23.

Cohen, Steve and Bryan Orme (2004): What's Your Preference?, *Marketing Research*, 16 (Summer): 32-37.

Finn, Adam and Jordan Louviere (1992): Determining the Appropriate Response to Evidence of Public Concern: The Case of Food Safety, *Journal of Public Policy & Marketing*, 11 (1): 12-25.

Flynn, Terry N., Jordan Louviere, Tim J. Peter and Joanna Coast (2007): Best-Worst Scaling: What It Can Do For Health Care Research and How To Do It, *Journal of Health Economics*, 26: 171-189.

Flynn, Terry N., Jordan Louviere, Tim J. Peters and Joanna Coast (2008): Estimating Preferences for a Dermatology Consultation Using Best-Worst Scaling: Comparison of Various Methods of Analysis, *BMC Medical Research Methodology*, 76 (8): 1-12.

Garver, Michael, Williams Zachary and Stephen A. LeMay (2010): Measuring the Importance of Attributes in Logistics Research, *The International Journal of Logistics Management*, 21 (1): 22-44.

Huber, Joel and Klaus Zwerina (1996): The Importance of Utility Balance in Efficient Choice Designs, *Journal of Marketing Research*, 33 (3): 307-317.

Johnson, Rich M. and Bryan K. Orme (1996): *How Many Questions Should You Ask in Choice-Based Conjoint Studies*?, Source: http://www.sawtoothsoftware.com/download/techpap/howmanyq.pdf.

Johnson, Rich and Bryan K. Orme (2003): *Getting the Most from CBC*, Source: http://www.sawtoothsoftware.com/education/techpap.shtml.

Lee, Julie Anne, Geoffrey N. Soutar and Jordan Louviere (2008): The Best-Worst Scaling Approach: An Alternative to Schwartz's Values Survey, *Journal of Personality Assessment*, 90 (4): 335-347.

Lee, Julie Anne, Geoffrey N. Soutar and Jordan Louviere (2007): Measuring Values Using Best-Worst Scaling, *Psychology & Marketing*, 24 (12): 1043-1058.

Louviere, Jordan, David Hensher, A. and Joffre Swait, D. (2000): *Stated Choice Methods: Analysis and Application*, University Press: Cambrigde.

Louviere, Jordan and Towhidul Islam (2008): A Comparison of Importance Weights and Willingness-to-Pay Measures Derived from Choice-Based Conjoint, Constant Sum Scales and Best-Worst Scaling, *Journal of Business Research*, 61: 903-911.

Louviere, Jordan, Deborah Street, Leonie Burgess, Nada Wasi, Towhidul Islam and Anthony A. J. Marley (2008): Modeling the Choices of Individual Decision-Makers by Combining Efficient Choice Experiment Designs with extra Preference Information, *Journal of Choice Modelling*, 1 (1): 128-163.

Louviere, Jordan and Georg Woodworth (1983): Design and Analysis of Simulated Consumer Choice or Allocation Experiments: An Approach Based on Aggregate Data, *Journal of Marketing Research*, 20 (2): 350-367.

Marley, Anthony, A. J. and Jordan Louviere (2005): Some probabilistic models of best, worst, and best-worst choices, *Journal of Mathematical Psychology*, 49 (6): 464-480.

McFadden, Daniel (1974): Conditional Logit Analysis of Qualitative Choice Behavior, in: P. Zarembka (ed.): *Frontiers in Econometrics*, Academic Press, New York, 105-142.

Meade, Adam, W. (2004): Psychometric Problems and Issues Involved with Creating and Using Ipsative Measures for Selection, *Journal of Occupational and Organizational Psychology*, 77 (3): 531-551.

Orme, B. (2009a): *MaxDiff Analysis: Simple Counting, Individual-Level Logit, and HB*, Source: http://www.sawtoothsoftware.com/education/techpap.shtml.

Orme, Bryan, K. (2009b): *Getting Started with Conjoint Analysis: Strategies for Product Design and Pricing Research*, Research Publishers, LLC: Chicago.

Paulhus, Delroy. L. (1991): Measurement and Control of Response Bias, in: John P. Robinson, Philip R. Shaver and Lawrence S. Wrightsman (eds.): *Measures of Personality and Social Psychological Attitudes* Academic, New York, 17-59.

Potoglou, Dimitris, Peter Burge, Terry N. Flynn, Ann Netten, Juliette Malley, Julien Forder and John Brazier (2011): Best-Worst Scaling vs. Discrete Choice Experiments: An Empirical Comparison Using Social Care Data, *Social Science & Medicine*, 72 (11): 1717-1727.

Sawtooth, Software (2012): *Sawtooth MaxDiff Analyzer Program*, Source: http://www.sawtoothsoftware.com/products/maxdiff/maxdiff.shtml.

Schlereth, Christian (2011): *DISE - Dynamic Intelligent Survey Engine*, Source: www.dise-online.de.

Skiera, Bernd and Sonja Gensler (2002): Berechnung von Nutzenfunktionen und Marktsimulationen mit Hilfe der Conjoint-Analyse (Teil II), *Wirtschaftswissenschaftliches Studium (WiSt)*, 31 (5): 258-263.

Tavares, Susana, Margarida Cardoso and José Dias, G. (2010): The Heterogeneous Best-Worst Choice Method in Market Research, *International Journal of Market Research*, 52 (4): 533-546.

Thurstone, Louis (1927): A Law of Comparative Judgement, *Psychological Review*,

34 (4): 278-286.

Vermunt, Jeroen, K. and Jay Magidson (2005): *Technical Guide for Latent GOLD 4.0: Basic and Advanced*, Statistical Innovatons Inc: Belmont Massachusetts.

Wedel, Michel and Wagner A. Kamakura (2000): *Market Segmentation: Conceptual and Methodological Foundations*, Kluwer Academic: Boston.

Wirth, Ralph (2010): *Best-Worst Choice-Based Conjoint-Analyse: Eine neue Variante der wahlbasierten Conjoint-Analyse*, Tectum: Marburg.

B Communication Behavior of Companies In Product Recalls Without Customer Identification Information

Raoul V. Kübler and Sönke Albers

Published in:

Kübler, Raoul V. and Sönke Albers (2010): Communication Behavior of Companies In Product Recalls Without Customer Identification Information, Marketing - Journal of Research and Management, 6 (1), 19-30.

Abstract: Although companies that sell products such as automobiles can easily track down their clients, companies that produce consumer goods usually lose direct contact with their customers at the retailer's cash register. This anonymity results in the obligation to communicate any product failure and/or recall action to a wide range of anonymous and potential customers, while running the risk of scaring off future buyers. Hypothesizing that companies know how to best communicate a recall, we investigate a sample of 104 German recall campaigns to identify common dimensions of recall messages and behavioral drivers that may minimize the risk of alienating future customers and reduce potential sales losses due to the product crisis. We identify six major underlying dimensions of recall messages and find empirical evidence for existing recall routines of companies depending on the degree of hazard and the probability of a product defect.

Keywords: Product Recall, Product Crisis, Communication, Retail Industry, Consumer Goods, Content Analysis, 3SLS

1 Introduction

Given the ever-increasing competitive pressure for innovation as well as continued improvement and development (e.g., Trommsdorff and Steinhoff 2007), companies are forced to introduce new products into the market in ever-shorter innovation cycles. This acceleration has led to an increased risk of introducing new products that have not been sufficiently engineered, which may involve increased risk for the consumer. This is supported by the observation that the number of recalls of defective products has increased significantly since the mid-nineties (Beamish and Bapuji 2008). With the German Appliances and Product Safety Law (Geräte- und Produktsicherheitsgesetz, GPSG) coming into effect in 2004, and the implementation of the European guideline 2001/95/EG with respect to product safety, we find an increased necessity for producers and retailers to recall dangerous or damaging products as soon as possible, as well as to identify and eliminate sold units (GPSG 2004 in combination with Felcher 2003 and Kersten 1992). The same is the case for American producers and retailers as recall rates in the US market steadily raise. According to Beamish and Bapuji (2008), toy recalls alone significantly increased from 17 in 2004 to 22 in 2007 in the US. This boost resulted in 22 toy-related deaths and an estimated 220,500 toy-related injuries in 2006. The US Consumer Protection and Safety Commission (CPSC) estimates that the yearly overall costs due to recalls and product failures in the US today are approaching 357 billion US Dollars (CPSC, 2007).

Products that have been sold directly by the company or that have been registered either by the producer or, in the case of the automobile industry, by a central authority can easily be traced. In such cases, recalling is rather simple and is now often accepted as normal business. Much more difficult is the situation where the customers cannot be traced, which happens with durables such as toys or apparel; this inability to track customers occurs because these products are sold by retailers who do not record the names of their customers. As a consequence, the defect of the product and the resulting recall action has to be communicated through public channels. This public communication leads to a problem for the manufacturer that not only customers

affected, but unaffected customers as well will be alienated. In the worst-case scenario, current non-customers will be deterred from purchase. Thus, companies face the question of how recalls should be designed and communicated in general to avoid possible damage to the companies' image. In addition, companies are interested in the context factors influencing the decision to start a recall.

The extant literature focuses primarily on responses by consumers. Dawar and Pillutla (2000) provide experimental evidence that recalls will generally decrease brand equity and thus cause decreases in sales. These findings also lead to a direct relationship between the value of company shares and recalls (Chu, Lin, and Prather 2005). Thus, companies face a high risk when they are involved in a recall.

Moreover, VanHeerde, DeKimpe and Helsen (2007) use a panel data set to demonstrate that advertising and price elasticities will be negatively affected by a recall. As a consequence, during a product crisis marketing must deal with a special situation in which communication routines are no longer effective or work only weakly. In particular, the company faces the risk that its messages will be ignored or perhaps even misunderstood. In such a situation, companies have to plan very carefully how to communicate the reason for their recall and how to carry out its execution.

Despite the high risks identified in previous research, research in business has not dealt with the question of how communication content can be used to minimize damage to the company's public image. The same is true for the question of how companies should design the execution of the recall to retain aggrieved customers despite the product's defects. Furthermore, our knowledge is very restricted regarding how context factors such as the degree of the product defect or the observed consequences of the faulty product impact customer behavior.

To find answers to our questions, we content-analyze actual cases of recall communications of German products distributed through retailers for which no customer identification is available. The goal of this study is to describe the behavior

of companies with respect to their communication and execution of recalls, and to analyze whether consistent behavior has been established across similar recall situations. With this analysis, we hope to identify types of content in recall messages that will minimize risks and losses. Furthermore, identified situation-specific recall strategies may help companies better address their clients and the public in the case of future recalls.

The following article is organized as follows: After a discussion of the literature and a derivation of hypotheses on the communication behavior, we describe our empirical classification of the observed behavior of companies. In the next section, we detail how we relate this behavior to context factors and provide a discussion of our results. We close with a summary, followed by the implications for research and practice as well as some limitations.

2 Research question and derivation of hypotheses

2.1 Derivation of the research question

Product recalls represent a special form of product crisis in which products are taken from the market because of defects or the potential to cause damage to consumers or third parties (Felcher 2003, Kersten 1992). Most national laws – such as the European Guidelines for Consumer Protection or the US Consumer Product Safety Act (CPSA) – specify four reasons for activating product recalls: a) a substantial potential of damage for the user of a product, b) the use of material that has a toxic or detrimental effect on consumers in the short or long run, c) the risk of serious injuries or even death by improper use of the product, and d) a violation of current product safety guidelines regarding upper limits with respect to noise and environmental pollution (Chu, Lin and Prather 2005, GPSG 2004 and CPSA, 1972). As stated in the introduction, products that have been distributed through retailers such as toys, appliances or apparel present an extremely complicated recall situation. Aside from the losses caused by reduced sales, companies face the additional costs associated with communication via mass media and possible punishment by authorities, as well as the

cost of coordinating and executing the recall (Beamish and Bapuji 2008, Bapuji, Standop 2006, Konken 2002, Standop 1995, Standop 1993, and Folkes 1984). Recently, the punishment has become even harsher: the Consumer Product Safety Improvement Act in 2008 increased the possible fines and even specified jail time for some violations for US manufacturers (CPSC, 2008). This new approach on the part of the US administration may be underlined by the example of RC2, a manufacturer of wooden toys who imported lead-poisoned toy trains into the US and has recently been charged a civil penalty of 1.25 million US Dollars. In addition to the risk of fines, one should not forget the high risks of private lawsuits that firms in the US market face when harming customers.

In addition to these problems, any product recall represents further risks for the company: recalls negatively influence the brand's image (Mowen, Jolly and Nickell 1981, Jolly and Mowen 1985, Siomkos and Shrivastava 1993, Rhee and Haunschild 2006 and DeMatos and Rossi 2006), its customer loyalty (Cleeren/DeKimpe/Helsen 2008) and its brand equity (Dawar and Pillutla 2000), along with the company's stock market value (Chu, Lin and Prather 2005). Table 1 gives an overview of recent studies on the effects of recalls on companies' market performance. Product recalls will generally be communicated by the affected company through press releases or advertisements in mass media channels. This material is the cornerstone for the press and the public in dealing with a product recall. It is therefore reasonable to assume that any communication about product recalls contains all relevant information pertaining to the recall, which is why we use such announcements as our basis for analyzing how companies let the public know about product recalls. Given the potential losses companies face from initiating a product recall, it can be assumed that they will attempt to communicate the product recall in such a way as to minimize the unwanted effects (Konken 2002, Cezanne 1999 and Haller 1998). This implies that companies will avoid a standard product recall form, and instead will have developed certain routines – depending on the type of product, who the customers are, the brand image and the defect – to communicate and execute the recall. This leads to the following

research questions that, to the best of our knowledge, have yet to be answered by empirical or theoretical studies:

1. What kinds of communication content and forms of product recall execution do companies use to minimize negative effects?
2. Which context factors (situational variables) play an important role in influencing the behavior of affected companies?

Table 1: Recent studies of recall effects on firm or marketing performance

Authors	Object of Study	Sample and Method	Endogenous Variable	Major results
Cleeren, Dekimpe, and Helsen (2008)	Impact of brand reputation on rebuy behavior after product recall	Hazard analysis of panel data for 3 Australian peanut butter brands over 36 months	First rebuy after product recall	- Significant influence of brand on rebuy time - Significant decrease in brand loyalty - Significant negative impact on advertising elasticity of non-branded product
Van Heerde, Dekimpe, and Helsen (2007)	Impact of product recalls on marketing effectiveness	Time-series analysis of panel data for 3 Australian peanut butter brands over 36 months	Elasticities of the used marketing mix variables	- Significant negative impact on advertising elasticity - Significant negative impact on price elasticity - Significant negative cross effects
Chu, Lin, and Prather (2005)	Impact of product recalls on stock value	218 recall-affected producers of consumer goods (1984-2003)	Abnormal returns	- Significant negative impact on stock value - Greater reactions for pharmaceuticals and cosmetics - Greater reactions for toy-producing companies
Dawar and Pillutla (2000)	Impact of product recalls on brand equity	Survey of 218 coffee consumers and laboratory experiments with a student sample	Change of brand perception	- Significant negative impact on brand equity
Folkes (1984)	Consumer reactions to product failures	Correlation analysis of experiments with n= 61 students	Demand for punishment of the affected company	- Confession of guilt by producer positively connected with the demand for compensation by customers - Confession of guilt by producer is positively connected with the demand for punishment

2.2 Derivation of Hypotheses

National laws such as the German Appliances and Product Safety Law (GPSG), as well as various guidelines for product safety established by the European Union, do not require companies to follow specific procedures in communicating the recall of defective products; some freedom is provided for companies to design their own individual product recall and its communication. We can further assume that companies will make use of these degrees of freedom and do everything to minimize the recall's negative effects.

Some studies have shown that a company's acknowledgment of responsibility for a product defect generates a strong desire among consumers, not only for the company's punishment but for compensation for inflicted damages as well (Folkes 1984 and Coombs 1998). Thus, we hypothesize that companies will attempt to cloud the question of guilt whenever possible. Studies have also found that the communication of social responsibility and – when true – an emphasis on the voluntary nature of the product recall, in addition to immediate action by the company to address the defect, also have a positive effect (DeMatos and Rossi 2006, Standop 1996, Siomkos and Shrivastava 1993 and Jolly and Mowen 1985).

Thus, we expect that product recalls communicate social responsibility in different ways and with various magnitudes. From the results by Jolly and Mowen (1983) and Standop (2006), it is also reasonable to hypothesize that companies vary the amount of information as well as the degree of detail they provide when communicating product recalls. This is done by either withholding or relativizing certain information; this practice also serves the purpose of countering possible image damage by altering the degree of transparency toward the public (Standop 2006 and Jolly and Mowen 1985). It is likely that companies even attempt to improve their image by designing a convenient and hassle-free product recall for their customers rather than simply conducting damage control with good communication regarding the product recall. Thus, product recall announcements will vary with respect to the transparency of communication and the degree of convenience for the affected customers.

Furthermore, the study by Mowen, Jolly, and Nickell (1981) shows a negative impact on company image caused by the length of time between the discovery of a defect and the beginning of communication. This study finds that the longer the company waits to begin communication, the greater the loss of image. We can also assume that customers expect a high degree of transparency and honesty (Standop 2006), so companies must behave accordingly. This effect implies that, although a certain degree of freedom is allowed for communicating a product recall, specific situations and the expectations of customers frequently reduce the amount of freedom a company has.

Finally, we have to be aware – as explained in chapter 2.1 – that companies are legally not completely free in how they communicate product defects in a recall. According to the German implementation of the guidelines of the European Union for consumer protection in amended Paragraph 10 of the German Appliances and Product Safety Law (GPSG 2004), if companies, or the concerned distributor, do not communicate the defect, or do not do so in an appropriate fashion, authorities can take over communication about the product recall and inform consumers directly. According to the CPSC, the same is the case for the US market, where affected retailers or manufacturers are forced by law to at least inform the official safety commission about a product defect. In the case of non-cooperation or communication of insufficient information, the CPSC is allowed to communicate the product defect on its own. Of course, it is unlikely that the authorities will take into account the effects of full disclosure regarding a recall on the image of the company. Thus, this regulation serves as a threat, making it in a company's best interest to communicate the product defect, thereby giving the company the opportunity to design the communication with minimal damaging effects. In fact, it has been empirically shown that consumers view product recalls more positively when they are announced by the affected company than when they are communicated through public agencies (Mowen 1979 and Mowen, Jolly, and Nickell 1981). Therefore, a company will design the communication of a product recall so that it follows the guidelines set forth by the public authorities but at the same time minimizes possible damage to the company's image.

Knowing how companies can design their product recall communication poses the question of how this design depends on context. Obviously, the communication may rely on the degree of hazard posed by a product defect and the probability of its occurrence. We pointed out previously that consumers tend to want to see some kind of punishment for companies recalling defective products. This desire may lead the company to employ a strategy of customer appeasement in cases of a low defect probability, an approach that tries to pacify consumers by claims about a company's social responsibility or by relativizing the defect. However, companies must not neglect their responsibility to minimize continued risks to consumers in the face of strict national laws, which will seriously punish any producer or retailer if further accidents occur. Moreover, each additional accident caused by the defective product will likely arouse increasing public interest and media pressure. So, companies will engage in a strategy that relativizes product defects and emphasizes their social responsibility only if the potential damage to consumers is small. Similarly, we can also assume that the transparency of communication will be higher when the potential damage to the consumer is higher and less transparent when risk of injury is lower.

These conclusions lead to the following hypotheses:

> H1a: *The lower the probability of a possible damage (hazard) caused by a product defect, the stronger a communication of social responsibility and offered transparency will take place.*
>
> H1b: *The higher the degree of a possible damage (hazard) caused by a product defect, the more transparent the communicated product recall will be.*
>
> H1c: *The higher the degree of a possible damage (hazard) caused by a product defect, the greater the emphasis will be placed on social responsibility.*

Communicating social responsibility or providing more information for achieving high transparency automatically leads to longer recall messages. We explain this by the fact

that companies will use more space for justification and presentation of the companies' own image. This leads to the following hypotheses:

> H2a: *The smaller the probability of a possible damage caused by a product defect, the more detailed and longer the product recall communication will be.*

> H2b: *The higher the degree of a possible damage caused by a product defect, the more detailed and longer the product recall communication will be.*

The early studies by Mowen (1979) and Mowen, Jolly, and Nickell (1981) suggest that the perception of a product recall is also influenced by the customer's expense, effort and discomfort caused by returning the product. In particular, the greater the convenience of returning the product is, the smaller the final damage to the image of the defective product will be. Of course, the higher the degree of hazard caused by the product defect and the greater the probability of hazard, the more likely it is that companies will offer convenient product return procedures. This means that in case of low risk of damage, a company will ask customers to return the product themselves, whereas in more dangerous situations, a company will offer more convenient ways to return the product. This leads to the following hypothesis:

> H3: *The higher the probability or the degree of damage by a product defect, the higher is the degree of convenience offered for returning the product.*

In general, product recalls do not affect the entire production of a product but rather smaller lots. In this case, companies communicate information that allows customers to distinguish between affected and unaffected products; this enables customers to identify dangerous products immediately, thus minimizing the risk of possible damages. At the same time, such a communication allows companies to emphasize the general safety of the product, with only specific lots being unfortunately affected. This leads to the following hypothesis:

H4: *The higher the degree of hazard or its probability, the more detailed the information communicated for consumer identification of defective products will be.*

Several studies provide evidence that brand awareness significantly influences the perception of product recall (Mowen, Jolly, and Nickell 1981, Jolly and Mowen 1985, Siomkos and Shrivastava 1993, Standop 1996, Rhee and Haunschild 2006, DeMatos and Rossi 2006). However, the study results are ambiguous with respect to the direction of that effect. While the Anglo-American studies by Mowen, Jolly and Nickell (1981), Jolly and Mowen (1985), Rhee and Haunschild (2006) and DeMatos and Ross (2006) find a negative influence on the perception, Standop (1996) finds empirical support that a positive effect exists in the case of German consumers. Taking this German study, we can suppose that companies with branded products will counter possible image losses by placing a strong emphasis on social responsibility and by providing a high level of communication transparency, better identification of defect products and greater convenience of product return. Thus, we can derive the following hypotheses:

H5a: *Companies with branded products communicate a higher degree of social responsibility compared to companies with non-branded products.*

H5b: *Companies with branded products will make their communication more transparent than companies with non-branded products.*

H5c: *Companies with branded products will communicate their product recall in a more detailed way than companies with non-branded products.*

H5d: *Companies with branded products will design the product return for consumers to be more convenient than companies with non-branded products.*

H5e: *Companies with branded products will provide more help in identifying defective products than companies with non-branded products.*

Finally, we note that the product-involvement of consumers has a substantial influence regarding their reception of a product recall (DeMatos and Rossi 2006). This study concludes that disappointment and anger about the product recall are higher for products with high involvement than for those with less involvement. The authors assume that companies will take this into account when designing their product recall. This leads to the following hypotheses:

> H6a: *Companies will communicate a higher degree of social responsibility for products with high involvement of consumers than for those with low involvement.*
>
> H6b: *Companies will communicate in a more transparent way the product recall for products with high involvement of consumers than for those with low involvement.*
>
> H6c: *Companies design the communication of a product recall in a more detailed way for products with high involvement of consumers than for products with low involvement.*
>
> H6d: *Companies design the product return to be more convenient for consumers of products with high involvement of consumers than for low-involvement products.*
>
> H6e: *Companies provide more assistance in identifying defective products with high involvement of consumers than for those with low involvement.*

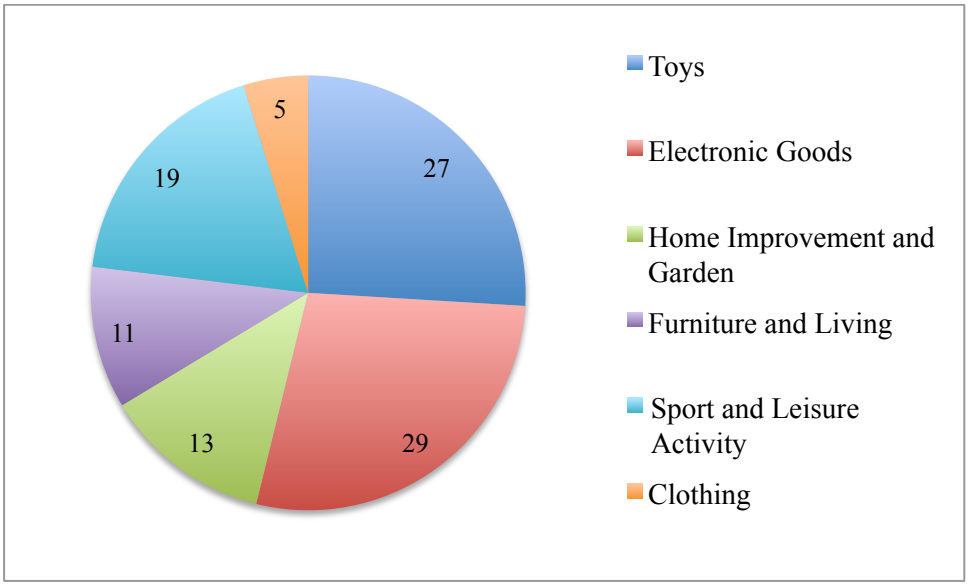
To test these hypotheses, we conducted an empirical study consisting of two parts. The first part aims to identify different content dimensions and corresponding levels in recall messages to deliver a database for the second part. The second part of the study uses this knowledge to test the influence and the moderating effect of context and product factors on the company's choice of the dimension levels according to our hypotheses. In the first part of the study, we content-analyze the sample of product

recalls for identifying dimensions and corresponding levels. In preparation for the second part of the study, we coded the sample of product recalls with respect to the identified content dimensions and levels. The second part of the study then tests the hypotheses with respect to a possible relationship between the coded recall characteristics and moderating effects.

3 Content analysis of product recalls

The scope of the first part of the study is to identify all typical content dimensions and levels that companies use in a product recall to deliver a specific database for the second part of the study. We assume that companies choose specific communication styles and content to minimize possible damage to their image or sales losses. Based on a sample of 104 product recalls, we identify their communication dimensions and individual levels.

Figure 1: Recall Frequencies by Product Categories in the Sample



3.1 Sample and data

Product recalls are generally communicated by the affected companies through press releases and by placing advertisements in mass media. This information also serves as

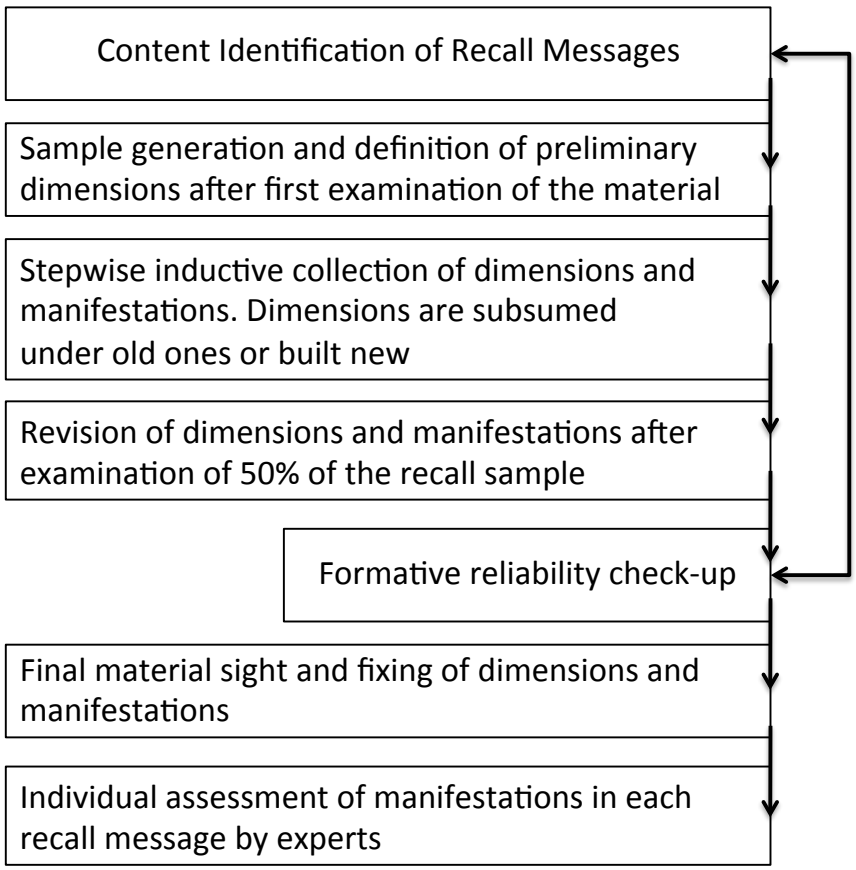
the basis for further press reports. We assume that these press releases and advertisements contain all the pertinent information that a company wishes to communicate to the consumer. To investigate the content of product recalls, we collected the information of all product recalls provided by the companies producing products whose consumers were not traceable between January 1, 2008, and October 1, 2009. We searched the Rapid Alert System for Non-Food (RAPEX) database (Rapex 2009) of the European Union, websites and databases of various consumer protection agencies of the German states (e.g., Bundesverband der Verbraucherzentralen), websites of several independent product testing foundations (e.g., Stiftung Warentest) and a few independent and private consumer portals such as www.produktrueckrufe.de. In total, we identified 104 product recalls coming from 51 different companies. Figure 1 presents an overview of the product categories and the number of product recalls in the respective categories. It is clear that the sample mainly consists of toys and appliances, which coincides with the observation by Beamish and Bapuji (2008) that product recalls are most frequently found in these categories.

3.2 Methodology

To find the content dimensions and levels as well as the communication patterns and moderating context factors, we applied a content analysis method to the texts of the product recalls. This method is especially suitable for pilot studies (Mayring 2008). Following the rules of content analysis (Krippendorff 1980, Gerbner 1969) – as depicted in Figure 2 – we analyzed the product recall messages in eight steps (Mayring 2000). We first looked for any possible dimensions of communication by inductively screening all recall messages. In a second step, we tried in an inductive way, again, to identify all possible occurrences of individual dimension levels in all recall messages. In the next step, we attempted to classify each occurrence into dimensions and then analyzed which levels per dimension are present in the various texts. After having conducted a categorization of 50% of the data material, we revised the definition of the communication dimensions and its levels according to the suggestion by Mayring

(2000, 2008). In the end, we were able to identify six content dimensions, which are listed in Table 2. According to the same inductive process, we then determined the levels per dimension for a subsequent coding of the product recall messages.

Figure 2: Qualitative Content Analysis (Source: based on Mayring, 2008)



3.3 Content dimensions and levels

As mentioned, we identified six content dimensions. Table 2 provides an overview of the dimensions and their corresponding levels.

We find that companies place the strongest emphasis on social responsibility with respect to the customer. The communication of social responsibility ranges from levels such as “defect despite quality control” to very detailed descriptions of social responsibility. We also find evidence that companies vary the degree of social responsibility in their communication.

Moreover, we discover that product recalls vary with respect to the degree of the level of convenience for the product return. In this dimension, we identify six levels ranging from “disposal at home” to “repair or exchange at client’s home.”

Product recalls often provide information on how to identify defective products in connection to the degree of hazard. Thus, we could identify a third dimension, “Help with identification.” The dimension contains six levels, ranging from “mentioning the product name” to the statement that “a call center will help with individual identification.”

All product recalls note the original discoverer of the product defect. Therefore, the fourth identified dimension was the source of the recall. Inductively, we derived six levels for this dimension, ranging from “in-house during quality controls” to “more than ten customers.” The fifth dimension is “transparency of communication.” The communication can be categorized in the following levels: “problems are identified, but not sufficiently,” “problem will be technically explained without discussing the outcome for the customer,” and “hazard is clearly communicated without being relativized.”

Following social responsibility, companies generally emphasize the fact that they are voluntarily engaging in a product recall. Despite many vague descriptions, we were able to determine three levels for the dimension “voluntariness of recall”: “own decision,” “because of public pressure,” and “by law.”

It is not surprising that product recalls provide information on the expected degree of hazard and its probability. We were able to find different levels of hazard in the product recall communication, ranging from “appliance not working” to “mortal danger.”

Table 2: Identified content dimensions and corresponding levels

Social Responsibility (Soc.Respon.)
1 = Defect occurred in spite of quality control 2 = Defect occurred in spite of quality control and high safety standards 3 = Defect occurred in spite of quality control and high safety standards, stressing that customer safety comes first 4 = Defect occurred in spite of quality control and high safety standards, stressing that customer safety comes first and that no customer has been harmed yet 5 = Defect occurred in spite of quality control and high safety standards, stressing that customer safety comes first, that no customer has been harmed yet and that recall is done voluntarily 6 = Defect occurred in spite of quality control and high safety standards, stressing that customer safety comes first, that no customer has been harmed yet and that recall is done voluntarily. Relativization of the defect
Convenience of Product Return (Con-Prod-Ret)
1 = Product can been disposed without any compensation 2 = Return in retail store with compensation in cash 3 = Return by mail with compensation by bank transfer 4 = Repair or exchange in retail store 5 = Repair or exchange by mail 6 = Repair or exchange at customer's home
Help with Identification (HwI)
1 = Only the name of the product is mentioned 2 = Name of the product and photo of affected product are shown 3 = Name of the product, photo of affected product and retail price are shown 4 = Name of the product, photo of affected product, retail price and batch or serial number are shown. 5 = Name of the product, photo of affected product, retail price and batch or serial number are shown. In addition, message contains guidance for identification. 6 = Name of the product, photo of affected product, retail price and batch or serial number are shown. In addition, message contains guidance for identification and a customer service number to call.
Source of Recall (SoR)
1 = Defect identification by in-house labs 2 = Defect identification by independent labs 3 = Defect identification by official authorities 4 = Defect identification by client without consequences 5 = Defect identification by fewer than 10 clients with serious consequences 6 = Defect identification by more than 10 clients with serious consequences
Transparency of communication (Transp.)
1 = Problems are addressed without any concretization 2 = Technical problems are discussed without taking on possible consequences for consumers 3 = Problems are explained but the possible consequences are given under relativization 4 = Problems are explained with all possible consequences without any relativization
Voluntariness of recall (VoR)
1 = By self-motivation 2 = By public or media pressure 3 = By legal statement

4 Communication behavior and moderators

With the first part of the study, we show that the information on product recalls uses a common set of building blocks. We also find evidence that companies substantially vary the levels of content dimensions. In the second part of the study, we investigate whether companies more or less arbitrarily choose a certain product recall communication or whether they purposely react to context factors.

4.1 Sample and operationalization

This second part of the study is based on the same sample as the first part. In the first part, we identified six content dimensions and their corresponding levels. Based on this operationalization of levels, five consumers between the ages of 21 and 70 with a household income of at least 1,000 Euros a month each coded all 104 product recalls. They had to decide which level of a content dimension – as specified in Table 2 – a company had chosen for each recall message. To avoid a single-rater bias, we aggregated the individual judgments with the help of the confidence-based weighted mean (Van Bruggen, Lilien, and Kacker 2002). Confidence had to be subjectively assessed on an increasing scale from 1 to 7. Van Bruggen, Lilien, and Kacker provide empirical support that this aggregation technique leads to better aggregate judgments. In addition to the coding of the content dimensions, we also asked our five experts to estimate the degree of hazard caused by the product defect and its probability on an increasing scale from 1 to 7. Although the content dimensions were categorized on an ordinal scale, we treat the variables as metric, because in using the confidence-weighted mean of the five judgments, the scale becomes quasi-metric. Of course, this assumes that the differences between neighboring scale points are the same.

In addition, we collected data on the product category (coded as dummy variables for the categories in Figure 1), length of text (number of words), branded product (dummy variable) and involvement (dummy variable). Our dummy variable "involvement" takes on the value "1" when the product recall comes from one of the three high-involvement categories: toys, clothing, and/or sport and leisure; otherwise, the variable

is coded "0." Table 3 provides an overview of the means, standard deviations and, for binary variables, the number of cases with a specific characteristic for all the variables. We realize that the means for the probability and the degree of hazard are fairly high in the recall messages. In contrast, the means for the dimensions social responsibility and help with identification are fairly low, whereas the means for the two dimensions "convenience of product return" and "transparency" are also fairly high.

Table 3: Sample Descriptives

	Scale	***Mean***	***Std.Dev.***	***Cases with***
Probability of Hazard	1 – 7	3.945	0.914	
Degree of Hazard	1 – 7	4.233	1.415	
Communication of Social Responsibility	1 – 6	2.497	1.081	
Offered Convenience of Product Return	1 – 6	3.611	1.176	
Offered Help with Identification	1 – 6	2.093	0.846	
Transparency of Communication	1 – 4	2.412	0.765	
Number of Words		335.048	428.578	
High-Involvement Goods				72%
Branded Products				55%

4.2 Methodology

To test the hypotheses, we ran two regressions. The first regression equations (1) – (3) model the type of communication depending on context factors that describe the situation the company faces. The second regression equations (4) – (5) deal with the design of the product return from the customer perspective, depending on the same context factors as just mentioned.

The communication of product recalls is characterized by the degree of the expressed social responsibility (Soc.Respon.), the transparency of the communication (Transp.) and the length of communication (No.Words).

$$(1) \quad Soc.Respon. = \alpha + \beta * Prob + \delta * Hazard + \gamma * Brand + \eta * Involve + \mu * No.Words + \pi * Transp. + \varepsilon$$

$$(2) \quad Transp. = \alpha + \beta * Prob + \delta * Hazard + \gamma * Brand + \eta * Involve + \mu * No.Words + \theta * Soc.Respon. + \varepsilon$$

$$(3) \quad No.Words = \alpha + \beta * Prob + \delta\ Hazard + \gamma * Brand + \eta * Involve + \theta * Soc.Respon. + \pi * Transp. + \varepsilon$$

These three dimensions of communication are explained by the context factors "degree of hazard" (Hazard), "individual probability" (Prob), "level of involvement" (Involve) and whether a branded product (Brand) is involved. The regressions also contain the other respective content dimensions to control for simultaneity.

With respect to the design of the product return, we work with two regression equations.

$$(4) \quad Con\text{-}Prod\text{-}Ret = \alpha + \beta * Prob + \delta * Hazard + \gamma * Brand + \eta * Involve + \tau * HoI + \varepsilon$$

$$(5) \quad HwI = \alpha + \beta * Prob + \delta * Hazard + \gamma * Brand + \eta * Involve + \varsigma * Com\text{-}Pro\text{-}Ret + \varepsilon$$

Regression 4 explains the convenience of the product-return procedure (Con-Prod-Ret), while Equation 5 explains the help with identification (HwI) of the defective product. Both choices of the company will be explained by the same context factors as above. Since 98% of the recalls in our sample were conducted voluntarily, we do not include the VoR variable in the model. The same is true for the variable "source of recall," because 97% of all defects were detected in-house. We estimate both structural equation models with the help of a 3 Stage Least Squares (3SLS) regression model (Green 2003, Schlichthorst 2009) due to the simultaneity and the contemporary correlation of the errors. To better assess the relationships of the variables in this study, Table 4 gives an overview of the correlations between all variables. Because the

intercorrelations are fairly low, we are not concerned with multicollinearity. Tables 5 and 6 present the results of the estimations.

4.3 Results

We tested our hypotheses by taking the respective regression coefficients of the variables in the regression equations and investigating whether a significant relationship exists. We find that Hypothesis 1 can be fully supported. In both cases, we find a negative and highly significant influence of the probability (-0.595), and a positive and highly significant influence of the degree of hazard (0.252) on social responsibility and transparency (see Table 5). For Hypothesis 2, which assumes that the probability has a negative impact and the degree of hazard has a positive impact on the overall length (No.Words) of a recall message, we find only weak empirical evidence. Both parameters in Table 5 show the correct directions but are significant only at the 10% level. With respect to Hypotheses 3 and 4 (i.e., with an increasing degree of hazard and probability, companies choose more information for identification of affected products as well as a higher convenience for product return), no empirical support was found because the respective coefficients for hazard and probability are insignificant in both tables. With respect to Hypotheses 5 and 6, our results are mixed. Except for help with identification, branded products exhibit the expected positive and significant effect on social responsibility arguments (0.528), transparency (0.491) and text length (238.001), all shown in Table 5, and convenience of product return (0.458), which is shown in Table 6. However, the results are inconclusive regarding involvement. The coefficients are not significant for social responsibility arguments, transparency and text length, whereas the coefficients for convenience of return (0.645) and help with identification (-0.468) in Table 6 point into opposite directions. Regarding the endogenous effects between the dependent variables, we can surprisingly identify counterintuitive effects.

Table 5: Results of Estimation 1 (Equations 1, 2 and 3)

	Coefficients	*Standard Error*	*Z*	*P > | z |*
Soc.Respon. (Equation 1) **R^2= 0,338**				
Prob.	**-0.595*****	0.136	-4.380	0.000
Hazard	**0.252*****	0.094	2.680	0.003
Brand	**0.528****	0.281	1.880	0.030
Involve	-0.124	0.249	-0.500	0.691
No.Words	-0.001	0.001	-0.990	0.161
Transp.	**-0.776****	0.401	-1.940	0.026
cons	5.723	1.192	4.800	0.000
Transp. (Equation (2) **R^2= 0,711**				
Prob.	**-0.415*****	0.141	-2.950	0.001
Hazard	**0.218*****	0.066	3.300	0.001
Brand	**0.491****	0.220	2.240	0.012
Involve.	0.078	0.198	0.390	0.340
Soc.Respon.	**-0.001*****	0.001	-2.350	0.009
No.Words	**-0.505****	0.245	-2.060	0.020
cons	4.492	1.032	4.350	0.000
No.Words (Equation (3) **R^2= 0,549**				
Prob.	**-132.133***	92.469	-1.430	0.076
Hazard	**74.861***	46.229	1.620	0.053
Brand	**283.001*****	102.324	2.770	0.003
Involve.	109.727	102.184	1.070	0.142
Soc.Respon.	-141.937	151.825	-0.930	0.824
Transp.	**-367.430****	165.649	-2.222	0.013
cons	1545.883	789.059	1.960	0.025

** significant at least at 0.10 Level (one-sided)*

*** significant at least at 0.05 Level (one-sided)*

****significant at 0.01 Level (one-sided)*

Regarding the influence of social responsibility on transparency, we identified a negative and significant coefficient (-0.001, see Table 5), which implies that talking about the willingness to care about customer safety results in less transparency. As expected, help with identification (1.373, see Table 6) is positively related to convenience of product return.

Table 6: Results of Estimation 2 (Equations 4 and 5)

	Coefficients	*Standard Error*	*z*	*P > |z|*
Con-Prod-Ret (Equation 4)		**R^2 = 0,729**		
Prob.	0.225	0.153	0.150	0.443
Hazard	-0.040	0.098	-0.040	0.484
Brand	**0.458***	0.284	1.610	0.054
Involve.	**0.645****	0.304	2.120	0.017
HwI	**1.373*****	0.205	6.670	0.000
cons	-1.866	1.265	-1.470	0.071
HwI (Equation 5)		**R^2 = 0,215**		
Prob.	-0.218	0.112	-0.190	0.425
Hazard	-0.001	0.072	-0.010	0.496
Brand	**-0.314***	0.221	-1.420	0.078
Involve.	**-0.468****	0.216	-2.170	0.015
Con-Prod-Ret	**0.705*****	0.095	7.460	0.000
cons	1.466	0.743	1.970	0.024

** significant at least at 0.10-Level (one-sided)*

*** significant at least at 0.05-Level (one-sided)*

****significant at 0.01-Level (one-sided)*

Tab. 4: Parameter Correlations

	Prob.	Hazard	Involve	Brand	Soc. Respon	Transp.	No. Words	Con-Prod-Ret	HwI
Probability	1	0.114	-0.09	**-.240***	**-.383****	**-.207***	-0.07	**-.286****	**-.199***
Hazard	0.114	1	-0.03	***-.267****	0.101	**.237***	-0.068	**-.279****	-0.185
Involve	-0.09	-0.03	1	-0.005	-0.081	0.035	0.138	0.048	-0.167
Brand	**-0.240***	**-.267****	-0.005	1	0.186	0.017	**.254****	**.446****	**.216***
Soc.Respon	**-0.383****	0.101	-0.081	0.186	1	***.290****	-0.034	0.036	0.13
Transp.	**-0.207***	**.237***	0.035	0.017	**.290****	1	0.099	0.091	**.242***
No.Words	-0.07	-.068	0.138	***.254****	-0.034	0.099	1	**.243***	0.011
Con-Prod-Ret	**-0.286****	**-.279****	0.048	***.446****	0.036	0.091	**.243***	1	**.240***
HwI	-0.185	-.185	-0.167	**.216***	0.13	**.242***	0.011	**.240***	1

** Correlations at 0.01 Significance Level (double-sided)
*. Correlations at 0.05 Significance Level (double-sided)

5 Discussion and implications

As predicted, we were able to detect some routines for communicating a product recall. Apparently, companies tend to differentiate between communicating a product recall and organizing the product return, and alter their behavior accordingly.

Although we find empirical evidence that (as hypothesized) companies overemphasize social responsibility when the probability of hazard is low or the degree of hazard is high, we cannot find any evidence that the probability or degree of hazard has an influence on the offered comfort of a product return, or on the amount of information customers are given to ease identification of affected products. It is possible that companies want their customers to perceive a product recall as a service activity rather than as a removal of defective products forced by law.

The same can be concluded for the degree of transparency of communication. Companies tend to design the communication more transparently when the degree of hazard is high. We assume that a clear and non-relativizing communication of problems is used when companies wish to avoid further consumer damages caused by the defective product — the consequence of which being the company would receive even more negative headlines in the mass media.

It seems that high involvement and brand products are recalled in a more convenient way than non-branded or low-involvement products. This may be explained by focusing on cost effects. Most high-involvement or brand products in the sample were high-priced products for which a repair is more worthwhile than simple replacement. Furthermore, we assume that these companies use the opportunity of a product recall to underline and prove their quality claims and beliefs toward their customers by trying to avoid any inconvenience and helping customers identify affected goods quickly and easily.

At the same time, brand-producing companies seem to attempt to keep customers' trust in their brands by emphasizing social responsibility and showing a maximum amount of transparency. This strategy may be explained by the fact that most brand-

producing companies try to avoid cross-effects to other products in their brand family to avoid damaging the long-term brand value. Interestingly, the findings for involvement are inconclusive. If at all, we find that companies offer a more convenient product return for high-involvement products than for low-involvement products because they fear stronger negative reactions. On the other hand, we cannot find a positive influence of involvement on the ease of identification. This rather inconsistent behavior may be caused by the fact that product recalls are very rare events, often catching companies unprepared for them. For example, companies cannot react quickly enough to install service centers or hotlines. Thus, companies can substantially improve their behavior in cases of necessary product recalls by preparing early routines or scenarios. Even if our study identifies some valuable insights as to how companies react to recall situations, more empirical investigations on the effects of communication patterns and the design of product returns on the perceptions of consumers as well as their subsequent buying behavior are necessary to provide more valuable insights for companies.

6 Summary

This study provides interesting results about the influence of situational factors on the behavior of companies and their communication of product recalls. Although studies in the literature focus on the effects of product recalls on brand image and sales, we analyze how companies choose their communication behavior and design their product return policy for consumers. Based on a content analysis, we find several unique dimensions of communication that companies follow, and we investigate whether the companies choose the levels of the respective communication dimensions according to situational variables such as the degree of hazard, its probability, whether the product is branded or if it represents a high-involvement product.

This study shows that companies usually follow a consistent communication behavior pattern across different situations. The degree of hazard and the probability of a

product defect play an important role in expressing social responsibility and showing transparency. Furthermore, we find empirical evidence for the fact that companies producing branded and high-involvement products organize their recalls to be more convenient for their customers. A limitation of the study is that we cannot make any suggestions about optimal recall strategies because we lack information about customers' reactions to the levels of the communication dimensions in different situations. Given this fact and the low number of product recalls and in light of the difficulties faced while collecting data, future research should focus on insights from experts with the help of conjoint analyses or policy capturing approaches.

References:

Bapuji, Hari and Paul W. Beamish (2007): *Toy Import and Recall Levels: Is There a Connection?*, Source: http://www.asiapacific.ca/analysis/pubs/pdfs/commentary/cac45.pdf.

Beamish, Paul W. and Hari Bapuji (2008): Toy Recalls and China: Emotion vs. Evidence, *Management and Organization Review*, 4 (2): 197-208.

Cezanne, Marc (1999): *Krisenmanagement und Komplexität*, Deutscher Universitäts-Verlag: Wiesbaden.

Chu, Tin-Teng, Lin Che-Chun and Larry J. Prather (2005): An Extension of Security Price Reactions Around Product Recall Announcements, *Quarterly Journal of Business & Economics*, 44 (3/4): 33-48.

Cleeren, Kathleen, Marnik G. Dekimpe and Kristiaan Helsen (2008): Weathering Product-Harm Crises, *Journal of the Academy of Marketing Science*, 36 (2): 262-270.

Coombs, Timothy (1998): An Analytic Framework for Crisis Situations: Better Responses From a Better Understanding of the Situation, *Journal of Public Relations Research*, 10 (3): 177-219.

CPSA (1972)*: Consumer Product Safety Act*, Washington, 1972, Source: http://www.cpsc.gov/cpscpub/pubs/reports/2003rpt.pdf.

CPSC (2007)*: Consumer Product and Safety Commission Annual Report*, Washington, 2007, Source: http://www.cpsc.gov/cpscpub/pubs/reports/2007rpt.pdf.

CPSC (2008): C*onsumer Product Safety Improvement Act,* Washington, 2008, Source: http://www.cpsc.gov/ABOUT/Cpsia/cpsia.html.

Dawar, Niraj and Madan M. Pillutla (2000): Impact of Product-Harm Crises on Brand Equity: The Moderating Role of Consumer Expectations, *Journal of Marketing Research*, 37 (2): 215-226.

de Matos, Celso Augusto and Carlos Alberto Vargas Rossi (2007): Consumer reaction to product recalls: factors influencing product judgement and behavioural intentions, *International Journal of Consumer Studies*, 31 (1): 109-116.

Felcher, E. Marla (2003): Product Recalls: Gaping Holes in the Nation's Product Safety Net, *Journal of Consumer Affairs*, 37 (1): 170-179.

Folkes, Valerie S. (1984): Consumer Reactions to Product Failure: An Attributional Approach, *Journal of Consumer Research*, 10 (4): 398-409.

Gerbner, George (1969): *The Analysis of Communication Content: Developments in Scientific Theories and Computer Techniques*, John Wiley & Sons Inc: New York.

GPSG (2004): *Gesetz über technische Arbeitsmittel und Verbraucherprodukte (Geräte- und Produktsicherheitsgesetz – GPSG)*, Source: http://bundesrecht.juris.de/-bundesrecht/gpsg/gesamt.pdf.

Greene, William H. (2003): *Econometric Analysis*, 5th edition, Prentice Hall: New Jersey.

Haller, Markus (1998): Reden wir darüber? Die Kommunikationspolitik der Deutschen Bank im "Fall Schneider", in: Günter Bentele, Lothar Rolke and Michael Huber (eds.): *Konflikte, Krisen und Kommunikationschancen in der Mediengesellschaft - Case Studies aus der PR-Praxis*, Vistas, Berlin, 57-109.

Jolly, David W. and John C. Mowen (1985): Product Recall Communications: The Effects of Source, Media and Social Responsibility Information, *Advances in Consumer Research*, 12 (1): 471-475.

Kersten, Jan (1992): Rückrufkosten - ein weltweites Risiko und seine Kosten, *Versicherungswirtschaft*, 7 (8): 537-539.

Konken, Michael (2002): *Krisenkommunikation. Kommunikation als Mittel der Krisenbewältigung*, FBV Medien-Verlags GmbH: Limburgerhof.

Krippendorff, Klaus (1980): *Content Analysis: An Introduction to its Methodology*, Sage: Newbury Park.

Mayring, Philipp (2000): Qualitative Inhaltsanalyse, *Qualitative Social Research*, 1 (2): 1-27.

Mayring, Philipp (2008): *Qualitative Inhaltsanalyse*, Beltz: Weinheim und Basel.

Mowen, John C. (1979): Further Information on Consumer Perceptions of Product Recalls, *Advances in Consumer Research*, 7 (1): 519-523.

Mowen, John C., David Jolly and Gary S. Nickell (1981): Factors Influencing Consumer Responses to Product Recalls: A Regression Analyses Approach, *Advances in Consumer Research*, 8 (1): 405-406.

Rapex (2009*):* *Rapid Alarm System for Non Food Consumer Goods*, Source: http://ec.europa.eu/consumers/safety/rapex/index_en.htm.

Rhee, Mooweon and Pamela R. Haunschild (2006): The Liability of Good Reputation: A Study of Product Recalls in the U.S. Automobile Industry, *Organization Science*, 17 (1): 101-117.

Schlichthorst, Marisa (2009): Mehrgleichungsmodelle: Schätzmethoden und Anwendungsbeispiele, in: Sönke Albers, Daniel Klapper, Udo Konradt, Achim Walter and Joachim Wolf (eds.): *Methodik der empirischen Forschung*, 3rd edition, Gabler, Wiesbaden, 349-364.

Siomkos, George and Paul Shrivastava (1993): Responding to Product Liability Crises, *Long Range Planning*, 26 (5): 72-79.

Standop, Dirk (1993): Sicherheitskommunikation, in: Ralph Berndt and Arnold Hermanns (eds.): *Handbuch Marketing-Kommunikation*, Gabler, Wiesbaden, 946-964.

Standop, Dirk (1995): Die Nebenwirkungen von Produktrückrufen: Ansätze der empirischen Analyse, in: Horst Albach (ed.): *Dynamik und Risikofreude in der Unternehmensführung*, Gabler, Wiesbaden, 71-94.

Standop, Dirk (2006): Der Verlust von Konsumentenvertrauen gegenüber Anbietern: der Fall von Produktrückrufen, in: Hans Bauer, Marcus M. Neumann and Anja Schüle (eds.): *Konsumentenvertrauen*, Vahlen, München, 95-104.

Trommsdorff, Volker and Fee Steinhoff (2007): *Innovationsmarketing*, Vahlen: München.

Van Bruggen, Gerrit H., Gary L. Lilien and Manish Kacker (2002): Informants in Organizational Marketing Research: Why Use Multiple Informants and How to Aggregate Responses, *Journal of Marketing Research*, 39 (4): 469-478.

Van Heerde, Harald, Kristiaan Helsen and Marnik G. Dekimpe (2007): The Impact of a Product-Harm Crisis on Marketing Effectiveness, *Marketing Science*, 26 (2): 230-245.

C The Impact of Product Recall Communication on Brand Image, Brand Attitude and Perceived Product Quality

Raoul V. Kübler and Sönke Albers

Published in: Kübler, Raoul V. and Sönke Albers (2012): *The Impact of Product Recall Communication on Brand Image, Brand Attitude and Perceived Product Quality*, Available at SSRN: http://ssrn.com/abstract=2060858

.

Abstract: The number of product recalls has considerably increased during the last decade resulting in escalating costs for the public and companies likewise. Companies are confronted not only with severe consequences of their short-term performance (e.g. sales) but also of long-term marketing metrics such as consumer's brand image, brand attitude and perceived quality. Therefore, companies need to know how they should design their recall communication to minimize negative impact. Using data of 450 consumers' reactions to 16 different and major German product recalls, the authors investigate the impact of recall communication characteristics like transparency and social responsibility on customers' brand evaluation. They further measure how the particular degree of hazard and its probability of a product failure moderate the impact of the chosen communication strategy. The findings provide companies with recommendations how to design their recall messages depending on the context of the recall necessity.

Key words: brand evaluation, effectiveness of communication instruments, product recall, seemingly unrelated regression

1 Introduction

The number of product recalls has considerably increased during the last decade (Beamish and Bapuji 2008). In the toy sector alone this boost resulted for the US in 22 toy-related deaths and an estimated 220,500 toy-related injuries in 2006. In addition, the public costs related to such product crises or recalls are immense. According to the U.S. Consumer Product Safety Commission crisis related costs due to product failures only in the United States sum up to a yearly bill of over 700 billion US-dollars (CPSC 2005).

The costs and other risks for affected companies have to be considered as substantive and threatening. Previous empirical research shows that a product crisis or a recall scenario has a considerable impact on a company's short term performance and marketing activities. Mowen, Jolly, and Nickell (1981) and Coleman (2011) show that companies who are affected by a recall necessity substantively loose in sales. While this seems to be obvious for affected products, previous research underlines that in the case of multiproduct- or multibrand-companies product recalls may even develop negative spill-over effects on the sales of non-affected but associated products (Van Heerde, Helsen, and Dekimpe 2007). Such sales losses further lead to a negative impact onto financial and the stock market performance of the affected companies (Cheah, Chan, and Chieng 2007, Chen, Ganesan, and Liu 2009).

Whereas the outcome of these short time effects already poses severe problems to marketers, one has to keep in mind that a specific product recall may also negatively influence long-term marketing metrics like the respective brand equity and customer's attitude toward the affected product (Dawar and Pillutla 2000, Siomkos and Shrivastava 1993). Cleeren, Dekimpe, and Helsen (2008) show that this effect in turn leads to a long-term decrease of future sales as well as a loss of market share.

As companies do not only bear responsibility for their customers but also for all their stakeholders, managers have to keep in mind that while communicating a product defect and warning consumers they also have to secure the future of the affected company and the related workplaces.

The major instrument to reduce the negative impact of a product harm crisis or product recall on the above mentioned brand metrics is certainly the recall communication. Classic recall communication addresses a wide public and informs possibly affected customers about the risks of the specific product defect and asks them to directly return the defect product to its manufacturers or retailers. Whereas recall campaigns may help to prevent further incidents, many marketers are afraid that non-affected customers may get scared or unsettled by any recall information. Surprisingly, only little light has yet been shed on the design and content of recall communication and how different patterns of communication behavior do affect customer's perception and evaluation of a recalling company. Kübler and Albers (2010) investigate how companies communicate product recalls. They identify four typical communication patterns of recall messages. By using different degrees of communicated social responsibility, transparency, identification help or excuse notes, marketers adapt their individual recall communication to the specific situational requirements of a product recall, like the degree of hazard or its probability and hope that this will help to minimize possible negative effects.

Although we observe that managers use these complex communication instruments, uncertainty about their effects is still high as – to the best knowledge of the authors – no study has yet examined how these instruments really moderate the impact of a product recall on the different brand metrics and the consumer's perception of the recalling company. As consumers' general perceptions of recalls already vary between different types of products and branded and non-branded products (de Matos and Rossi 2007, Jolly and Mowen 1985) one must as well assume that the impact of any communication will also vary between different product categories, types of recall scenarios or types of defects. Thus recalling companies have to be aware that the impact of their recall communication might be moderated by these situational variables and that they cannot just implement a universal communication strategy and hope that it will solve all problems. Hence, to generally help marketers to optimally communicate a product failure and to minimize any negative effects of a product recall

on the different valuable brand metrics, one must find answers to the following two yet unanswered questions:

1. How does the specific design of a recall communication campaign moderate the impact of the product recall on brand image, brand attitude and perceived product quality?
2. How is the impact of the communication moderated by the recall specific situational components such as the type of product defect or the degree and probability of the defect-induced hazard?

We are the first to analyze the impact of different recall communication patterns (stated social responsibility, degree of transparency, help with identification, and provided comfort of product return) on key marketing measures like brand image, brand attitude and perceived product quality. Our study also analyzes the moderating impact of the hazard degree and its probability on the impact of the above described communication patterns. The results are of high managerial relevance as they provide insights into the importance of different context-specific recall communication strategies and will help managers to better target and inform affected customers without having to fear a sustainable hurt of own marketing key measures.

The paper is structured as follows. The next section presents the conceptual framework and develops the research hypotheses. We then outline our empirical study and measures, and report the results of our analyses. The paper concludes with a discussion of managerial implications, limitations, and directions for future research.

2 Theoretical Framework

2.1 The Impact of Stated Social Responsibility

Empirical knowledge on how consumers react to product recalls and how companies can influence these reactions is still small. Earlier studies focused more on how consumers in general react to product recalls (Mowen 1979, Mowen, Jolly, and Nickell 1981). While these studies mostly revealed a plausible negative impact of a product recall on consumer perceptions, succeeding research could show that

companies who react early to defects and who try to immediately communicate the product failure face less negative responses than slower companies (Vassilikopoulou, Lepetsos, Siomkos, and Chatzipanagiotou 2009). This is due to consumers apparently assessing the corporate social responsibility communicated by the recalling company. On the one hand, companies who react fast are considered to fulfill their responsibility toward the public as they judge consumer protection higher than their own marketing or image issues. On the other hand hesitating companies are evaluated as less responsible because of being selfish and only self-advantage-orientated. These assumptions are in line with the general findings by Jolly and Mowen (1985) and Vanhamme and Groben (2009) who both show that emphasizing the social responsibility of the recalling company helps to reduce the negative effects of a product recall. In addition, Lin, Chen, Chiu, and Lee (2011) find that consumer's rebuy-decisions are positively influenced by the social responsibility shown by the affected company during a product harm crisis. Kübler and Albers (2010) find that companies are well aware of this effect because many recall campaigns already contain information about the social responsibility of the recalling company and the social and ethical claims a company shares. The stated degree thereby varies from low by only emphasizing the importance of consumer safety to high by stating responsibility for consumer safety together with the accentuation of high corporate safety standards and the relativization of the occurred product defect. Thus, we hypothesize that companies who mask their legal obligation to recall a defect product with higher degrees of stated social responsibility will be rewarded with less negative consumer reactions:

> H1: *Consumers react in terms of (1) brand image, (2) brand attitude, and (3) perceived product quality more positively to product recalls with a higher communicated degree of social responsibility than to product recalls with lower communicated degrees of stated social responsibility.*

According to previous research on the impact of a product recall on customer's perception, the degree of hazard affects the overall evaluation of a product recall (de

Matos and Rossi 2007). In addition, van Heerde, Helsen, and Dekimpe (2007) show that a recall does not only affect consumers' perceptions but also the effectiveness of the employed marketing instruments like advertising or pricing. Therefore one can assume that the degree of hazard will not only have a direct impact on consumers' perceptions but might as well affect the effectiveness of the used communication strategy and its primary tools like the stated degree of social responsibility. As higher degrees of hazard will cause more uncertainty and thus greater levels of perceived personal insecurity it can been assumed that it will increase the various negative effects and decrease the various positive effects of the different recall communication elements. Furthermore, Dawar and Pillutla (2000) demonstrate that consumers' reactions toward a product recall are directly moderated by their expectations toward the recalling company. Thus higher degrees of hazard and the related higher degree of uncertainty will thereby lead to a higher expectancy of stated social responsibility, which will thus increase the impact of this variable. Hence, we can assume that:

> $H1_M$: *Higher degrees of hazard will increase the impact (as in H1) of the stated degree of social responsibility.*

2.2 The Impact of Chosen Degree of Transparency

Previous research shows that consumers' negative reactions to product recalls are mostly caused by the uncertainty about possible personal consequences resulting from a product failure (Coombs 1998, de Matos and Rossi 2007). Whereas the stated social responsibility is obviously only intended to improve consumers' reactions toward the recalling brand, the chosen degree of communicated transparency can be understood as a tool to minimize customer's uncertainty and fears by providing information about the recall cause and the threats of the specific product defect (Simola 2003). Choosing the right degree of transparency is not an easy task for marketers. On the one hand, the communication of higher risks may scare customers so that they stop using a defect product, which may prevent further consumer casualties. On the other hand, companies have to fear that higher degrees of transparency may also scare unaffected

customers or even future buyers and will thus increase sales losses and the negative impact of a product recall on the various long-term marketing metrics (Cleeren, Dekimpe, and Helse 2008). Providing suitable and detailed information about the product defect and its consequences may therefore help to reduce these fears. Thus higher degrees of communicated transparency will help consumers to better understand the product recall and make customers think that the recalling company really cares for its customers. This may also increase customer satisfaction and thus help decrease the overall negative effects on brand attitude, brand image and perceived quality. Hence we end up with the following hypothesis:

> H2: *Consumers react in terms of (1) brand image, (2) brand attitude, and (3) perceived product quality more positively to product recalls communicated with higher degrees of transparency than to product recalls with lower degrees of transparency.*

One can assume that this positive effect may additionally be increased by the hazard probability. In case of a low hazard probability consumers are more likely to evaluate their own risk as small. So it can be assumed that they are more likely to forgive a lack of transparency. In contrast, consumers will perceive a high hazard probability as more frightening and unsettling and will thus call for higher degrees of communicated transparency. Hence we derive the following hypothesis:

> $H2_M$: *Higher degrees of hazard probability will increase the positive impact (as in H2) of the stated degree of transparency.*

2.3 The Impact of Product Return Convenience

Examples from the automotive sector show that product recalls can be used as an opportunity to demonstrate a brand's service capabilities (Rhee and Haunschild 2006). Car manufacturers nowadays offer their clients new car models, which they can test-drive for free during the time that their own car is repaired or upgraded. Thereby, well-executed product recalls can be understood as a promotion tool as they allow marketers to demonstrate how intensively they care for their customers. If the recall

itself has to be communicated publicly it will reach an even broader audience. Demonstrating a high degree of customer orientation by making recalls as easy and comfortable as possible or by providing a compensation for the inconveniences caused by the initial recall, will help marketers to improve customer satisfaction and may even help to attract future buyers. By all means, higher degrees of conveniences will certainly decrease the negative effect of a product recall on long-term marketing metrics. Hence, we derive the following hypothesis:

> H3: *Consumers react in terms of (1) brand image, (2) brand attitude, and (3) perceived product quality more positively to product recalls with higher degrees of product return convenience than to product recalls with lower degrees of return convenience.*

2.4 The Impact of the Source of Recall

Not only the elements of a product recall but also the used media channel has a severe impact on consumers' reactions (Jolly and Mowen 1985). Furthermore, not only the type of media but also the source of defect identification might affect brand assessment. As consumers tend to lay more trust into public media than into corporate communication, it seems plausible to hypothesize a less negative impact on brand evaluation if the source where someone becomes aware of a recall is public media (high level) compared to companies' press releases (low level). Hence we end up with the following hypothesis:

> H4: *Consumers react in terms of (1) brand image, (2) brand attitude, and (3) perceived product quality more positively to product recalls if the source of the recall is public media rather than press releases by the company itself.*

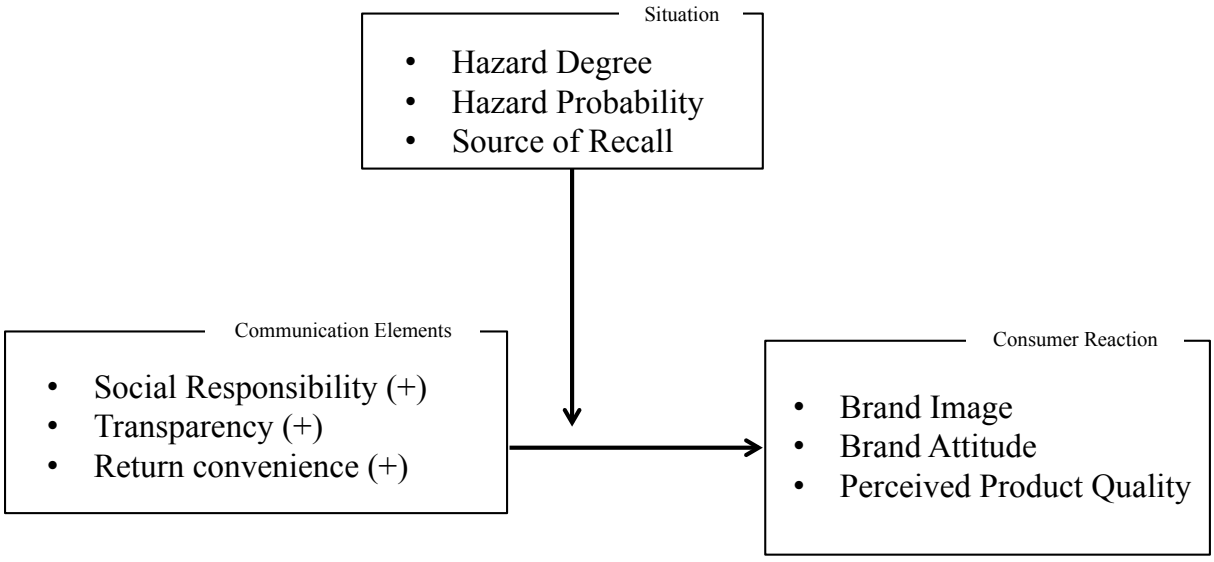
The credibility-lead of public media may not only directly influence the overall evaluation of a product recall by the customer, but may also in- or decrease the effectiveness of the used communication strategy of the recalling company. As consumers lay more trust in messages published by official media compared to corporate communication (Jolly and Mowen 1985), one can assume that this effect

will also increase the effectiveness of the specific message content when published by official media. Kübler and Albers (2010) argue that customers become suspicious if social responsibility statements of a company are only blind excuses that are solely meant to delude customers. As consumers in general lay more trust in the press and believe that public media will carefully check any corporate statements before printing or airing them, the impact of social responsibility statements will be increased when being published by the press. So we end up with the following hypothesis:

> $H4_M$: *The impact of any Social Responsibility information (as in H4) will be stronger when the initial source of the recall is a media product compared to when it is a corporate publication.*

Figure 1 gives a final overview of our research model.

Figure1: The Research Model



3 Sample and Measurement Development

To empirically test our hypotheses we collect data on different product recalls and measure the impact of the related recall communication on the overall brand evaluation by consumers who were aware of the respective recalls. To ensure enough variation in our data we identified 16 different suitable German product recall

campaigns that do not only come from different product categories but also differ in how they were communicated. All recall campaigns were not older than 18 months and the defect products originate from six different product categories, such as toys, tools or electronic goods. In addition, the data has been balanced between branded and non-branded products and between recall cases with high hazards and low hazards. Additional information about the different recall cases can be obtained from table 3.

To gather data about the impact of the recall communication on brand image, brand attitude and perceived product quality, the authors developed an online questionnaire. Participants were acquired in cooperation with Germany's largest product recall web platform www.produktrueckrufe.de and in cooperation with several other highly frequented but not recall related German web platforms through newsletter advertising and special banners. To keep the questionnaire as short as possible, the product recall sample was divided into two equally sized and balanced samples with 8 different recall scenarios within each questionnaire version. Participants were randomly assigned to one of the two questionnaires.

Table 1 gives detailed information about our study participants. More than 40% of all participants range between an age of 31 and 45 and more than 57% of our participants are employed. In contrast to many other studies, the number of students in the sample is rather low with a share of only 18 percent. The number of male participants (48%) is slightly higher than the number of female participants (52%).

Participants were first asked to indicate which of the particular products they knew. Based on the three well-established scales developed by Low and Lamb (2000) participants were then asked to rate the brand attitude, the brand image and their perception of the particular product quality of the known products. Table 2 gives an overview over the nine different items used to measure the three different constructs. Item aggregation was obtained by simple addition of the particular items.

Afterwards, participants were asked to state if they were aware of the fact that the particular products had been recalled. This specific question order secured that the stated brand image, brand attitude and perceived product quality measures could not

be biased by the possible recall information. Furthermore, respondents who were aware of a product recall could indicate where they first heard about it (source of recall). Participants could choose between 9 different media types and information sources like e.g. daily regional newspapers, national daily newspapers, weekly magazines, TV shows, through friends or through point of sale information.

In addition, an open blank field was used to gather possible non-mentioned contact sources. Not all respondents were aware of all the different recall cases. Thus, we obtain different numbers of observations for each of the recall cases. Table 3 gives an overview about how many observations we could gather for each of the cases. Overall, we end up with a total of 2262 observations for the 16 different recall scenarios.

Table 1: Characteristics of Participants

	Total	Per Cent
Sex		
male	216	48,00
female	234	52,00
Marital status		
single	223	49,56
married	227	50,44
Age		
Younger than 24	38	8,44
24-30	102	22,67
31-45	197	43,78
46-60	78	17,33
Older than 60	35	7,78
Number of children		
No children	239	53,11
1 child	91	20,22
2 children	86	19,11
3 children	17	3,78
More than 3 children	17	3,78
Occupation		
student	83	18,44
employee	258	57,33
self-employed	51	11,33
without occupation	24	5,33
retired	34	7,56

Table 2: Measuring Items for Brand Image, Brand Attitude and Perceived Product Quality According to Low and Lamb 2000

		Question	Scale
Brand Image		For me, the product is/has	
	Item1	not useful (1) / useful (5)	5-point-scale
	Item2	non-durable (1) / very durable	5-point-scale
	Item3	a bad image (1) / a good image (5)	5-point-scale
Brand Attitude		For me the product is…	
	Item1	unpleasant (1) / pleasant (5)	5-point-scale
	Item2	bad (1) / good (5)	5-point-scale
	Item3	worthless (1) / valuable (5)	5-point-scale
Perceived Product Quality		For me, the product is…	
	Item1	inferior (1) / superior (5)	5-point-scale
	Item2	poor (1) / excellent (5)	5-point-scale
	Item3	low quality (1) / high quality (5)	5-point-scale

To obtain content specific recall communication measures the authors followed the approach by Kübler and Albers (2010) and deployed a content analysis for all 16 recall messages. Based on the different dimensions provided by those authors, 6 different recall and public relation experts evaluated the 16 different recall messages. Experts had been randomly targeted by the authors and asked for evaluation. The intercoder-reliability (based on an ICC approach) lies at .856 and can be thus considered as satisfactory. The ICC has been computed according to the well-established approach of McGraw and Wong (1996). For the aggregation of the different evaluations, the authors follow the approach of confidence-based weighted mean taking into account the informants' self-assessed confidence in the accuracy of each response estimate. The individual response gets than multiplied by the confidence weight that is divided by the sum of all confidence weights. Finally, the weighted responses get summed up. Hence, the response from more confident informants weights more heavily than those from less confident informants (van Bruggen, Lilien, and Kacker 2002). Table 3 also gives an overview over the mean aggregated values for each of the different communication measures.

Table 3: Recall Specific Information and CBWM Aggregated Expert Ratings

Recall No.	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
Product	Tea Cup	Ski	Climbing Harness	HDD	Hair-Dryer	Baby Buggie	Music Player	Shower Tub	Electric Kettle	Grinding Wheel	Bath Tub Handle	Tea Cup	GPS Device	Electric Generator	Vacuum Cleaner	Kids Socks
Category	Kitchen	Sport	Sport	Elec. Goods	Home	Toys/ Kids	Elec. Goods	Home	Kitchen	Tools	Home	Kitchen	Elec. Goods	Tools	Home /	Toys / Kids
Brand	No	Yes	No	No	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	Yes	Yes	No
Observations	165	98	74	87	197	75	194	197	200	91	100	202	121	76	140	245
Social Responsibility	4.60	1.27	3.05	4.05	4.79	5.40	1.26	1.00	3.89	4.50	1.33	4.43	5.47	4.33	1.68	5.17
Transparency	3.13	2.39	2.00	3.71	3.47	3.33	2.67	1.50	2.71	2.83	3.00	3.78	3.53	2.72	2.47	1.59
Help with Identification	5.74	4.84	1.68	4.89	5.50	5.50	4.17	1.00	4.11	4.16	4.11	3.87	4.06	5.26	3.53	5.74
Convenience of Product Return	2.00	3.22	4.50	5.16	4.79	1.38	4.00	4.25	4.00	4.88	4.65	2.00	5.00	4.28	5.07	2.00
Hazard Degree	4.00	4.31	2.67	3.20	5.44	3.80	1.00	2.27	4.00	4.18	3.63	2.62	2.78	4.00	5.68	3.20
Hazard Probability	2.68	2.00	3.89	4.79	3.50	2.00	2.41	2.50	2.78	3.00	2.84	3.14	2.44	3.73	2.53	2.80

4 Model Development

We test our research hypotheses with the help of the three different regression models by explaining how the communication elements and their moderators affect the three marketing metrics brand image, brand attitude, and perceived product quality:

$$(1) \quad \begin{aligned} BrImg = {} & \alpha + \beta_1 \cdot SocR + \beta_2 \cdot Trans + \beta_3 \cdot Con\,\mathrm{Re} + \beta_4 \cdot SoR + \beta_5 \cdot HazD + \\ & \beta_6 \cdot Haz\,\mathrm{Pr} + \beta_7 \cdot SocR \cdot HazD + \beta_8 \cdot Trans \cdot Haz\,\mathrm{Pr} + \beta_9 \cdot SocR \cdot SoR + \varepsilon \end{aligned}$$

$$(2) \quad \begin{aligned} BrAtt = {} & \alpha + \beta_1 \cdot SocR + \beta_2 \cdot Trans + \beta_3 \cdot Con\,\mathrm{Re} + \beta_4 \cdot SoR + \beta_5 \cdot HazD + \\ & \beta_6 \cdot Haz\,\mathrm{Pr} + \beta_7 \cdot SocR \cdot HazD + \beta_8 \cdot Trans \cdot Haz\,\mathrm{Pr} + \beta_9 \cdot SocR \cdot SoR + \varepsilon \end{aligned}$$

$$(3) \quad \begin{aligned} PerQual = {} & \alpha + \beta_1 \cdot SocR + \beta_2 \cdot Trans + \beta_3 \cdot Con\,\mathrm{Re} + \beta_4 \cdot SoR + \beta_5 \cdot HazD + \\ & \beta_6 \cdot Haz\,\mathrm{Pr} + \beta_7 \cdot SocR \cdot HazD + \beta_8 \cdot Trans \cdot Haz\,\mathrm{Pr} + \beta_9 \cdot SocR \cdot SoR + \varepsilon \end{aligned}$$

The dependent variables BrImg, BrAtt and PerQual are standing for Brand Image, Brand Attitude and Perceived Quality. On the right-hand side of the equations, SocR, Trans, ConRe, and SoR, stand - as depicted in our formal research model - for the independent variables Social Responsibility, Transparency, Convenience of Product Return, Source of Recall, while HazD and HazPr represent the moderators Hazard Degree and Hazard Probability.

We estimate all three equations simultaneously with a seemingly unrelated regression (SUR) model. Implementing interaction terms into a regression by simple multiplication can lead to severe multicollinearity-issues (Belsley, Kuh, and Welsch 1980). To avoid such effects we follow the well-documented approach of Aiken and West (1991) and orthogonalize all interaction terms.

5 Results

We first try to replicate the initial findings of Mowen (1979) and test if a product recall generally has a negative impact on the different brand metrics. Therefore, we use the two groups of our sample (participants who are aware and not aware of the recall) and test if these two groups differ in how they judge the average brand image, brand

attitude, and perceived quality with the help of an independent-sample t-test. The results in table 4 show that there is a slight but significant difference in the perception of a brand's perceived product quality, based on a 5 point semantic differential, between subjects who remember a certain recall (m=3.1068) and those who do not (m=3.2436). Hence, people who are aware of a past product failure tend to judge the product of less quality than subjects without previous knowledge about recalls. We obtained similar results for brand image and brand attitude although the difference for brand attitude is not significant. While this result is not surprising, it justifies a closer look at the determinants of brand evaluation with respect to recall characteristics.

Table4: Results of Independent-Sample t-test

	Recall	mean	sd	t-value	Sig. (2-tailed)
Brand Image	not know	3.3305	1.2845	2.320[b]	0.035
	know	3.1752	1.2754		
Brand Attitude	not know	3.3716	1.0848	0.202[ns]	0.840
	know	3.3597	1.1391		
Perceived Product Quality	not know	3.2436	1.2415	2.111[b]	0.021
	know	3.1068	1.2353		
	know	3.1752	1.2754		

[b] significant at 10%; [ns] not significant

We test our hypotheses H1 to $H4_M$ based on the regression coefficients and t-values of the above described SUR-model. For this purpose we use the part of the sample that is able to remember a certain product recall. With an acceptable fit (R^2 of at least 0.339) each of the three marketing metrics can be explained by six main effects and three interaction terms. Most of the estimated coefficients are significant at the 1%-level. The results in Table 5 show that all three models obtain similar estimates for almost all recall characteristics across brand evaluation measures.

Surprisingly, we do not find empirical support for our first hypothesis, which states that higher degrees of stated social responsibility would lead to a better evaluation of the company. Counterintuitive to our assumption the effect is negative and significant

(at 1%-level) for all of the three brand metrics (Brand Image: -.339, Brand Attitude: -.183, and Perceived Product Quality: -.412). In addition to the negative main effect of social responsibility, our model identifies a significant negative interaction between the hazard degree and the degree of stated social responsibility (Brand Image: -.459, Brand Attitude: -.357, and Perceived Product Quality: -.310). Additionally to the interpretation of the regression coefficients we probed the pattern of the interaction by examining the simple slope of social responsibility for high (mean + one std. dev.) and low degrees of hazard (mean – one std. dev.) in case of the three different brand metrics (Aiken and West, 1991). Figures 2, 3 and 4 graphically show the results of this analysis. In all of the three cases the negative slope of the straight line plotting the high hazard condition is steeper. In addition, the straight line of the high hazard condition lies under the straight line of the low hazard condition, indicating that higher degrees of hazard do enforce the negative impact of social responsibility on Brand Image, Brand Attitude and the Perceived Product Quality. Consistent with our estimates this effect seems to be true for all of the three different brand metrics. In line with our second hypothesis transparency of communication has a significant and positive impact on brand image, brand attitude, and perceived product quality (.897, .760, and .991 respectively). Hence, a communicated high level of transparency during the recall leads to a more positive brand evaluation. Surprisingly, the interaction between hazard probability and transparency turns out to be negative (Brand Image: -1.785, Brand Attitude: -1.373, and Perceived Product Quality: -1.826). Very likely, customers get more scared by a detailed information policy when facing higher risks of being affected.

This assumption is again underlined by the difference of slopes when you plot the straight line for transparency for the three different brand metrics. Figures 5, 6 and 7 clearly show that – in contrast to our initial hypotheses – higher degrees of hazard probability increase the overall negative impact on attitude, image and quality. So we have to reject hypothesis $H2_M$.

The impact of convenience of product return has - as assumed in hypothesis 3 - a positive and highly significant (at 1%-level) impact on the brand evaluation (Brand Image: 1.242, Brand Attitude: 1.117, and Perceived Product Quality: 1.341). Thus we find strong empirical support for our third hypothesis. At last, we assume a positive relation between the source of recall and the particular brand evaluation. Our model delivers for all of the three different brand metrics significant and positive betas (Image: .466, Brand Attitude: .556, and Perceived Product Quality: .340), providing empirical evidence for our fourth hypothesis. In addition, our model delivers - in line with hypothesis $H4_M$ - positive betas for the interaction between social responsibility and the source of recall in case of the Brand Image-model (.184) and the Brand Attitude (.297), while the interaction effect remains insignificant in case of the Perceived-Product Quality-model. Figures 8 and 9 show the corresponding plots of the two significant effects. In line with our assumptions the straight line for the public media condition lies above the straight line of the corporate communication situation (like e.g. an advertisement, a message on the corporate webpage or a message at the point of sale), indicating that consumers react more positively to any social responsibility information when they find them for the first time in the media and not in some sort of corporate communication. In addition, one can observe – in case of Brand Image - that the slope of the straight line for the media condition appears to be significantly higher, than the slope for the straight line of the corporate media condition. So we find at least particular support for hypothesis $H4_M$.

Table5: Regression Coefficients of Main and Interaction Effects on the Three Dimensions of Brand Evaluation

	Coef.	Std.Err.	z	P>|z|	[95% Conf.Interval]	
Brand Image (R^2 = .383)						
Social Responsibility	**-0.339**	0.078	-4.35	0.000	-0.491	-0.186
Transparency of Communication	**0.897**	0.155	5.8	0.000	0.594	1.200
Convenience of Product Return	**1.242**	0.068	18.27	0.000	1.108	1.375
Source of Recall	**0.466**	0.069	6.75	0.000	0.330	0.601
Hazard Degree	-0.064	0.066	-0.98	0.328	-0.193	0.064
Hazard Probability	-0.169	0.135	-1.25	0.211	-0.433	0.095
IASocR x HazD	**-0.459**	0.100	-4.57	0.000	-0.656	-0.263
IATrans x HazP	**-1.785**	0.177	-10.07	0.000	-2.132	-1.437
IASocR x SoR	**0.184**	0.064	2.86	0.004	0.058	0.310
constant	**3.747**	0.411	9.12	0.000	2.942	4.553
Brand Attitude (R^2 = .339)						
Social Responsibility	**-0.183**	0.078	-2.34	0.019	-0.336	-0.030
Transparency of Communication	**0.760**	0.155	4.9	0.000	0.456	1.064
Convenience of Product Return	**1.177**	0.068	17.27	0.000	1.044	1.311
Source of Recall	**0.556**	0.069	8.03	0.000	0.420	0.692
Hazard Degree	**-0.146**	0.066	-2.22	0.026	-0.275	-0.017
Hazard Probability	**-0.263**	0.135	-1.94	0.052	-0.528	0.002
IASocR x HazD	**-0.357**	0.101	-3.54	0.000	-0.555	-0.160
IATrans x HazP	**-1.373**	0.178	-7.72	0.000	-1.722	-1.025
IASocR x SoR	**0.297**	0.065	4.6	0.000	0.170	0.424
constant	**4.447**	0.412	10.78	0.000	3.639	5.256
Perceived Product Quality (R^2 = .402)						
Social Responsibility	**-0.412**	0.078	-5.3	0.000	-0.565	-0.260
Transparency of Communication	**0.991**	0.155	6.41	0.000	0.688	1.294
Convenience of Product Return	**1.341**	0.068	19.74	0.000	1.207	1.474
Source of Recall	**0.340**	0.069	4.94	0.000	0.205	0.476
Hazard Degree	-0.069	0.065	-1.05	0.295	-0.197	0.060
Hazard Probability	-0.172	0.135	-1.28	0.202	-0.436	0.092
IASocR x HazD	**-0.310**	0.100	-3.08	0.002	-0.506	-0.113
IATrans x HazP	**-1.826**	0.177	-10.3	0.000	-2.173	-1.479
IASocR x SoR	0.029	0.064	0.46	0.648	-0.097	0.156
Constant	**3.557**	0.411	8.66	0.000	2.752	4.362

Figure 2: Interaction Effect Between Social Responsibility and Hazard Degree in Case of Brand Image

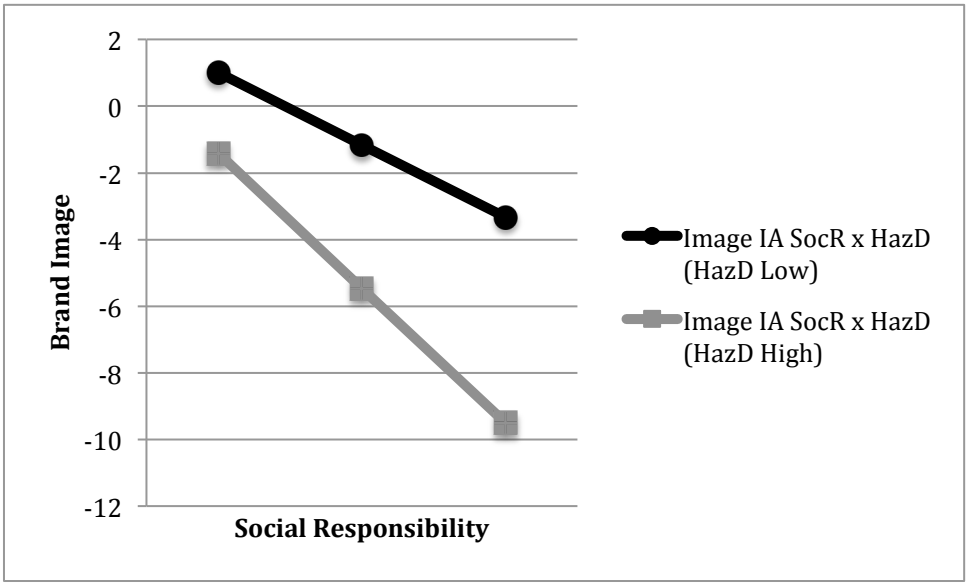


Figure 3: Interaction Effect Between Social Responsibility and Hazard Degree in Case of Brand Attitude

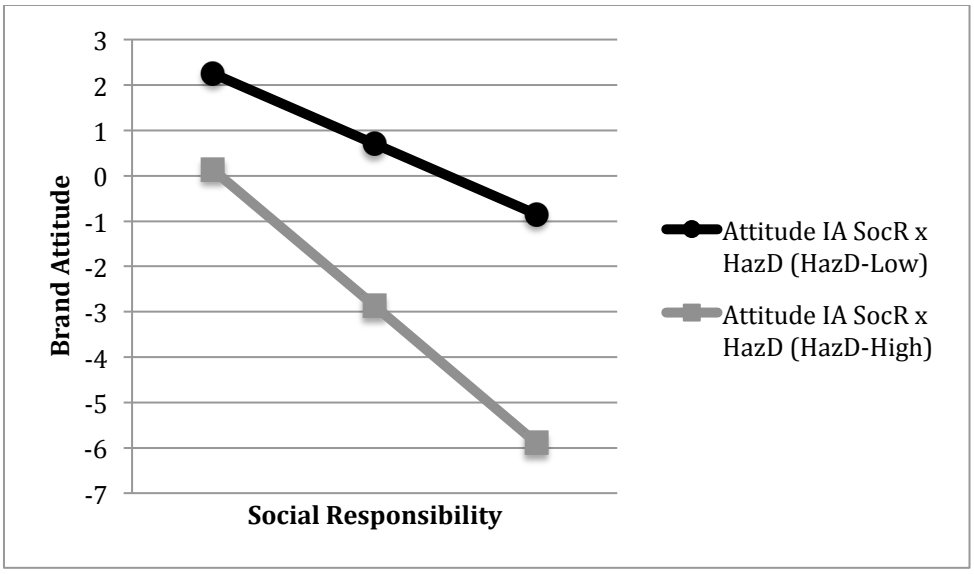


Figure 4: Interaction Effect Between Social Responsibility and Hazard Degree in Case of Perceived Product Quality

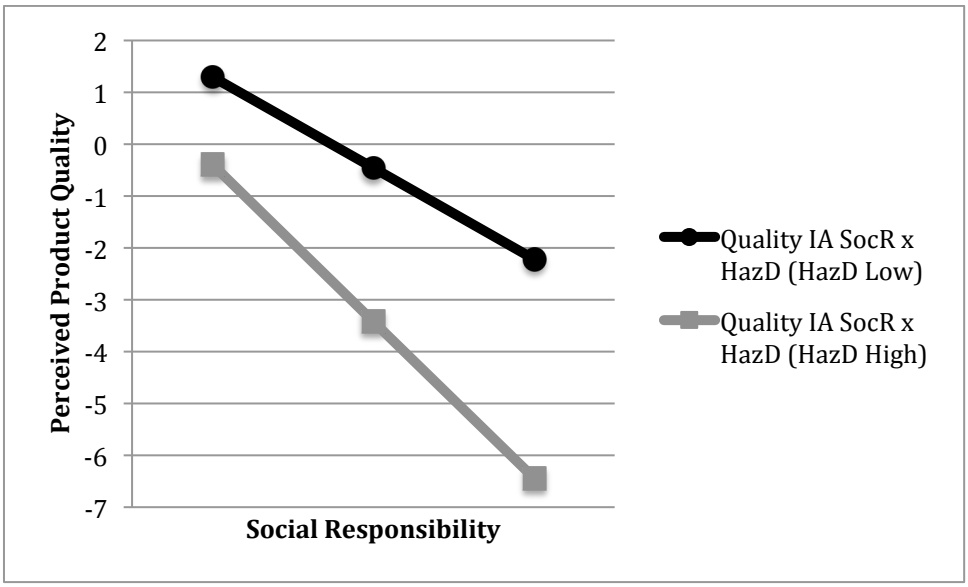


Figure 5: Interaction Effect Between Transparency and Hazard Probability in Case of Brand Image

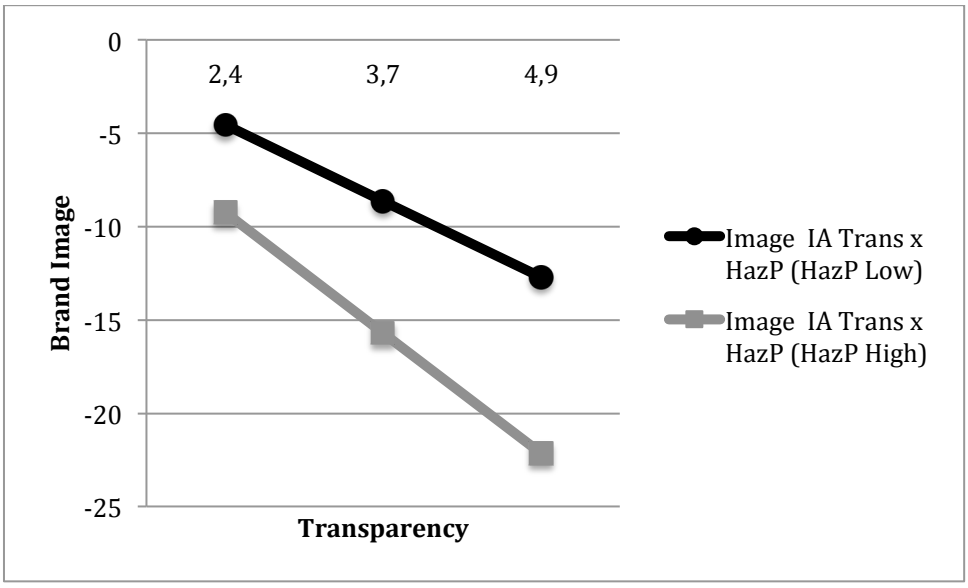


Figure 6: Interaction Effect Between Social Responsibility and Hazard Degree in Case of Brand Attitude

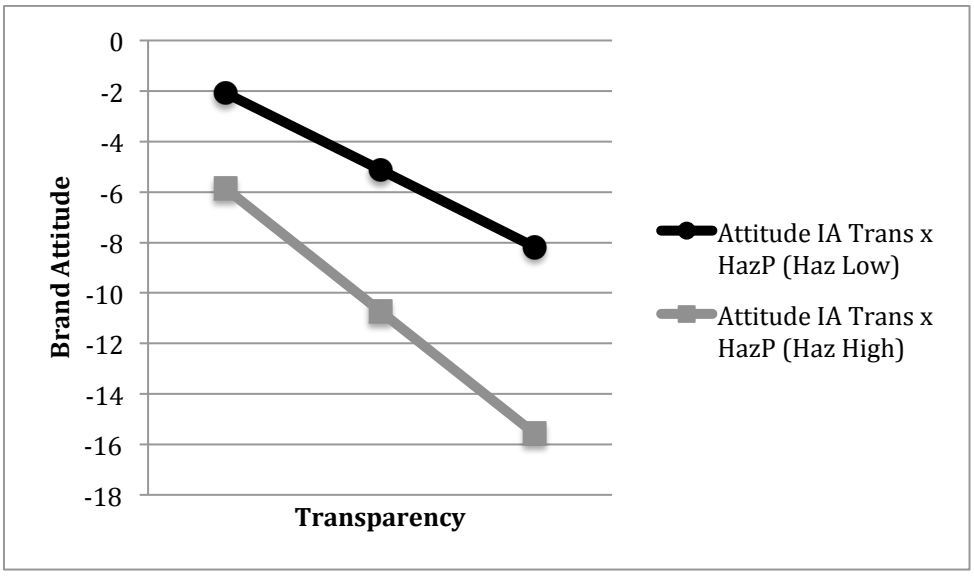


Figure 7: Interaction Effect Between Social Responsibility and Hazard Degree in Case of Perceived Product Quality

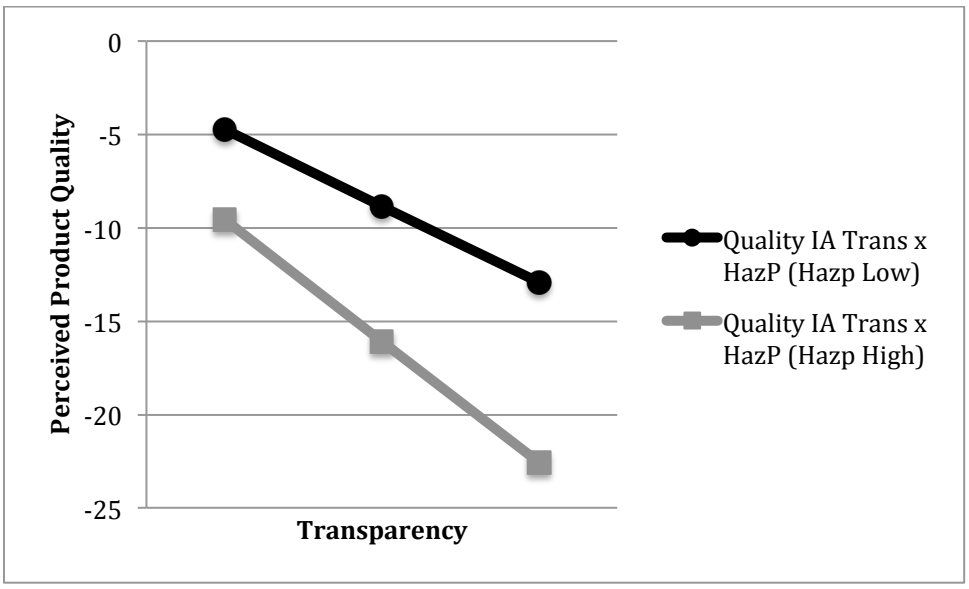


Figure 8: Interaction Effect Between Social Responsibility and Source of Recall in Case of Brand Image

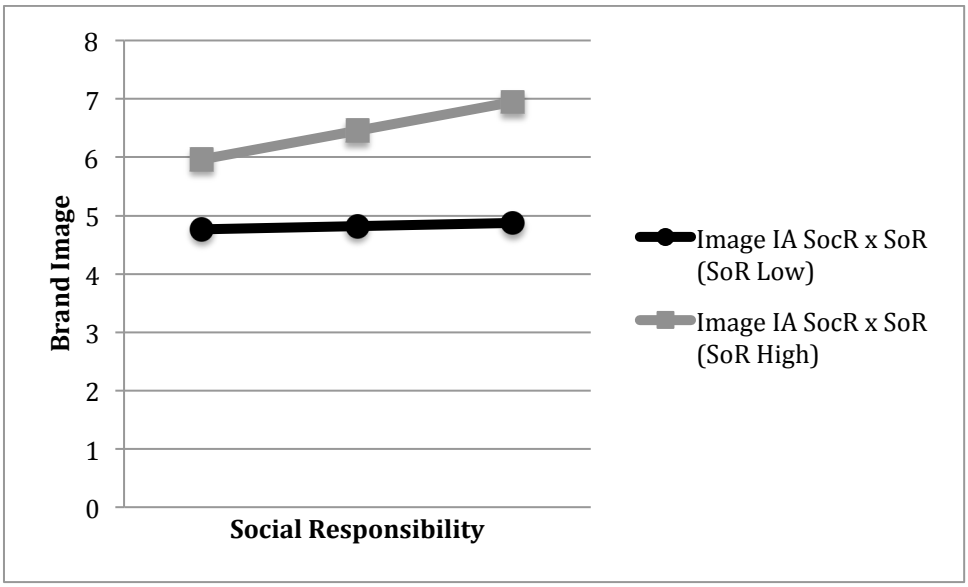
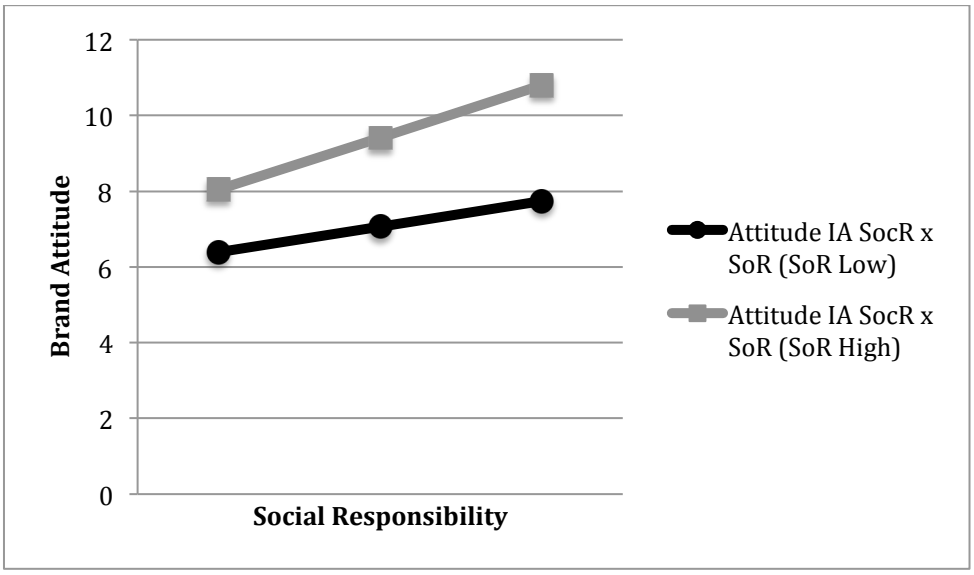


Figure 9: Interaction Effect between Social Responsibility and Source of Recall in Case of Brand Attitude



6 Discussion

The outcome of the independent-sample t-test clearly shows that people tend to judge a brand of less quality and inferior image, when a company has to recall a defective product. Therefore, from a company's perspective it is crucial to know which recall communication strategy will diminish the negative effect on brand assessment. The majority of publicly aired recall messages contains information about the communicated social responsibility of the affected company. The negative estimate for such social responsibility indicates that a corporate excuse apparently makes consumers wish to punish the affected company. This result might be perceived as counterintuitive as people in general tend to reward social responsible behavior. But in line with the former findings of Folkes (1984), we are able to show that the reverse relationship applies. Consumers may understand the over-emphasizing of social responsibility as some kind of guilt rejection or even a corporate attempt of excuse. Therefore, companies should abstain from communicating high levels of social responsibility in recall messages as consumers tend to misperceive that. In contrast, by rendering high levels of transparency and offering convenience of product return, a company can reduce or even avoid possible draw-backs of recall communication.

The degree of hazard caused by defective products, taken for itself, has no significant impact on brand evaluation at all. Taking into account interaction effects, both, an increasing degree of hazard and a communicated high degree of social responsibility lead to a significant and positive effect on marketing metrics. It thus appears that both context variables attenuate the negative impact of social responsibility, indicating that consumers – only being overwhelmed by a risky and dangerous situation appreciate social responsibility statements. This result contradicts previous empirical findings from different experimental studies (e.g. Mowen, Jolly, and Nickell 1981 or Vassilikopoulou, Chatzipanagiotou, Siomkos, and Triantafillidou 2011), which stated, that social responsibility statement always lead to a better evaluation of the recalling company. Also contrary to the extant knowledge stated above, transparency not only

fosters consumers' brand evaluation but also possibly diminishes it if the probability of hazard due to defective products is increasing. Therefore, too much information about the product defect and its consequences will over emphasize the threat by defective products if people are aware of a high probability of risks. As shown in table 5, the source of recall in terms of whether a defective product is communicated by companies' press releases or via public media has a significant effect on brand evaluation. Even when it is beyond the reach of corporate influence, companies will be well served by integrating public media and authorities to emphasize the reliability and sincerity of recall messages. This effect will be intensified by high degrees of social responsibility such as emphasizing safety pattern as shown above. Hence, companies emphasizing customer safety together with high safety standards in recall messages are able to gain an advantage if the source of identification is beyond their scope.

7 Conclusions, Outlook and Future Research

For the first time in marketing research our study investigates the impact of different communication strategies (stated degree of social responsibility, transparency and convenience of product return) as well as certain situational variables (source of recall, degree and probability of hazard) onto consumers' brand evaluation. Our findings provide insight in how to design recall messages and thus help marketers to reduce the risk of long-lasting image losses. Taking this into account our study may also encourage companies to communicate a recall to a broader audience. This again may help to secure higher levels of consumer protection, as more consumers will get aware of the affected products and the related risks.

Our results indicate that companies can reduce the negative impact of recall messages on customers' brand evaluation by taking into account various effects of different communication strategies and situational factors. As shown in the former paragraphs, the results differentiate between the use of single and combined variables in terms of their impact on customers' brand evaluation. In addition, the integration of the several interaction effects allow us to derive more generalizable implications that allow

managers in different situations – in terms of e.g. different degrees or probabilities of hazard – to design the best possible and most harm-minimizing communication strategy.

While most of the recall related research focuses on the impact of product recalls on short-term performance such as sales (Folkes, 1984; Mowen, Jolly, and Nickell, 1981), our study investigates for the first time the effect of different communication strategies as well as situational variables on brand evaluations that are considered long-term marketing metrics. Of course our study is - due to the cross-sectional character of our data - not able to provide insights into the occurring dynamic effects during a recall scenario. Thus to gather more knowledge on time-depending effects, future research should more rely on a longitudinal approach and try to measure how persistent the different communication induced effects are and how long customers really remember a product recall. Such information would allow managers to better plan recall campaigns, as they would know when to stop communicating the product defect.

In addition, our study shows for the first time, that the impact of the different elements of the recall communication is significantly moderated by the media-channel in which the message is published. Thus, companies should better cooperate early with public media to ensure a positive coverage. Taking into account that press professionals are believed to be more suspicious and informed than ordinary consumers, future research should also examine how these persons do react to recall communication and how their reaction is moderated by the different situational components. Such knowledge would certainly encourage more marketers to liberally communicate any defect issues to a broader audience and thus in the long-term help to ensure higher degrees of consumer protection and public welfare.

References:

Aiken, Leona S. and Stephen G. West (1991): *Multiple Regression: Testing and Interpreting Interactions*, Sage Publactions: Newbury Park.

Beamish, Paul W. and Hari Bapuji (2008): Toy Recalls and China: Emotion vs. Evidence, *Management and Organization Review*, 4 (2): 197-208.

Belsley, David A., Edwin Kuh and Roy Welsch (1980): *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, Wiley: New York.

Cheah, Eng, Wen Chan and Corinne Chieng (2007): The Corporate Social Responsibility of Pharmaceutical Product Recalls: An Empirical Examination of U.S. and U.K. Markets, *Journal of Business Ethics*, 76 (4): 427-449.

Chen, Yubo, Shankar Ganesan and Yong Liu (2009): Does a Firm's Product-Recall Strategy Affect Its Financial Value? An Examination of Strategic Alternatives During Product-Harm Crises, *Journal of Marketing*, 73 (6): 214-226.

Cleeren, Kathleen, Marnik G. Dekimpe and Kristiaan Helsen (2008): Weathering Product-Harm Crises, *Journal of the Academy of Marketing Science*, 36 (2): 262-270.

Coleman, Les (2011): Losses from Failure of Stakeholder Sensitive Processes: Financial Consequences for Large US Companies from Breakdowns in Product, Environmental, and Accounting Standards, *Journal of Business Ethics*, 98 (2): 247-258.

Coombs, Timothy (1998): An analytic Framework for Crisis Situations: Better Responses From a Better Understanding of the Situation, *Journal of Public Relations Research*, 10 (3): 177-218.

CPSC (2005): US Consumer Product Safety Commission 2005 Performance and Accountability Report, Source: http://www.cpsc.gov/cpscpub/pubs/reports/2005rpt.pdf.

Dawar, Niraj and Madan M. Pillutla (2000): Impact of Product-Harm Crises on Brand Equity: The Moderating Role of Consumer Expectations, *Journal of Marketing Research*, 37 (2): 215-226.

de Matos, Celso Augusto and Carlos Alberto Vargas Rossi (2007): Consumer Reaction to Product Recalls: Factors Influencing Product Judgement and Behavioural Intentions, *International Journal of Consumer Studies*, 31 (1): 109-116.

Folkes, Valerie S. (1984): Consumer Reactions to Product Failure: An Attributional Approach, *Journal of Consumer Research*, 10 (4): 398-409.

Jolly, David W. and John C. Mowen (1985): Product Recall Communications: The Effects of Source, Media and Social Responsibility Information, *Advances in Consumer Research*, 12 (1): 471-475.

Kübler, Raoul V. and Sönke Albers (2010): Communication Behavior of Companies in Product Recalls Without Customer Identification Information, *Marketing - Journal of Research an Management*, 6 (1): 59-69.

Lin, Chieh Peng, Shwu Chuan Chen, Chou Kang Chiu and Wan Yu Lee (2011): Understanding Purchase Intention During Product-Harm Crises: Moderating Effects of Perceived Corporate Ability and Corporate Social Responsibility, *Journal of Business Ethics*, 102 (3): 455-471.

Low, George and Charles Lamb (2000): The Measurement and Dimensionality of Brand Associations, *Journal of Product & Brand Management*, 9 (6): 350-356.

McGraw, Kenneth O. and Steven Wong (1996): Forming Inferences About Some Intraclass Correlation Coefficients, *Psychological Methods*, 1 (1): 30-46.

Mowen, John C. (1979): Further Informaion on Consumer Perceptions of Product Recalls, *Advances in Consumer Research*, 7 (1): 519-523.

Mowen, John C., David Jolly and Gary S. Nickell (1981): Factors Influencing Consumer Responses to Product Recalls: A Regression Analyses Approach, *Advances in Consumer Research*, 8 (1): 405-406.

Rhee, Mooweon and Pamela R. Haunschild (2006): The Liability of Good Reputation: A Study of Product Recalls in the U.S. Automobile Industry, *Organization Science*, 17 (1): 101-117.

Simola, Sheldene (2003): Ethics of Justice and Care in Corporate Crisis Management, *Journal of Business Ethics*, 46 (4): 351-361.

Siomkos, George and Paul Shrivastava (1993): Responding to Product Liability Crises, *Long Range Planning*, 26 (5): 72-79.

Van Bruggen, Gerrit H., Gary L. Lilien and Manish Kacker (2002): Informants in Organizational Marketing Research: Why Use Multiple Informants and How to Aggregate Responses, *Journal of Marketing Research*, 39 (4): 469-478.

Van Heerde, Harald, Kristiaan Helsen and Marnik G. Dekimpe (2007): The Impact of a Product-Harm Crisis on Marketing Effectiveness, *Marketing Science*, 26 (2): 230-245.

Vanhamme, Jolle and Bas Grobben (2009): Too Good to be True: The Effectiveness of CSR History in Countering Negative Publicity, *Journal of Business Ethics*, 85 (2): 273-283.

Vassilikopoulou, Aikaterini, Apostolos Lepetsos, George Siomkos and Kalliopi Chatzipanagiotou (2009): The Importance of Factors Influencing Product-Harm Crisis Management Across Different Crisis Extent Levels: A Conjoint Analysis, *Journal of Targeting, Measurement & Analysis for Marketing*, 17 (1): 65-74.

Vassilikopoulou, Aikaterini, Kalliopi Chatzipanagiotou, George Siomkos and Amalia Triantafillidou (2011): The Role of Consumer Ethical Beliefs in Product-Harm Crises, *Journal of Consumer Behaviour*, 10 (5): 279-289

D Faking or Convincing: Why Do Some Advertising Campaigns Win Creativity Awards?

Raoul V. Kübler and Dennis Proppe

Published in:
Kübler, Raoul V. and Dennis Proppe (2012): Faking or Convincing: Why Do Some Advertising Campaigns Win Creativity Awards?, *BuR Business Research*, 5(1), Available at: http://www.business-research.org/early_view/kuebler-proppe-faking-early.pdf

Web Appendices:

1. Exemplary for Fake Campaigns
2. List of all Print/Guerrilla/Ambient/New-Media advertisements in the Sample
3. List of all TV/Cinema/Radio advertisements in the Sample
4. Exemplary Online-questionnaire for print campaigns
5. Exemplary Online-questionnaire for tv campaigns

Abstract: Since the Sarbanes-Oxley Act was passed in 2002, it has become commonplace in the advertising industry to use creativity-award-show prizes instead of gross income figures to attract new customers. Therefore, achieving a top creativity ranking and winning creativity awards have become high priorities in the advertising industry. Agencies and marketers have always wondered what elements in the advertising creation process would lead to the winning of creativity awards. Although this debate has been dominated by pure speculation about the success of different routines, approaches and strategies in winning creativity awards, for the first time our study delivers an empirical insight into the key drivers of creativity award success. We investigate what strategies and which elements of an advertising campaign are truly

likely to lead to winning the maximum number of creativity awards. Using a sample of 108 campaigns, we identify factors that influence campaign success at international advertising award shows. We identify innovativeness and the integration of multiple channels as the key drivers of creativity award success. In contrast to industry beliefs, meaningful or personally connecting approaches do not seem to generate a significant benefit in terms of winning creativity awards. Finally, our data suggest that the use of so-called "fake campaigns" to win more creativity awards does not prove to be effective.

Key words: Advertising Agencies, Advertising Awards Shows, New Business, Creative Success, "Fake Campaigns," Partial Least Squares

1 Introduction and Research Concept

1. Introduction and Research Concept

Being in the top 10 in various agency ranking systems has become high priority for many advertising agencies. This desire is explained primarily by the understanding that major brands and marketers that are looking for new agencies often pre-select their candidates according to the placement of an ad agency in various ranking systems (Butkys and Herpel 1992, Helgesen 1994). This type of pre-selection is based on an old tradition in the advertising industry and was reported as early as the 1960s (Ogilvy 1963). Since then, the industry has developed several ranking systems, which judge agencies based on income figures, the overall success of an agency at award shows such as the Cannes Lions or the One Show, or a variety of individual expert evaluations (e.g., Gross Income Rankings, National Creativity Rankings, or subjective agency rankings developed by individual marketing managers of leading brands).

Until 2001, the vast majority of rankings was based on various income figures such as gross income or accumulated media volume of an agency, under the assumption that these figures clearly reflect the competence of an agency. In the wake of the Enron and WorldCom scandals, the US government enacted the so-called Sarbanes-Oxley Act (SOX) in 2002. This legislation increased investor protection and, under the threat of a jail sentence (up to 20 years for managers who were found to be responsible), banned companies from publishing incorrect information (Sarbanes-Oxley-Act 2002). With these new regulations in place, major networks were afraid that individual subsidiaries might provoke severe consequences for the parent company if they independently published incorrect or unverified figures. Thus, all agencies in the network were forced to first (or only) report their income figures to their particular parent company and to stop publishing income-related figures, for example, for ranking purposes. Up to that point, most national income rankings had been based on disaggregated and voluntary information on a national subsidiary level. Since 2002, and as a consequence of

increasing concerns of the holding companies, income figures became available only on an aggregated level and could no longer be used for national rankings.

Because of the lack of reliable individual income figures for the individual subsidiaries and the national agencies, the industry shifted its focus more toward creativity (Myers 2004). This shift explains why creativity rankings have gained importance in agency pre-selection. Thus, winning creative awards and achieving top results in creative rankings have become the key promotion tools, and the industry today also uses increasing amounts of resources for participation in creativity award shows (Wentz 2005) and strives for top positions in annual national creative rankings (e.g., the Ad Age ranking (Ad-Age-Ranking 2009) in the US or the W&V/Horizont ranking (W&V-Kreativ-Ranking 2009) in Germany).

In general, these rankings are based on the annual overall results of an advertising agency in the most prestigious international award shows, such as the Cannes Lions, the Clio Awards, the One Show, the European ADC Awards and a group of various content-specific or national award shows. The rise in importance of winning awards in these shows has led to a shift in agency behavior. As winning creativity awards became a major topic in the industry, agencies started to implement different routines to maximize their number of creativity awards. In addition to making serious financial efforts to attract the industry's top creative people, some agencies even started to compete by offering years of free work to clients, who were known to approve the production of more creative ads, if they switched agencies. These offers clearly underline the importance to advertising agencies of award shows and creative reputation.

Most agencies believe that their clients are less creative or that they restrict the agencies' creative competence by their strategic concerns or risk aversion (El-Murad and West 2003). To showcase their creative skills, some copywriters and art directors started to develop campaigns for "alibi" or fictitious clients. As another strategy, agencies at times try to maximize their outcomes at award shows by submitting

advertisements not only for their real clients but also for alibi clients. Producing advertising for fictitious clients allows agencies to avoid daily business restrictions or long discussions with clients or product managers, who may want to replace creative concepts with less creative approaches (West 1993).

In most award shows, submissions must be aired or published only once to qualify, so agencies started to hire pro-bono clients such as bakeries or pet shops to fulfill the minimum show requirements for fictitious campaigns. To be able to submit these so-called fake campaigns, agencies mostly work for free and even pay for the one-time airing fees.

Today, copywriters and graphic artists continue to speculate as to which factors and elements in the process of idea generation and ad planning, conceptualization, design, and execution play decisive roles in winning creativity awards.

Furthermore, the knowledge that agencies try to maximize their success at award shows leads to the question of whether creativity award shows or creative rankings can successfully reflect the abilities and competence of an advertising agency. In other words, what does award-show success portray? Does winning creativity awards indicate real market competence and an ability to combine the individual communication requirements of a client with the creative competence of the agency, or does it only indicate the artistic feeling for great ideas reflected in outstanding but fake campaigns without any business background?

In summary, one can observe that agencies invest significant time and human and capital resources for the sole purpose of winning awards. Given the fact that marketers currently tend to pre-select agencies according to their success at award shows, to the extent that winning creative awards is crucial for commercial success, three important questions arise:

> First, which factors and strategies in the ad development process result in campaigns and agencies winning creativity awards?

Second, what exactly does winning creativity awards reflect? In other words, does working creatively guarantee winning creativity show awards?

Answering the first and the second questions might also simultaneously deliver the answer to a third question: Are creativity rankings the appropriate instrument for marketers to use to pre-select agencies?

2. Research Gap and Contribution

Creativity research has a long and fruitful tradition in psychology and marketing. Whereas most studies from the marketing side were oriented more toward effectiveness, psychological studies focused more on defining and measuring creativity (Haberland and Dacin 1992). The early field of psychological studies can be divided into three different approaches: The first – and earliest – approach aims at identifying the traits of creative people (Barron 1955, Barron and Harrington 1981, MacKinnon 1987). Creativity in these studies was based mostly on three main characteristics: originality, the adaptiveness to reality and development and elaboration.

The second approach of fundamental psychological studies focused on process-oriented factors, which may influence or enhance creativity or stages of the creative process (Bruner 1962, Newell, Shaw, and Simon 1962). In these studies, the fundamental concept of "novelty" was introduced as a key characteristic of creativity. Finally, the third and smallest area of psychological studies attempted to identify the particular characteristics of creative outcomes (Besemer and O'Quin 1986, Besemer and Treffinger 1981). Following this school of thought, Jackson and Messick (1965) introduced a multi-item approach that identifies four underlying dimensions of creativity: unusualness, appropriateness, transformation and condensation. Finally, Amabile (1983) started a new development in psychological creativity research by focusing on creativity-enhancing factors in terms of working atmosphere (Amabile, Conti, Coon, Lazenby, and Herron 1996) or employee development (Mumford and Simonton 1997).

Although earlier studies relied on various items to measure the specific creativity aspects, the study by Amabile and later studies returned to a simple measurement approach, which asks participants to directly judge overall creativity. Relying on the different results of the three approaches, further psychological studies took a closer look at the different outcomes of creativity. Thereby, the focus lay primarily on the influence of creativity on employee development (Andrews and Smith 1996), better working atmosphere (Amabile, Tighe, Hill and, Hennessey 1994), better sales (Moorman and Miner 1997, Song and Montoya-Weiss 2001) and enhanced product or corporate performance (Deshpandé, Farley, and Webster 1993).

In terms of marketing research, creativity is addressed from two perspectives: The first school of thought focuses on the impact of creativity on product development (Sethi, Smith, and Park 2001) and new product performance (Im and Workman 2004, Moorman and Miner 1997). These studies relied heavily on the results from the various psychological studies mentioned above. The second approach to marketing research in terms of creativity focuses more on the impact of creativity on marketing performance and, in particular, on creativity in advertising. This research area can also be divided into two different approaches. Whereas the first group of studies attempted to identify specific factors that may influence or produce the perceived creativity of an advertisement, the other group is oriented more toward output and questions particularly the efficiency of creative advertising. The first group focused mainly on the factors that may help increase advertising creativity (Ang, Lee, and Leong 2007, Kilgour and Koslow 2009, Pieters, Warlop, and Wedel 2002). This focus may lead to a better understanding of how customers perceive advertising creativity (Haberland and Dacin 1992, Michell 1984) and may explain the differences in assessing advertising creativity between agency workers (Nyilasy and Reid 2009) and their clients (Devinney, Dowling, and Collins 2005, Koslow, Sasser, and Riordan 2003) and the creative differences between copywriters and art directors (Young 2000).

Another area of advertising-related creativity research focused on the positive impact of advertising creativity on advertising effectiveness in general (Rossiter 2008) and on

sales gains in particular (Bell 1992, Bogart, Tolley, and Orenstein 1970, El-Murad and West 2003, Smith, MacKenzie, Yang, Buchholz, and Darley 2007). In addition to these effectiveness-oriented advertising studies, research also addressed other outcomes of creativity in terms of market share gains (Buzzell 1964), differences in long- and short-term impacts of creative advertising (Jones 1995), increased brand recall (Stone, Besser, and Lewis 2000), better brand attention (Pieters, Warlop, and Wedel 2002), greater brand awareness (Gibson 1996), the link between ad likeability and sales performance (Bergkvist and Rossiter 2008), better persuasion rates (Till and Baack 2005) and an enhancement of the perceived brand quality (Dahlén, Rosengren, and Törn 2008). Finally, research explored the positive effects of advertising creativity on brand image and trust enhancements (Hairong, Wenyu, Guangping, and Nan 2008).

Like most product development studies, the majority of this communication-oriented research operationalized creativity using items or measurement approaches from basic psychological studies. Although this measurement approach seems to be well established and appropriate for use, other studies on creativity efficiency can be found that operationalize "creative advertising" by defining it as "advertising that won creative awards" (Csikszentmihalyi 1998, El-Murad and West 2003, Kover, Goldberg, and James 1995, Saffert and Reinartz 2011). Although this approach seems to be well accepted, it is surprising that, to the best of our knowledge, no study has yet explored whether this operationalization is correct. Hence, if winning creativity awards is influenced not only by the particular degree of creativity used but also by other – situational or strategic – factors such as personal connections with award-show juries or the ability to produce fake campaigns of the respective agency, one must be aware that such an operationalization might be biased by these non-observable factors. Therefore, our study will not only identify different factors that will help agencies win creativity awards, but also clarify whether creativity awards can simply be used as an easy, appropriate and, in particular, *tested* measure for creative advertising.

The contributions of this study are as follows: First, we theoretically derive a novel set of factors that influence the success of an individual campaign in a creativity award

show allowing us to find an answer to our first question. Second, we test the influence of several strategic factors and routines on creativity-award-show success. The results reveal which individual factors, elements, and routines in the ad-creation process drive the overall creativity-award-show outcomes and may help us to find answers to our second and third research questions. Our results will show agencies which of these factors should be enforced and which factors can be neglected and will thus help agencies to maximize their award-show outcome.

In the following section, we develop a conceptual model and identify hypotheses regarding the impact of creative, executional (e.g., the layout, the approach or the choice of the media channels) and organizational factors on award-show performance. We aim to answer the three research questions by using data from 108 campaigns of more than 40 international and highly ranked award shows. In the third section, based on our empirical findings we develop several implications for ad agencies to improve their individual creativity-award-show outcomes and to maximize the individual placement in annual creativity rankings. These implications should help agencies to more effectively attract new business and new clients. At the end of our paper, we provide suggestions for further research.

3. Conceptual Framework and Hypotheses

Provided that winning awards and being a leader in creative business rankings is an important reference tool for agencies to promote their competence, academic studies should also focus on the success factors of agencies in these award shows. Our study defines creative success as being successful in creativity award shows. A top position in a creativity ranking can be secured only by winning as many awards as possible. To identify the key drivers of this creative success, we introduce a conceptual model as depicted in Fig. 1.

To develop our research hypotheses toward the question of which elements, factors or routines help win creativity awards, we review a wide range of literature addressing

creativity and advertising. Thus, we rely on two different sources of literature and theories. First, hypotheses will be derived from empirically oriented academic literature focusing on factors that influence creativity in general, but not on winning award shows. Second, we also consider more practice-oriented publications such as educational books for copywriters or art directors, which deal in a more detailed way with the topic of advertising and award shows, but may lack solid empirical justification.

To answer to our first research question, i.e., whether creativity is actually a key driver for success in creativity award shows, we first examine the psychological understanding of creativity. According to this understanding, creativity in general consists of three key elements (Im and Workman 2004): *newness or novelty* (Amabile 1983, Amabile 1988, Amabile 1993), *meaningfulness* (Goldenberg, Mazursky, and Solomon 1999), and *connectedness or originality* (Sternberg and Lubart 1995).

In terms of advertising, novelty can be defined as a fundamental breaking away from existing schemes or routines (Ang, Lee, and Leong 2007, Ang 2000). This process leads to a significant change or adaptation in the cognitive structure of the viewer and will help to enhance the attentive process and the memory effect. Earlier studies showed that a creative ad has to use *innovative*, *fresh*, *unique, different* and *new* ways to break from pre-existing schema and to deploy a sense of the unexpected (Bogart, Tolley, and Orenstein 1970, Haberland and Dacin 1992). These findings are supported by the empirical results of Ang and Low (2000), who found evidence for a positive correlation between the degree of innovation and the overall creative appearance of an ad. According to the most prominent textbooks for copywriters or art directors, newness or innovativeness in creative advertising can be expressed by the degree of *layout innovation*, the *innovativeness of the overall approach* of a campaign, or at least by the *innovativeness of media usage* (Aitchinson 1999, Sullivan 2003).

Thus, it can be assumed that the use of novel ideas, new layouts, and innovative concepts plays an important role in creative success and will lead to more creative

awards (Ang and Low 2000, Blasko and Mokwa 1986). Summarizing the aspects of novelty in the ad-creation process leads to our first hypothesis:

> H1: Campaigns that use a more novel or innovative approach are more successful in creativity award shows than campaigns that do not use such an approach.

Previous research noted that novelty alone might not be sufficient to define creativity (Ang, Lee, and Leong 2007, Baack, Wilson, and Till 2008, Haberland and Dacin 1992). Andrews and Smith (1996), therefore, introduced the additional concept of meaningfulness. This aspect of creativity is defined as the extent to which an idea deviates in a *meaningful* way from ordinary praxis. In terms of advertising creativity, meaningfulness can be understood as the link between an idea and the *relevance* of this idea to its audience (Lee and Mason 1999). In addition, Goldenberg, Mazursky, and Solomon (1999) emphasized that random advertising creativity without any relevance may be harmful at worst and inefficient at best. These assumptions were also made in the study by Smith, MacKenzie, Yang, Buchholz, and, Darley (2007). Advertising creativity might thus be understood as the combination of *relevance* and divergence. In terms of advertising creation, this definition implies that the ad not only has to rely on a novel communication approach (that is divergent from existing forms of communication), but the message also has to convey product-related information relevant to the customer. This means that the overall idea, the message, and the information about the advertised product must convey a singular message that is relevant to the targeted customer (Ang, Lee, and Leong 2007, Baack, Wilson, and Till 2008, Smith, MacKenzie, Yang, Buchholz, and Darley 2007).

Later research on advertising creativity defined meaningfulness as the central idea or theme of an advertisement (Thorndyke 1977). *Meaningfulness* can therefore be defined as the relevant *key benefits* on which the specific product is positioned in the marketplace (Haberland and Dacin 1992). In other words, meaningfulness is secured when the key message of an ad can rely on the *clear, appropriate and non-replaceable benefits* (Runco and Charles 1993) of the advertised product. Therefore, advertisements for products with these non-replaceable benefits perform better in creativity award shows because they can create original, unique, and new campaign approaches that will secure meaningfulness. The practice-oriented

advertising literature confirms this effect (Aitchinson 1999, Sullivan 2003). Thus, we hypothesize the following:

> H2: Campaigns that use a more meaningful approach for the creative implementation of an idea are more successful at winning creativity awards than campaigns that do not rely on such an approach.

The third and final criterion for defining advertising creativity is *connectedness* (Amabile, Conti, Coon, Lazenby, and Herron 1996, Ang, Lee, and Leong 2007, Im and Workman 2004). In terms of advertising creativity, connectedness can be understood as the ability of an advertisement to build an enduring link or a lasting connection between the viewer and the promoted product or brand (Dahlén, Rosengren, and Törn 2008). Previous research showed that such a connection can be best realized when a particular ad causes an *emotional* or *rational resonance* (Lubart 2000) in the consumer. This resonance should enforce the connection between the viewer and the advertisement's message or at least the advertised product (Sasser and Koslow 2008).

Previous research showed that this resonance can be best obtained when the creative idea is linked to a product specific and unique *benefit* that satisfies a certain *need* (Sternberg and Lubart 1995). Thus, information perceived as inappropriate or useless does not build a positive connection with its viewer and vice versa. These findings are in line with the suggestions of Keil (1975), who claimed that a creative advertisement should always follow a main strategy of connecting the product to a certain *benefit*. Again, the practice-oriented advertising literature confirmed this effect. That is, both, leading textbooks and agency guidelines for copywriters suggest that the search for a creative idea should always be based on the *benefit* of the respective product because only this *individual benefit* will lead to an enduring *connection* to the advertisement's audience (Aitchinson 1999, Sullivan 2003).

In summary, it can be assumed that if an advertisement is to win creative awards, it will have to connect to its audience by communicating a meaningful *product benefit* using a novel communication approach. This leads to our third hypothesis:

> H3: Campaigns that use a more connective approach as operationalized by the communication of a certain product benefit will perform better in creativity award shows than campaigns that do not rely on such an approach.

Although simply working creatively when designing an ad should result in better creativity awards performance, the authors suspect that agencies also use several other instruments and strategies to enhance their creativity-award-show outcomes. This assumption is underlined by the findings of Amabile (1997), who noted that people easily learn to adapt their creative behavior to reward systems. When rewarded, people stop trying to identify new and innovative ways to generate creative approaches or creative ideas and begin to rely on the rewarded and successful approach they have already developed. In the case of advertising creativity, this phenomenon indicates that creativity in award shows is bound to a small number of different creative patterns and is influenced by specifically developed strategies.

Transferring these assumptions to agencies, one must be aware that the industry has managed to establish several reward systems for winning creativity awards. Agencies, for example, reward their employees with bonuses for winning creative awards. In addition, in most agencies, the professional advancement of copywriters or art designers is closely related to their success in award shows.

As a consequence, one must assume that these different incentives and rewards programs will lead to the development of specific strategies to maximize the individual award-show outcomes. In the following section, we develop additional hypotheses concerning these additional variables, strategies and instruments.

In the last decade, integrated thinking has become an important issue in advertising (Kotler and Keller 2006, Naik and Raman 2003). Most clients today challenge their agencies to create ideas not for only one communication channel but for all possible channels. Thus, creative

people are forced to find the "one" idea that works for all channels. In multi-channel campaigns, the main creative idea has to be as flexible as possible if it is to succeed and needs to be adapted to all communication channels; we can thus assume that these multi-channel ideas will be judged as more creative (Aitchinson 1999). Therefore, we believe that work from multi-channel campaigns tends to fare better in creativity award shows (creative performance or award-show performance) than other campaigns (Johnson 2003). This trend is also reflected by the official calls of festival juries for an emphasis on more integrated campaigns. Thus, the authors assume the following:

> H4: Advertisements designed for a multi-channel campaign are more successful at creativity award shows than are advertisements designed for a single-channel campaign.

Although creativity is understood as a key element in attracting new clients, most agencies fear that their creative ability in daily business is restricted by the clients' strategy or the clients' overall creative competence, which is generally assumed to be lower than that of the agencies. In addition, several academic studies noted a serious divergence in the perception of creativity between agencies and their clients (El-Murad and West 2003, Helgesen 1994, Koslow, Sasser, and Riordan 2006, White 2003). These authors asked leading creative directors to explain the old industry belief and any discrepancy with the truth. They found that many marketers seem to fear that too much creativity may not be sufficiently product-oriented and will therefore be either misinterpreted or misunderstood by the customer. West, Kover, and Caruana (2008) arrived at similar findings in their study. Previous research showed that, in some cases, marketers even tend to interpret higher forms of creativity as a type of risk taking (El-Murad and West 2003). As a result, they are afraid that too much creativity might even hurt future sales or market share (Hairong, Wenyu, Guangping, and Nan 2008). In addition, clients are often believed by their own agencies to be more reluctant to take risks and to be less creative than the workforce of the agency (Kover, Goldberg, and James 1995). This phenomenon results in the belief that marketers may not have the

ability to identify the potential of a creative campaign or concept (Devinney, Dowling, and Collins 2005, Kover 1995).

Although marketers find creative agencies by pre-selecting them according to their rank, they are nevertheless said to be overburdened by the cutting-edge concepts of some copywriters. This dilemma puts agencies in a difficult situation: On the one hand, they need awards to attract new clients, but on the other hand current clients will likely not allow the agency, or only rarely, to work on a creative level that is sufficient to win further awards. As a result of this difficult situation, agencies sometimes stopped trying to convince their conservative clients to accept more creative or otherwise outstanding campaigns and rather started to produce cutting-edge concepts on their own without clients. To maximize their potential number of awards and to have the opportunity to showcase their unbounded creative potential, some agencies are even said to have complete campaigns, or at least some strong concepts, in reserve. The agencies then seek alibi clients that fit these concepts or campaigns. For the opportunity to submit these ideas to award shows, the agencies are sometimes willing to work for free and even to pay media fees. In general, typical pro-bono or "fictitious" clients for these concepts are small or local shops, such as bakeries or pet shops (Koremans 2007). Online Appendix A shows some typical fake campaigns developed by leading international agencies such as BBDO or DDB for small German bakeries and local charitable societies.

Because these campaigns are not restricted by the client's strategy objections, the cognitive creative potential of the marketer's management, or market and media restrictions, it can be hypothesized that fake campaigns allow a higher level of creativity than ordinary campaigns. Therefore, we postulate the following:

> H5: Campaigns created solely for creativity award shows ("fake campaigns") perform better in creative advertising award shows than ordinary campaigns created for real clients.

In addition to the abovementioned strategies in the ad-development process, agencies can use other strategies to maximize their creativity-award-show outcomes: Design, layout and conceptual elements can influence the number of creativity awards for a single ad as well as situational factors such as workplace atmosphere, routines or management techniques used by the designing ad agencies.

Numerous empirical studies focused on the influence of factors such as atmosphere, motivation, and leadership style on creativity (Amabile 1998, Amabile, Conti, Coon, Lazenby, and Herron 1996, Amabile, Tighe, Hill, and Hennessey 1994, Blasko and Mokwa 1986, El-Murad and West 2004, Koslow, Sasser, and Riordan 2003, Kover, William, and Sonner 1997, Ruscio, Whitney, and Amabile 1997).

As mentioned above, some agencies have begun to acquire special clients as well as special "creative" individuals who are well known for creativity-award-show success and outstanding creativity. The larger networks such as DDB, Jung von Matt and BBDO even started to integrate special departments into their networks to accommodate these experts. The only objective of these "creative task forces" is to create award-winning campaigns and to maximize the number of creativity awards won by the agencies. Because of cost effects, we also expect large and experienced companies to offer more attractive workplaces and more sophisticated management techniques (Aitchinson 1999). Furthermore, because of pathway effects (Barney 1991, Barney 2001), we expect older agencies to have established more creative routines than younger ones. Moreover, experienced industry managers with a strong reputation have suggested that a link between the age of an agency and its contacts with jury members in award shows might exist because these jury members are primarily recruited from experienced agencies or from associations such as the national Art Directors Club. Knowing jury members and their special preferences might also increase the overall performance of an advertisement in creativity award shows. Therefore, we integrate these ideas into the following three hypotheses:

> H6: Campaigns created by larger (a) and more experienced (b) agencies that have more contact with associations (c) are more successful in winning creativity awards than campaigns from smaller (a) and less experienced (b) agencies that have less contact with associations (c).

In addition to the mentioned direct effects of the different variables on creativity-award-show performance, we consider several indirect effects of these different variables. Below, we incorporate these indirect effects into the structural equation model.

As mentioned, the so-called fake campaigns are created solely for creativity award shows. By doing so, agencies attempt to demonstrate their creative skills without being restrained by real clients' wishes, briefings, or strategies. Therefore, it could be assumed that these campaigns will be based on stronger creative approaches than other campaigns. If this assumption is true, it must also be assumed that all factors that define creativity in general will also have an indirect impact on these fake campaigns. Therefore, we incorporate indirect effects from the variables "novelty," "meaningfulness" and "connectedness" in a "fake campaign" in the model. This leads to the following hypotheses:

> H7: Fake campaigns use more novel (a), more meaningful (b), and more connecting (c) approaches than other campaigns submitted to creativity award shows.

We also mentioned above that typical "*fake campaigns*" are aired only once in local media to fulfill the minimum submission criterion of creativity award shows. Given this and the fact that these special advertisements are based on a single concept, it must also be assumed that there is a negative indirect association between the variable "*multi-channel approach*" and the variable "*fake campaign.*" This leads to our eighth hypothesis:

> H8: Fake campaigns are less likely to be multi-channel campaigns than other campaigns submitted to creativity award shows.

In addition, older agencies will have developed better routines to maximize their creativity-award-show outcomes. One reason for this trend is that more experienced

agencies might have more resources to perform this extra work. Another reason might be that more experienced agencies with more connections to the award-show industry and their respective associations will also know how to place good "*fake campaigns*" without being blamed for doing so. We account for this advantage of older agencies by including further indirect effects from the variables "*agency age,*" "*agency size*" and "*association members*" in the variable "*fake campaign.*" Additionally and as a consequence of pathway effects, we assume that older agencies will be larger and will have more contacts with creative associations. We account for these effects by incorporating indirect effects from the variable "*agency age*" in the variables "*agency size*" and "*association members.*" This leads to our ninth hypothesis:

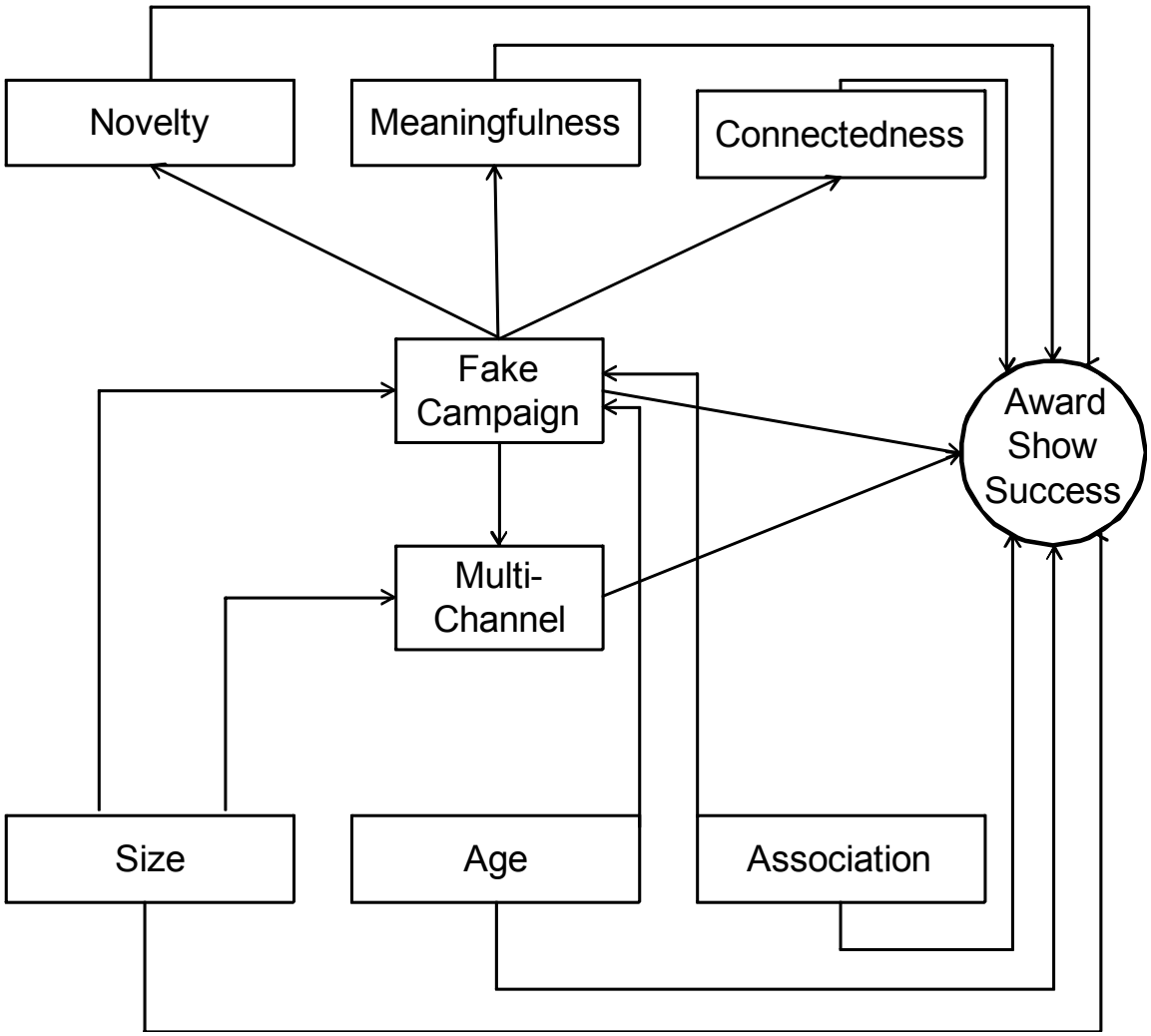
> H9: The use of fake campaigns is positively influenced by the age (a), size (b), and number of contacts with associations (c) of the agency.

In addition to fake campaign-related effects, we address agencies' capability to produce multi-channel campaigns. Thus, we assume that agency size may also have a positive influence on the ability of an agency to work on multi-channel campaigns. This assumption may be justified primarily by the fact that small agencies tend to specialize themselves for individual channels such as online marketing or below-the-line marketing, whereas larger agencies and network agencies try to gather all the aspects of marketing under one roof. We incorporate this assumption by an indirect effect from the variable "agency size" on the variable "*multi-channel approach.*" Thus, we derive our final hypothesis:

> H10: The use of a multi-channel approach is positively influenced by the size of an agency.

On the basis of this conceptual framework (see Fig. 1), we test our hypotheses (see Tab. 1 for a summary) to address the question of which of these factors drive the overall creativity-award-show outcomes of an agency.

Figure 1: Conceptual Model



4. Empirical Analysis

To empirically test our conceptual model, we collected data from 108 German campaigns from a total of 40 creativity award shows in 2006. Data collection was mainly divided into two parts: factual data gained by intense investigation and the results of an expert questionnaire that evaluated several campaign features. In the following section, we describe in detail our data sample and followed by a description of our expert survey and measurement development. The section closes with

specifications and explanations of our estimation process and a description of our empirical results.

Table 1: Summary of the Hypotheses

H_x	**Hypothesis formulation**
H1	Campaigns that use a more novel or innovative approach are more successful in creativity award shows than are campaigns that do not use such an approach.
H2	Campaigns that use a more meaningful approach to the creative implementation of an idea are more successful in winning creativity award than are campaigns that do not rely on such an approach.
H3	Campaigns that use a more connective approach by communicating a product benefit will perform better in creativity award shows than will campaigns that do not rely on such an approach.
H4	Advertisements designed for a multi-channel campaign are more successful at creativity award shows than are advertisements designed for a single-channel campaign.
H5	Campaigns created solely for creativity award shows ("*fake campaigns*" or "*gold ideas*") perform better in creative advertising award shows than do ordinary campaigns created for real clients.
$H6_{a-c}$	Campaigns created by larger (a) and more experienced (b) agencies that have more contacts with associations (c) are more successful in winning creativity award than are campaigns from smaller (a) and less experienced (b) agencies that have less contact with associations (c).
$H7_{a-c}$	Fake campaigns use more novel (a), more meaningful (b), and more connecting (c) approaches than other campaigns submitted to creativity award shows.
H8	Fake campaigns are less likely to be multi-channel campaigns than are other campaigns that are submitted to creativity award shows.
$H9_{a-c}$	The use of fake campaigns is positively influenced by the age (a), the size (b), and the number of contacts with associations (c) of the agency.
H10	The use of a multi-channel approach is positively influenced by the size of an agency.

4.1 Sample

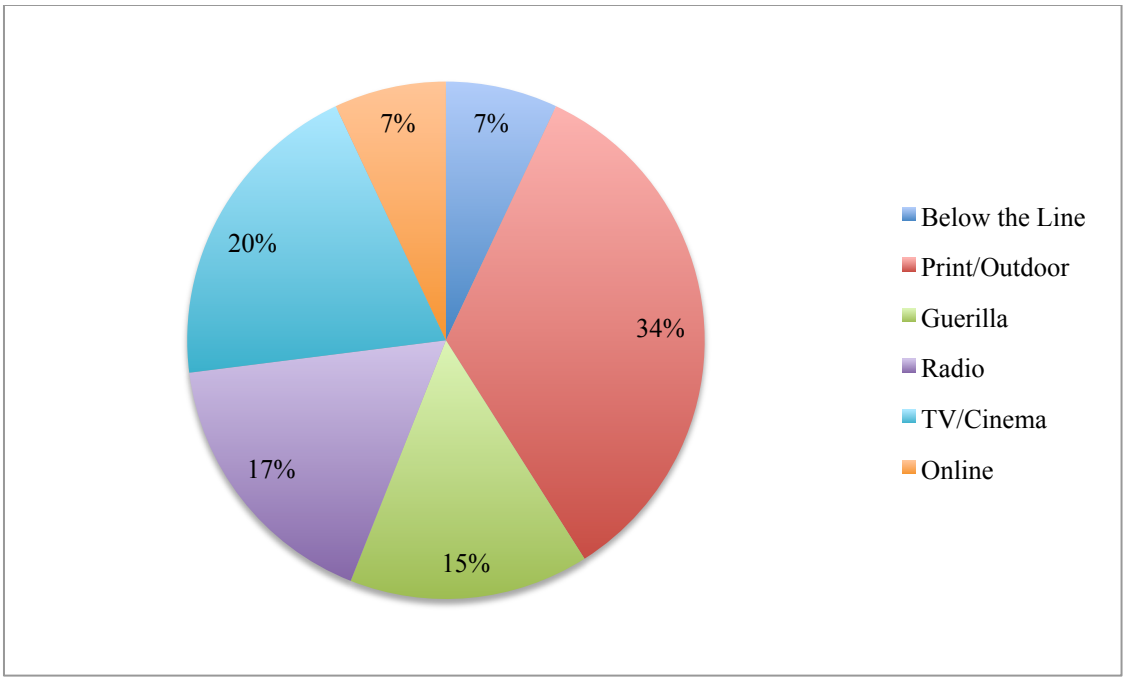
All campaigns in the sample were at least shortlisted in one of the 40 considered award shows. The award shows were chosen based on their importance for the national German creativity ranking. The award-show sample considered only rankings acknowledged by the German Art Directors Club – the most important advertising association in Germany – as a base for the national agency ranking. This approach promises unique data with high internal validity because juries in all award shows are

said to rely on identical criteria when judging submissions. These data were collected from online sources and with the help of the German Art Directors Club. All 108 campaigns in the sample can be watched and listened to in online Appendix B (Print/Ambient/Internet/BtL) and online Appendix C (Video/Cinema/Radio). Both appendices also show the creative performance of the individual advertisements and a separate fake index indicating the number of experts who judged the advertisement as a fake campaign. The sample consists mainly of print and outdoor advertisements, followed by the categories TV/cinema and radio (see Figure 2). Campaigns were submitted by 27 German advertising agencies ranging in size from 25 to 1,400 employees (see Table 2).

4.2 Expert Survey

To gather data regarding some of the campaign-specific features (e.g., innovativeness, whether fake or not and use of USP), the authors developed a web questionnaire and presented it to experts from the advertising industry (for further details, see online Appendices D, E, and F). To reduce possible bias, we used a panel of five experts, who rated each campaign in the sample. This approach has the advantage that the judgments do not rely on the potentially subjective ratings of any single expert. All experts are highly reputed in the advertising industry and were either leading creative directors or leading managers of advertising agencies. In addition, all experts were members of a festival or an award-show jury in the five years prior to our inquiry, thus ensuring that all experts had an insight into how juries select winners. None of these experts was or is on the jury of any of the award shows in the sample, which guarantees some degree of neutrality. The intercoder reliability (ICR) of the five different expert lies at .876, which can be considered as acceptable. We calculated the ICR according to the well-established approach of McGraw and Wong (1996) using the ICC procedure in SPSS. To ensure comparability with other studies we thereby only relied on the experts' evaluations and did not include their stated confidence. Including this measure would lead to an even higher ICR measure.

Figure 2: Distribution of Media Types Used for Campaigns



We aggregated these individual ratings by their "confidence-based weighted mean" (Van Bruggen, Lilien, and Kacker 2002). This measure accounts for the confidence that experts have in their specific ratings. The expert can thereby express his or her confidence in the judgments by choosing a level between one and five. The higher the confidence level, the higher the weight the respective expert receives in the subsequent score calculation. Van Bruggen, Lilien, and Kacker (2002) delivered empirical evidence that this technique leads to improved judgments and better aggregation.

The next section describes the operationalization of the variables according to the research hypotheses and the collection of the variable information. Because a large portion of the constructs from the hypotheses are not directly observable, we use latent variable operationalizations for them.

Table 2: Sample statistics

Item	Percentage
Agency size	
1–49 employees	18.5 %
50–99 employees	37.0 %
100–499 employees	33.3 %
500–999 employees	7.0 %
1,000–1,600 employees	3.5 %
Agency type	
Owner-managed agency	59.3 %
Network-based agency	40.7 %
Association members	
0	25.9 %
1	7.0 %
2	11.1 %
3	11.1 %
4	11.1 %
5	18.5 %
6	7.0 %
7	7.0 %
Product type	
Durable	38.9 %
FMCG	26.8 %
Consumer good	34.3 %
Client type	
B2B	11.1 %
B2C	75.0 %
Charitable	13.9 %

Note: $N_{agencies} = 27$ and $N_{campaigns}=108$

4.3 Measure Development

Following the standard procedures for scale development (Rossiter 2002), we based our scales on a review of academic literature, advertising textbooks and interviews with senior creatives (e.g., copywriters, art directors, and creative directors) from

several leading German advertising agencies. Because we wanted to account for as many perspectives and methods of the executional and conceptual aspects of ad creation as possible, we were not able to rely completely on existing scales and, therefore, had to identify and use new items that would properly measure the intricacies of the ad-creation process.

We measure the performance of an ad quite straightforwardly as the overall "creativity-award-show outcome" of the individual advertisement in all 40 award shows. The data were gathered by intense offline and online searches in the individual award-show databases. The majority of award shows are based on a similar ranking system. Individual placements range from a "Grand Prix" (the highest possible rating) resulting in five points, to a shortlist nomination (the lowest possible rating) resulting in one point. Thus, we could simply summarize the individual outcome of a campaign in a creativity award show. Because some awards are very well known and have numerous applicants, we weighted the individual awards according to the official weights of the German Creativity Ranking supplied by the German Art Directors Club.

Data concerning the construct "novelty" of an ad campaign (hypothesis 1) were gathered through expert judgments. Because each expert had to rate all 108 campaigns, we tried to keep the questionnaire as short as possible. Previous research showed that creatives and especially advertisers have a skeptical view of research and science (Chong 2006, Kover 1996). Those studies state that creatives believe that research is not definable and that creativity is a more intuitive process that cannot be separated into different aspects or parts. Moreover, some creatives even believe that research constrains their own creativity. Thus, the authors tried to avoid any turning-away behavior or rejection by using well-known and straightforward measures from the practice, which were easy for our experts to understand. Following some industry textbooks and previous research, all experts were asked to directly rate the degree of novelty of a campaign with respect to (1) media use, (2) graphic design and (3) content approach (Aitchinson 1999, Sasser, Koslow, and Riordan 2007, Sullivan 2003). All items were measured on a five-point Likert scale.

To measure the second construct, "*meaningfulness*" (hypothesis 2), we also used three different items based on expert judgments. As shown earlier, "*meaningfulness*" stands in close relationship to the *relevance* of the advertisement's message to consumers. To measure this relevance, we relied on three different items derived from the well-known Resource-Based View (RBV) and the concept of the Unique Selling Proposition (USP). According to Barney (1991), each of the following criteria has to be fulfilled to create a relevant and unique resource. First, the message of the campaign has to focus on an aspect of a product that is (1) unique in supply. In addition, this aspect has to be (2) non-imitable and (3) non-substitutable. All five experts were asked to separately rate each advertisement based on these three criteria. Expert judgments were based on a five-point Likert scale from 1 (no agreement) to 5 (high agreement).

As shown during the development of the hypothesis, an advertisement should build a lasting connection between the viewer and the advertised product or brand. To ensure such a connection, the message conveyed has to evoke a rational or emotional resonance in the viewer. Previous research showed that such a resonance can be evoked by presenting the viewer with a lasting and relevant product or brand *benefit* that *connects* the viewer with the advertisement (Im and Workman Jr 2004) and leads to an enduring memory of the ad or the advertised product. Following the need for easy-to-understand items and a short questionnaire, we operationalize our construct "connectedness" – in accordance with the findings of Bergkvist and Rossiter (2007) – with a single item that measures the connecting benefit. We asked our experts straightforwardly if a campaign communicates some sort of *connecting benefit* for the advertised product. Experts could rate the individual appearance of a communicated connecting benefit on a five-point Likert scale from 1 (no agreement) to 5 (high agreement).

The information for hypothesis 4 regarding whether a campaign's main idea followed a *multi-channel approach* was measured by a dummy variable that took the value of 0 when the concept was created for one channel alone and that took the value of 1 when the idea was integrated into more than one channel. Again, single-item measurement is

in accordance with the findings of Bergkvist and Rossiter (2007). Data were gathered from a search of special ad databases and the web presences of the agencies.

The independent variable "fake campaign" measures whether a campaign was created exclusively for creativity award shows ("*fake campaign*") (hypothesis 5), as determined by subjective judgment (yes/no) from our five experts.

Finally, the influence of the agency-related variables (hypotheses 6a, 6b, and 6c) is measured in accordance with Bergkvist and Rossiter (2007) by an individual item for the variables "age of the company in years," the "number of employees" to control for the size, and the "number of Art Directors Club members in the organization" for the variable "creative association." The Art Directors Club is a German association of successful professionals from the media industry. Thus, membership is a good indicator of (past) creative achievements of that person and the network this person may be connected with in the industry. These variables were collected using online sources such as company websites or other industry or association websites; the individual variables and their operationalizations are summarized in Table 3.

4.4 Specification and Estimation

Given the latent nature of most of the variables, the use of reflective and formative constructs and the causal structure of our research hypotheses, we use the partial least squares (PLS) approach for estimating the data (Chin 1998, Chin and Todd 1995). This approach is advocated for path models with latent variables that are observed by multiple indicators (Fornell and Bookstein 1982, Fornell and Cha 1994). In addition, PLS allows for moderate sample sizes in contrast with the classic OLS estimation (Fornell and Cha 1994). Thus, the PLS approach seems well suited for this study. The structural model derived from the hypotheses is shown in Figure 1. We estimated our model using the software application SmartPLS (Ringle, Wende, and Will 2005). The results are displayed in Table 5.

Table 3: Variable Measurement and Data Collection

Variable	Operationalization	Data collection	H_x
Novelty	Three items: • Personal rating regarding novelty of media use (five-point scale) • Personal rating regarding innovativeness of graphical content (five-point scale) • Personal rating regarding the novelty of the content approach (five-point scale)	Expert judgment	H_1 H_7
Meaningfulness	Three items: • Personal rating regarding the uniqueness of the communicated key product aspect (five-point Likert scale) • Personal rating regarding the imitability of the communicated key product aspect (five-point Likert scale) • Personal rating regarding the substitutability of the communicated key product aspect (five-point Likert scale)	Expert judgment	H_2 H_7
Connectedness	Personal rating regarding the utility generated by the communicated key benefit (five-point Likert scale)	Expert judgment	H_3 H_7
Multi-Channel Campaigns	Dummy variable indicating whether the campaign's main idea is used in multiple channels (1/0)	Web search in online databases and in the web presence of the creating agency	H_4 H_9
Fake Campaign	Expert judgment regarding whether the campaign is fake (1/0).	Expert judgment	H_5 H_9
Age	Single item: 1. Age of the agency (years)	Web search in online databases and in the web presence of the creating agency	H_{6a} H_9
Size	Single item: 1. Number of employees	Web search in online databases and in the web presence of the creating agency	H_{6b} H_9
Association Membership	Single item: 1. Number of ADC members	Web search in online databases and in the web presence of the creating agency	H_{6c} H_9

4.5 Results

Our model explains nearly 24% of the variance for creativity-award-show success, which can be considered to be substantive. Furthermore, our model delivers a corrected R^2 of .146 for multi-channel campaigns, which we also consider to be satisfactory. According to Chin (1998), this is a reasonable fit for a PLS model. Additionally, we validated the estimation by testing the predictive validity with an estimation sample of 80 cases (approximately 75% of the sample) and a holdout sample of 28 cases (approximately 25% of the sample). Thereby, we followed Chin and Todd (1995), who proposed that an average correlation between predicted and real values in the holdout sample higher than .3 and .5, respectively, can be considered to be satisfactory. Following Haitovsky (1969), we predicted a holdout sample with all parameters that had a t-value of at least one. With a correlation value of r = .513, the prediction satisfies all quality criteria and can be considered good.

To ensure that the measures used for the two reflective constructs are reliable, we calculated their composite reliabilities, Cronbach's alpha and the average variance extracted (AVE). Table 4 provides an overview of all three measures for both constructs. We find all measures to be highly satisfactory. The individual composite reliability values are all higher than 0.8, which complies with the suggestions made by Churchill (1979). In accordance with the guidelines provided by Nunnally and Bernstein (1994), both Cronbach's alpha values remain higher or near 0.7. At .584 (meaningfulness) and .674 (novelty), the AVE values for both reflective constructs are significantly higher than 0.5 and therefore fulfill the quality criteria mentioned by Fornell and Larcker (1981).

Table 4: Quality Criteria of the Reflective Constructs

	AVE	Composite Reliability	Cronbach's Alpha
Meaningfulness	0.584	0.805	0.700
Novelty	0.674	0.861	0.756

Our results are mixed, and to some extent, we face surprising and counterintuitive observations. All estimation results are reported in Table 5.

Table 5: Estimation Results

	Standard Error	Path Coefficient	T Statistics
Agency Size -> Creativity Award Show Performance	.090	.002	0,024 ns
Agency Size -> Fake Campaign	.090	.128	1.382 ns
Agency Size -> Multi-Channel	.068	-.280	4.106 ***
Agency Age -> Creativity Award Show Performance	.076	-.128	1.508 ns
Agency Age -> Fake Campaign	.107	.125	1.089 ns
Association Members -> Creativity Award Show Performance	.103	-.056	.523 ns
Association Members -> Fake Campaign	.097	0.170	1.740 *
Fake Campaign -> Creativity Award Show Performance	.076	-,211	2.775 ***
Multi-Channel -> Creativity Award Show Performance	.099	.259	2.484 ***
Multi-Channel Approach -> Fake Campaign	.078	-.289	3.868 ***
Meaningfulness -> Creativity Award Show Performance	.095	-.023	.230 ns
Fake Campaign -> Meaningfulness	.108	.231	1.973 *
Novelty -> Creativity Award Show Performance	.112	.239	1.961 *
Fake Campaign -> Novelty	.137	.193	1.493 ns
Connectedness -> Creativity Award Show Performance	.088	.054	0.578 ns
Fake Campaign -> Connectedness	.087	-.134	1.417 ns
ns not significant; ** = significant at least at the .95 level; *** = significant at least at the .99 level			

The assumed positive influence of "novelty" (pc: .239; t = 1.961) on creativity-award-show success as postulated in hypothesis 1 is confirmed. Although novelty seems to be a key driver for winning creativity awards, one has to be aware that path coefficients for "meaningfulness" (pc: -.023; t = .230) and "connectedness" (pc: .054; t = .578) remain insignificant. These findings are counterintuitive to our initial assumptions and surprising, as connectedness and meaningfulness are believed to be basic key components of the understanding of creativity. Our results show that award-show success is explained by presenting something that is simply novel rather than novel, meaningful and connecting. Thus, award-show creativity has to be considered as different from the general academic understanding of creativity that postulates the

holistic combination of all three characteristics. Therefore, hypotheses 2 and 3 must be rejected.

In contrast to the classic definition of creativity, our results suggest that a campaign that aims only at winning creativity awards has to find novel ways of communication or layout techniques, whereas relying on a relevant or meaningful and connecting message does not seem to be necessary. Although our finding – that working creatively does not automatically lead to winning creativity awards – contradicts conventional wisdom from the advertising industry, it supports the criticism of many marketers that creativity award shows have nothing in common with real, creative, persuading and effective advertising.

Hypothesis 4 postulates a positive effect from multi-channel campaigns on creativity-award-show success. With a path coefficient of .259 and a t-value of 2.484, the hypothesis is confirmed.

Hypothesis 5 states that campaigns created solely for creativity award shows win more creativity awards on average because their creative approach is not restricted by the client. Contrary to our assumption, the data identify a negative (pc: -.211) and significant (t = 2.775) direct effect of the construct “fake campaign” on creativity-award-show success. This finding is quite surprising because the vast majority of creatives, practitioners and critics of creativity award shows have frequently and vigorously asserted the opposite. We assume that this effect is explained primarily by the effect that juries easily identify these “fake campaigns” because of their extensive industry experience. In addition, it can be assumed that the same juries try to protect their shows from external criticism by punishing these obvious fakes. Although this negative direct effect implies that faking dos not seem to pay off, one has to be aware that a fake campaign may pose a better opportunity to show new and more creative approaches in terms of novelty, meaningfulness and connectedness. To account for this phenomenon, the authors calculate the particular net effect for the construct fake campaign by applying formula (1) to the data set.

(1) $$NFake = FaNov * novelty + FaMea * meaningfu\ln ess + FaConn * connectedness + Fake$$

In the equation, NFake is the net effect in terms of extra creativity for a fake campaign. FaNov, FaMea and FaConn are the mean differences in novelty, meaningfulness and connectedness, respectively, between a fake campaign and an ordinary campaign that is submitted to a creativity award show. Fake, novelty, meaningfulness and connectedness represent the direct effects of the particular constructs on creativity-award-show success. It should be kept in mind that most fake campaigns are produced for small alibi clients such as bakeries, pet shops or social initiatives. In exchange for free work, these clients also give their agencies a free hand in terms of the creative execution of their ads. Thus, any additional degree of novelty in the case of a fake campaign can be explained primarily by the absence of business restrictions, prejudices or risk aversion regarding creative ideas on the part of the alibi client. One also has to remember that the direct effect of faking proved to be negative. Therefore, if faking is effective in terms of winning creativity awards, the extra amount of possible creativity (viz. in novelty, meaningfulness and connectedness) will have to outperform the direct negative effect. To obtain the individual mean differences (in this case, equal to the particular unstandardized regression coefficients), we executed three different OLS regressions using the individual latent variable scores of the three latent constructs as dependent variables and the latent variable scores of the latent construct "fake campaign" as an independent variable. The obtained mean differences are reported in Table 6.

To calculate the direct effects from the four latent variables, we followed Reinartz, Krafft, and Hoyer (2004) and used the latent variable scores of the four constructs "fake campaign", "connectedness", "meaningfulness" and "novelty" as independent variables in an OLS regression and the latent variable scores of the construct "creativity-award-show success" as a dependent variable. Table 6 presents an overview of the individual results for all of the variables and the result of the net effect. Because NFake is negative, hypothesis 5 finally has to be rejected. Neither the

direct effect of a fake campaign nor the extra amount of novelty, meaningfulness or connectedness helps these types of campaign to win creativity awards.

Table 6: Net Effect Calculation of Fake Campaign

Constructs	Mean Difference	Direct Effect	Gains per Fake Campaign (mean differences * direct effects)	
Novelty	.210	20.609	4.328	
Meaningfulness	.190	-3.971	-.755	
Connectedness	-.172	4.202	-.723	
		Net effect		
Fake Campaign		-27.350	2.850	**-24.500**

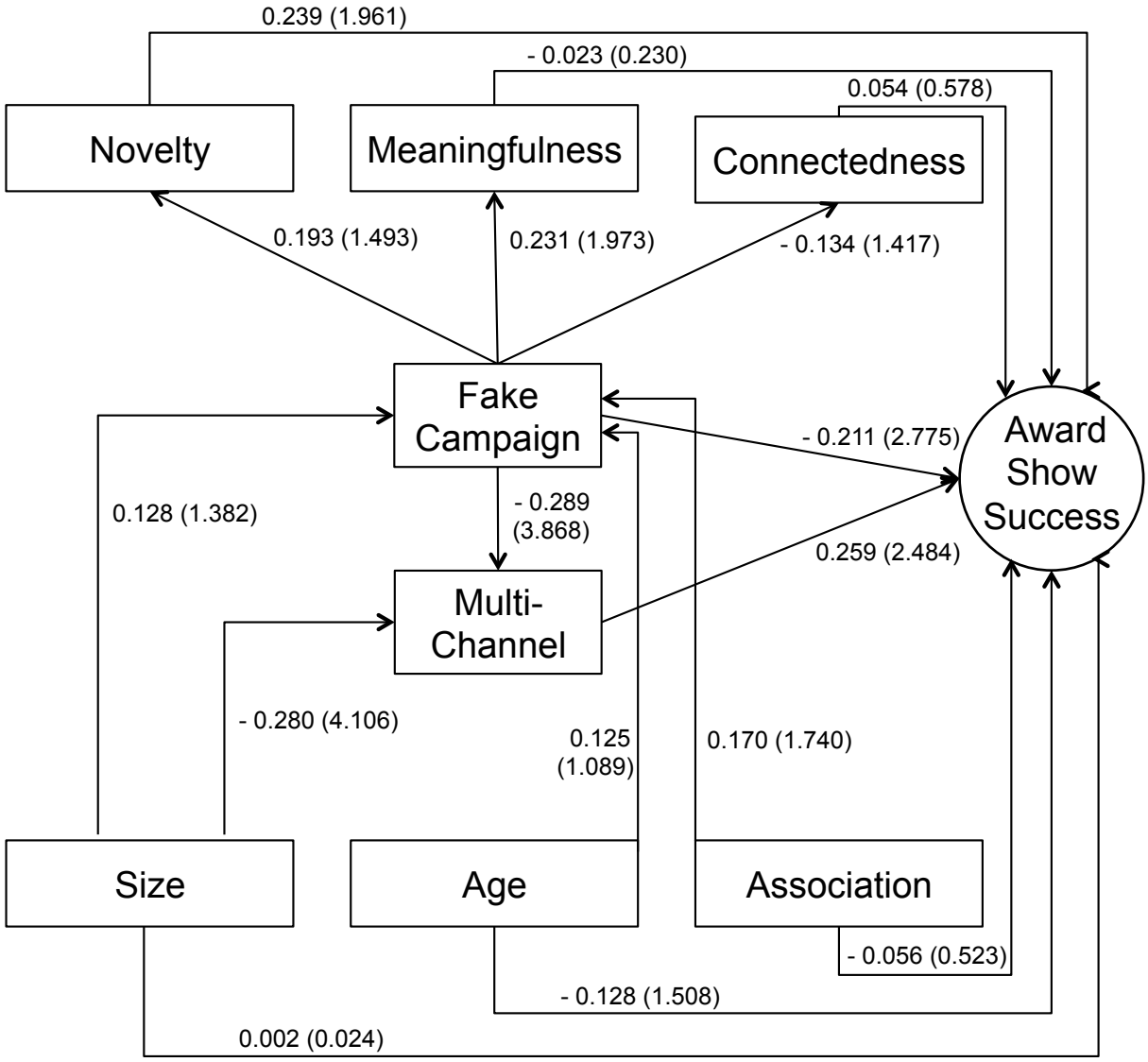
Hypotheses 6a through 6c postulate various relationships between agency-related factors and winning creativity awards. In all three cases, the path coefficients remain insignificant, with agency age taking a t-value of 1.508 (pc: -.128), agency size taking a t-value of .024 (pc: .002) and association membership taking a t-value of .523 and a path coefficient of -.056. Hence, hypothesis 6 has to be rejected. This finding is also surprising because it contradicts some old industry myths that claim that having good networking skills and experience in various associations is the primary reason for winning creativity awards. Although this finding contradicts our assumptions, it offers an important and valuable insight into the advertising world and refutes an old controversial notion.

Hypothesis 7 consists of three different assumptions and postulates that fake campaigns may use more novel, more meaningful and more connecting approaches. Our empirical findings are mixed. Although the effects for novelty (pc: .193; t = 1.493) and connectedness (pc: -.134; t = 1.417) remain insignificant, our data reveal a positive and significant influence on fake campaigns of meaningfulness (pc: .231; t = 1.973). Thus, it seems that fake campaigns are better able to communicate relevant product attributes than their "real" counterparts.

Furthermore, as we had assumed, our results reveal a negative (pc: -.289) and significant (t= 3.868) connection between fake campaigns and multi-channel campaigns. This result clearly gives empirical proof to the industry's definition of fake campaigns as typically singular products for small clients that are aired only once in inexpensive media channels that lack a

large audience simply to fulfill the submission criteria of the various creativity award shows. Thus, hypothesis 8 can be fully supported.

Figure 3: Estimation Results of the Conceptual Model



Hypothesis 9 consists of three different assumptions about fake campaigns. In the case of two of the assumptions, our results do not validate our hypothesis. Thus, the path coefficients for the impact from agency size (pc: .128; t= 1.382) and agency age (pc: .125; t= 1.089) on fake campaign remain insignificant. However, our model finds proof for a significant relationship between association membership and fake campaigns (pc: .170; t= 1.740). Thus, our results

deliver an indication that agencies with better connections with associations and therefore also with award-show juries more often attempt to maximize their creativity-award-show outcomes with faked campaigns. Again, this finding gives proof to some enduring industry myths and justifies complaints of the numerous critics of the existing award-show business. Finally, Figure 3 presents an overview of the particular effects.

5 Discussion

Although some of our results are surprising, our data deliver important answers to the various research questions raised at the beginning of this paper: First, our results suggest that one does not have to rely on an idea that simultaneously combines novelty, meaningfulness and connectedness to win creativity awards. In fact, winning these awards is more likely to reflect an agency's ability to produce novel approaches in terms of layout generation, conceptualization and media usage. Whereas novelty appears to be a key driver for creativity-award-show success, meaningfulness and connectedness play only a tangential role. This finding is quite appealing because most advertising practitioners – in line with academia –would claim that the most important elements of a successful campaign are a good idea, excellent art direction and a rich dash of creativity. Furthermore, our results acknowledge some old myths and criticisms by many marketers. Because meaningfulness and connectedness do not seem to be relevant to creativity-award-show success, winning these awards does not necessarily imply that an agency is able to create convincing campaigns, which rely on the well-investigated key benefits of products. Creativity award shows, however, seem to focus more on artistic elements than on hard but important strategy issues such as the relevance of the information or the highlighting of a clear product benefit. Therefore, marketers who pre-select agencies based on their creativity-award-show success or any creativity ranking data should keep this in mind.

In addition to these market-oriented insights, our results suggest some interesting topics for future research into advertising creativity. Until now, empirical studies have simply operationalized advertising creativity by the fact that these campaigns have

won some type of creativity award in various international award shows. Our results suggest that this definition may not be sufficient. Researchers should keep in mind that winning creativity awards primarily reflects the novelty of an approach, whereas the other two aspects of the construct of creativity, i.e., meaningfulness and connectiveness, seem to be ignored by festival juries.

In additional, our study delivers some important answers to our second research question. Our results give further insight into the business of creativity award shows and indicate the significant key drivers of awards success. First, using a multi-channel approach appears to be superior in the eyes of the award-show jurors. This finding is understandable because it is harder to create a campaign that works well in different channels (e.g., radio advertisements vs. print ads), and mastering this difficult task seems to be rewarded in the competition.

Second, agencies should stop producing so-called fake campaigns because they do not pay off. Rather, our results indicate that a negative effect exists for "fake campaigns" on creativity-award-show success. We suspect that one main reason behind this counterintuitive effect might be that as pure "award campaigns" are easily identified by the jurors, they might be "punished" by the jury. This phenomenon seems plausible because virtually all creativity award-show juries have an interest in avoiding contributions that were clearly created specifically for the award show. Most creativity awards are heavily discussed in public and in the industry's major magazines. Thus, awarding campaigns that are easily identified as "fake" could result in some loss of reputation of the individual award show. As a consequence, juries may be reluctant to reward campaigns that were created specifically for the award show rather than to attract real customers. Surprisingly, most agencies still try to maximize their performance in creativity award shows by using fake campaigns, even though they should know better because, in several international award shows, the leading creative officers of an agency are also jury members. This trend is impressively underlined by the large number of fake campaigns detected in our sample, in which 32 campaigns were rated as fake by at least 4 of 5 experts.

Our study hereby offers an important insight for both agencies and marketers in the ad generation process. First, the myth and criticism by marketers that winning creativity awards can be explained primarily by good faking abilities and good contacts with the individual award-show juries can be falsified by the empirical results of our study. Second, our results recommend that agencies instead look for novel approaches to ad design and media use. To produce only outstanding but easy-to-identify fake ideas for fictitious and non-paying clients is not sufficient.

Finally, our study provides valuable insight for marketers and answers our third research question: creativity rankings do not necessarily reflect the creative abilities of an ad agency. Marketers should keep this in mind when pre-selecting their agencies. In other words, these rankings do not provide information about an advertising agency's ability to identify and profitably communicate the relevant benefits of a product. However, the rankings offer insight into the capability of the creative workforce to identify novel ways to communicate.

6 Conclusions and Implications

Given the results of our study, we can point out important novel insights into the business of advertising award shows: First, our results challenge anecdotal industry knowledge in that we do not find any support for the assertion that fake work is more likely to win creativity awards. Considering that agencies consume many resources in terms of manpower and money for airing time or entrance fees in preparing fake campaigns, we recommend that their spending on these types of projects should be carefully scrutinized because of the unclear effects they have on the overall creative image of the company. Furthermore, the time used to produce real campaigns and to convince marketers is normally paid by the client. Thus, it seems advisable for agencies to switch resources from agency-paid fake projects to client-paid projects, which can be submitted to more than one award show.

The time and resources saved should be invested in creating multi-channel campaigns with innovative and groundbreaking layouts and approaches. Our results show that a suitable and promising concept that will win creativity awards can be identified by its ability to be effective not only in one media channel but in all possible channels. Knowing this fact may give the creative workforce or the agencies' consultants a good control measure to identify promising campaign approaches. Thus, they can determine whether an approach or a concept is promising and whether the project should be carried forward or cancelled.

Another industry myth about advertisement submissions in creativity award shows is that they rarely have anything in common with real-life advertising and market restrictions. This wisdom could be partly underlined by the results of this study. We could not identify a positive link between the emphasis on a meaningful product benefit – which effectively connects the ad to its audience – and winning awards. Clients might question the benefit of hiring an agency that wins creative awards with fake ads because, in the broad majority of cases, it must be assumed that clients will prefer agencies that are able to deliver creative solutions within the (often tight) range of the client's guidance.

Apart from these recommendations, we follow the anecdotal industry knowledge that was confirmed by our study. Thus, we recommend the submission of multi-channel campaigns that have an innovative appeal, especially in terms of the design and the overall creative approach. Of course, this practice is easily said but often difficult to do. We suspect that the massive amount of resources devoted to attracting the best creative people in the industry might truly be justified because creative innovation is the factor that separates excellent campaigns from merely good ones. Thus, an advertising agency that aims to earn creative awards should do everything it can to attract creative people and to build an idea-friendly work atmosphere. When hiring new copywriters, art directors, and creative directors to maximize their chances to win creativity awards, agencies should seek experienced creative professionals from older and larger agencies. Potential hires should have worked intensively on multi-channel

campaigns and be able to adapt a single idea to all possible channels of advertising. Moreover, these abilities should be focused on more than the membership in famous and prestigious creativity associations because copywriters and art directors from these established associations tend to rely more on fake campaigns rather than on convincing real clients to believe in their concepts.

Furthermore, we would like to encourage the research community to conduct more studies on the link between creativity and the effectiveness of advertising campaigns. At the moment, it does not seem worthwhile for creative professionals to tailor campaigns to an advertised product. If more evidence were available regarding the influence of content fit on ad effectiveness, award-show jurors might take this into consideration. Doing so would eventually lead to more creativity awards for campaigns that offer more than just a "crazy idea," namely, a very good idea that fits the advertised product.

References:

Aitchinson, Jim (1999): *Cutting Edge Advertising. How to Create the World´s Best Ads For Brands in the 21st. Century*, Prentice Hall: Singapore.

Amabile, Teresa M. (1983): The Social Psychology of Creativity: A Componential Conceptualization, *Journal of Personality and Social Psychology*, 45 (2): 357-376.

Amabile, Teresa M. (1988): A Model Of Creativity And Innovation In Organizations, *Research in Organizational Behavior*, 10: 123-169.

Amabile, Teresa M. (1993): What Does a Theory of Creativity Require?, *Psychological Inquiry*, 4 (3): 179-182.

Amabile, Teresa M. (1997): Motivating Creativity in Organizations: On Doing What You Love and Loving What You Do, *California Management Review*, 40 (1): 39-58.

Amabile, Teresa M. (1998): How to Kill Creativity, *Harvard Business Review*, 76 (5): 76-87.

Amabile, Teresa M., Regina Conti, Heather Coon, Jeffrey Lazenby and Michael Herron (1996): Assesing the Work Enviroment for Creativity, *Academy of Management Journal*, 39 (5): 1154-1184.

Amabile, Teresa M., Elizabeth M. Tighe, Karl G. Hill and Beth A. Hennessey (1994): The Work Preference Inventory: Assessing Intrinsic and Extrinsic Motivational Orientations, *Journal of Personality & Social Psychology*, 66 (5): 950-967.

Andrews, Jonlee and Daniel C. Smith (1996): In Search of the Marketing Imagination: Factors Affecting the Creativity of Marketing Programs for Mature Products, *Journal of Marketing Research*, 33 (2): 174-187.

Ang, See Hoon; Sharon, Low (2000): Exploring the Dimensions of Ad Creativity, *Psychology & Marketing*, 17 (10): 835-854.

Ang, Swee Hoon, Yih Hwai Lee and Siew Meng Leong (2007): The Ad Creativity Cube: Conceptualization and Initial Validation, *Journal of the Academy of Marketing Science*, 35 (2): 220-232.

Ang, Swee Hoon and Sharon Y.M Low (2000): Exploring the Dimensions of Ad Creativity, *Psychology & Marketing*, 17 (10): 835-854.

Baack, Daniel W., Rick T. Wilson and Brian D. Till (2008): Creativity and Memory Effects, *Journal of Advertising*, 37 (4): 85-94.

Barney, Jay B. (2001): Is the Resource-Based 'View' a Useful Perspective for Strategic Management Research? Yes, *Academy of Management Review*, 26 (1): 41-56.

Barney, Jay, B. (1991): Firm Resources and Sustained Competitive Advantage, *Journal of Management*, 17 (1): 99-120.

Barron, Frank (1955): The Disposition Towards Originality, *Journal Abnormal Psychology*, 51 (1): 478-485.

Barron, Frank and David M. Harrington (1981): Creativity, Intelligence, and Personality, *Annual Review of Psychology*, 32 (1): 439-476.

Bell, Jack A. (1992): Creativity, TV Commercial Popularity, and Advertising Expenditures, *International Journal of Advertising*, 11 (2): 165-172.

Bergkvist, Lars and John R. Rossiter (2007): The Predictive Validity of Multiple-Item Versus Single-Item Measures of the Same Constructs, *Journal of Marketing Research*, 44 (2): 175-184.

Bergkvist, Lars and John R. Rossiter (2008): The Role of Ad Likability in Predicting an AD's Campaign Performance, *Journal of Advertising*, 37 (2): 85-97.

Besemer, Susan and Karen O'Quin (1986): Analyzing Creative Products: Refinement and Test of Judging Instrument, *The Journal of Creative Behavior*, 20 (2): 115-126.

Besemer, Susan and Donald Treffinger (1981): Analysis of Creative Products: Review and Synthesis, *Journal of Creative Behavior*, 15 (3): 158-178.

Blasko, Vincent, J. and Michael Mokwa, P (1986): Creativity in Advertising: A Janusian Perspective, *Journal of Advertising*, 15 (4): 43-72.

Bogart, Leo, Stuart Tolley and Frank Orenstein (1970): What One Little Ad Can Do, *Journal of Advertising Research*, 10 (4): 3-13.

Bruner, Jerome (1962): The Conditions of Creativity, in: Howard Grober, Glenn Terrell and Michael Wertheimer (eds.): *Contemporary Approaches to Creative Thinking*, Atherton Press, New York, 1-30.

Butkys, Adolph S. and George Herpel (1992): How Advertising Agencies Handle Their Own Advertising Strategy: An Industry-Wide Overview of its Self-Promotion Efforts, *Journal of Advertising Research*, 32 (5): 18-22.

Buzzell, Robert D. (1964): Predicting Short-Term Changes in Market Share as a Function of Advertising Strategy, *Journal of Marketing Research (JMR)*, 1 (3): 27-31.

Chin, Wynne (1998): The Partial Least Squares Approach to Structural Equation Modeling, in: G.A. Marcoulides (ed.): *Modern Methods for Business Research*, London, 34-75.

Chin, Wynne and P. A. Todd (1995): On the Use, Usefulness, and Ease of Use of Structural Equation Modeling in MIS Research - a Note of Caution, *Mis Quarterly*, 19 (2): 237-246.

Chong, Mark (2006): How do Advertising Creative Directors Perceive Research?, *International Journal of Advertising*, 25 (3): 361-380.

Churchill, Gilbert, A. (1979): A Paradigm for Developing Better Measures of Marketing Constructs, *Journal of Marketing Research*, 16 (1): 64-73.

Csikszentmihalyi, Mihaly (1998): *Creativity: Flow and the Psychology of Discovery and Invention*, Harper Collins: New York.

Dahlén, Micael, Sara Rosengren and Fredrik Törn (2008): Advertising Creativity Matters, *Journal of Advertising Research*, 48 (3): 392-403.

Deshpandé, Rohit, John U. Farley and Frederick E. Webster Jr. (1993): Corporate Culture, Customer Orientation, and Innovativeness in Japanese Firms: A Quadrad Analysis, *Journal of Marketing Letters*, 57 (1): 23-37.

Devinney, Timothy, Grahame Dowling and Michael Collins (2005): Client and Agency Mental Models in Evaluating Advertising, *International Journal of Advertising*, 24 (1): 35-50.

El-Murad, Jaafar and Douglas C. West (2004): The Definition and Measurement of Creativity: What Do We Know?, *Journal of Advertising Research*, 44 (2): 188-201.

El-Murad, Jaafar and Douglas C. West (2003): Risk and Creativity in Advertising, *Journal of Marketing Management*, 19 (5/6): 657-673.

Fornell, C. and D. F. Larcker (1981): Evaluating Structural Equation Models with Unobservable Variables and Measurement Error, *Journal of Marketing Research*, 18 (1): 39-50.

Fornell, Claes and Fred L. Bookstein (1982): Two Structural Equation Models: LISREL and PLS Applied to Consumer Exit-Voice Theory, *Journal of Marketing Research*, 19 (4): 440-452.

Fornell, Claes and Jaesung Cha (1994): Partial Least Squares, in: Richard P. Bagozzi (ed.): *Advanced Methods of Marketing Research*, Blackwell, Cambridge.

Gibson, Lawrence D. (1996): What Can One TV Exposure Do?, *Journal of Advertising Research*, 36 (2): 9-18.

Goldenberg, Jacob, David Mazursky and Sorin Solomon (1999): Creative Sparks, *Science*, 285 (5433): 1495-1497.

Haberland, Gabriele S. and Peter A. Dacin (1992): The Development of a Measure to Assess Viewers' Judgments of the Creativity of an Advertisement, *Advances in Consumer Research*, 19 (1): 817-826.

Hairong, Li, Dou Wenyu, Wang Guangping and Zhou Nan (2008): The Effect of Agency Creativity on Campaign Outcomes, *Journal of Advertising*, 37 (4): 109-120.

Haitovsky, Y. (1969): A Note on Maximization of R2, *American Statistician*, 23 (1): 2-22.

Helgesen, Thorolf (1994): Advertising Awards and Advertising Agency Performance Criteria, *Journal of Advertising Research*, 34 (4): 43-53.

Im, Subin and John P. Workman Jr (2004): Market Orientation, Creativity, and New Product Performance in High-Technology Firms, *Journal of Marketing*, 68 (2): 114-132.

Jackson, Philip, W. and Samuel Messick (1965): The Person, the Product, and the Response: Conceptual Problems in the Assesing of Creativity, *Journal of Personality*, 33: 309-329.

Johnson, Michael (2003): All Together Now, *Design Week*, 18 (48): 14-15.

Jones, John Philip (1995): Single-Source Research begins to fulfill its promise, *Journal of Advertising Research*, 35 (3): 9-16.

Keil, John M. (1975): Can You Become a Creative Judge?, *Journal of Advertising*, 4 (1): 29-31.

Kilgour, Mark and Scott Koslow (2009): Why and How Do Creative Thinking Techniques Work?: Trading off Originality and Appropriateness to Make More Creative Advertising, *Journal of the Academy of Marketing Science*, 37 (3): 298-309.

Koslow, Scott, Sheila L. Sasser and Edward A. Riordan (2003): What Is Creative to Whom and Why? Perceptions in Advertising Agencies, *Journal of Advertising Research*, 43 (1): 96-110.

Koslow, Scott, Sheila L. Sasser and Edward A. Riordan (2006): Do Marketers Get the Advertising They Need or the Advertising They Deserve?, *Journal of Advertising*, 35 (3): 81-101.

Kotler, Philip and Kevin Keller (2006): *Marketing Management*, Prentice Hall: Upper Saddle River.

Kover, Arthur , Stephen Goldberg and William James (1995): Creativity vs. Effectiveness? An Integrating Classification for Advertising, *Journal of Advertising Research*, 35 (6): 29-40.

Kover, Arthur J. (1995): Copywriters' Implicit Theories of Communication: An Exploration, *Journal of Consumer Research*, 21 (4): 596-611.

Kover, Arthur J. (1996): Why Copywriters Don't Like Advertising Research--and What Kind of Research Might They Accept, *Journal of Advertising Research*, 36 (2): 8-11.

Kover, Arthur, James William and Brenda Sonner (1997): To Whom Advertising Creatives Write? A Inferential Answer, *Journal of Advertising Research*, 37 (1): 41-53.

Lee, Yih Hwai and Charlotte Mason (1999): Responses to Information Incongruency in Advertising: The Role of Expectancy, Relevancy, and Humor, *Journal of Consumer Research*, 26 (2): 156-169.

Lubart, Todd (2000): Models of the Creative Process: Past, Present and Future, *Creativity Research Journal*, 13 (3/4): 295-303.

MacKinnon, Donald W (1987): Some Critical Issues for Future Research in Creativity, in: Scott G. Isakse (ed.): *Frontiers of Creativity Researc*, Bear, Buffalo, 120-130.

McGraw, K. O. and S. P. Wong (1996): Forming Inferences About Some Intraclass Correlation Coefficients, *Psychological Methods*, 1 (1): 30-46.

Michell, Paul, C. (1984): Accord and Discord in Agency-Client Perceptions of Creativity, *Journal of Advertising Research*, 24 (5): 9-25.

Moorman, Christine and Ann S. Miner (1997): The Role of Organizational Memory in New Product Performance and Creativity, *Journal of Marketing Research*, 34 (2): 91-106.

Mumford, Michael, D and Dean Simonton, K. (1997): Creativity in the Workplace: People, Propblems, and Structure, *Journal of Creative Behavior*, 31 (1): 1-6.

Myers, Jack (2004): Advertising Accountability, R-O-I & Sarbanes-Oxley Concerns Spawn Audit Industry, Source: http://www.mediavillage.com/jmr/2004/07/07/jmr-7-7-04/

Naik, P. A. and K. Raman (2003): Understanding the Impact of Synergy in Multimedia communications, *Journal of Marketing Research*, 40 (4): 375-388.

Newell, Allen, I.C. Shaw and Herbert A. Simon (1962): *The Process of Creative Thinking*, Atherton Press: New York.

Nunnally, Jum, C. and Ira Bernstein, H. (1994): *Psychometric Theory*, McGraw-Hill: New York.

Nyilasy, Gergely and Leonard N. Reid (2009): Agency Practitioner Theories of how Advertising Works, *Journal of Advertising*, 38 (3): 81-96.

Ogilvy, David (1963): *Confessions of an Advertising Man*, Atheneum: New York.

Pieters, Rik, Luk Warlop and Michel Wedel (2002): Breaking Through the Clutter: Benefits of Advertisement Originality and Familiarity for Brand Attention and Memory, *Management Science*, 48 (6): 765-781.

Reinartz, Werner, Manfred Krafft and Wayne D. Hoyer (2004): The Customer Relationship Management Process: Its Measurement and Impact on Performance, *Journal of Marketing Research*, 41 (3): 293-305.

Ringle, Christian, Sven Wende and Alexander Will (2005): SmartPLS, Source: http:www.smartpls.de

Rossiter, John R. (2002): The C-OAR-SE Procedure For Scale Development in Marketing, *International Journal of Research in Marketing*, 19 (4): 305-334.

Rossiter, John R. (2008): The Necessary Components of Creative, Effective Ads, *Journal of Advertising*, 37 (4): 139-143.

Runco, Mark A. and Robyn E. Charles (1993): Judgments of Originality and Appropriateness as Predictors of Creativity, *Personality and Individual Differences*, 15 (5): 537-546.

Ruscio, John, Dean Whitney, and Teresa Amabile (1997): How Do Motivation and Task Behaviors Affect Creativity? An Investigation in Three Domains, Source: http://search.ebscohost.com/login.aspx?direct=true&db=buh&AN=9013580&site=ehost-live

Saffert, Peter and Werner Reinartz (2011): Creativity in Advertising and Implications for Product Sales Performance, Proceedings of the 33rd INFORMS Marketing Science Conference, Houston, USA.

Sarbanes-Oxley-Act (2002): § 1519. Destruction, Alteration, or Falsification of Records in Federal Investigations and Bankruptcy.

Sasser, Sheila L. and Scott Koslow (2008): Desperately Seeking Advertising Creativity, *Journal of Advertising*, 37 (4): 5-19.

Sasser, Sheila L., Scott Koslow and Edward A. Riordan (2007): Creative and Interactive Media Use by Agencies: Engaging an IMC Media Palette for Implementing Advertising Campaigns, *Journal of Advertising Research*, 47 (3): 237-256.

Sethi, Rajesh, Daniel Smith and Whan C. Park (2001): Cross-Functional Product Development Teams, Creativity, and the Innovativeness of New Consumer Products, *Journal of Marketing Research*, 38 (1): 73-84.

Smith, Robert, Scott MacKenzie, Xiaojing Yang, Laura Buchholz and William Darley (2007): Modeling the Determinants and Effects of Creativity in Advertising, *Marketing Science*, 26 (6): 819-836.

Song, X. Michael and Mitzi Montoya-Weiss (2001): The Effect of Perceived Technological Uncertainty on Japanese New Product Development, *Academy of Management Journal*, 44 (1): 61-89.

Sternberg, Robert and Todd Lubart (1995): The Concept of Creativity: Prospects and Paradigms, in: R Sternberg (ed.): *Handbook of Creativity*, Cambridge University Press, Cambrigde, 400-429.

Stone, Gerald, Donna Besser and Loran E. Lewis (2000): Recall, Liking, and Creativity in TV Commercials: A New Approach, *Journal of Advertising Research*, 40 (3): 7-18.

Sullivan, Luke (2003): *Hey, Whipple, Squezze This. A Guide to Create Great Ads*, Wiley: New York.

Thorndyke, Perry (1977): Cognitive Structures in Comprehension and Memory of Narrative Discourse, *Cognitive Psychology*, 9 (1): 77-109.

Till, Brian D. and Daniel W. Baack (2005): Recall and Persuasion: Does Creative Advertising Matter?, *Journal of Advertising*, 34 (3): 47-56.

Van Bruggen, Gerrit H., Gary L. Lilien and Manish Kacker (2002): Informants in Organizational Marketing Research: Why Use Multiple Informants and How to Aggregate Responses, *Journal of Marketing Research*, 39 (4): 469-478.

W&V-Kreativ-Ranking (2009): W&V Kreativ Ranking, *Werben und Verkaufen*, 83 (24): 1-4.

Wentz, Laurel (2005): At Cannes the Lions Say "Grrrr", *Advertising Age*, 76 (26): 3.

West, Douglas C. (1993): Cross-National Creative Personalities, Processes, and Agency Philosophies, *Journal of Advertising Research*, 33 (5): 53-62.

West, Douglas C., Arthur J. Kover and Albert Caruana (2008): Practitioner and Customer Views of Advertising Creativity, *Journal of Advertising*, 37 (4): 35-45.

White, Erin (2003): Agencies Face New Accountability, *Wall Street Journal*, 10(2): 16.

Young, Charles, E. (2000): Creative Differences between Copywriters and Art Directors, *Journal of Advertising Research*, 40 (2): 19-25.

Persönliche Daten

Name	Raoul V. Kübler
Geburtstag und –ort	06. März 1981, Stuttgart

Ausbildung

10/2009 – 09/2012	Wissenschaftlicher Mitarbeiter im DFG Projekt „Optimierung von Rückruf-Verfahren im Gebrauchsgütersegment“ am Lehrstuhl für Innovation, Neue Medien und Marketing der Christian-Albrechts-Universität zu Kiel
10/2008 – 08/2009	Wissenschaftlicher Mitarbeiter am Institut für Innovationsforschung der Christian-Albrechts-Universität zu Kiel
04/2008 – 09/2008	Stipendiat des Graduiertenkollegs ’Betriebswirtschaftliche Aspekte lose gekoppelter Systeme und Electronic Business’ der Christian-Albrechts-Universität zu Kiel
04/2004 – 01/2008	Studium der Betriebswirtschaftslehre an der Christian-Albrechts-Universität zu Kiel mit den Spezialisierungen: Innovation, Neue Medien und Marketing, Organisation und Politik, Diplomnote: 2,4
10/2001 – 03/2004	Grundstudium der Wirtschaftswissenschaften an der Universität Hohenheim
06/2000	Abitur am Philipp-Matthäus-Hahn Gymnasium, Leinfelden-Echterdingen, Notendurchschnitt: 2,2

Berufserfahrung

Seit 07/2007	Freier Berater für verschiedene europäische Konzerne in den Sparten Beauty, Mode und Consumer-Electronics im Bereich Marketing-Strategie, Marktforschung und Marketing-Kommunikation
04/2005 – 10/2005	Wissenschaftliche Hilfskraft am Lehrstuhl für „Innovation, Neue Medien und Marketing“ der Christian-Albrechts-Universität zu Kiel
04/2004 – 06/2006	Freier Mitarbeiter in Konzeption und Text für verschiedene Werbeagenturen (u.A. für Springer&Jacoby, The Ad-Store und Zum goldenen Hirschen)
03/2003 – 04/2004	Praktikant und anschließend Konzeptioner und Werbetexter bei der Werbeagentur „Zum goldenen Hirschen“, Hamburg
10/2002 – 03/2004	Wissenschaftliche Hilfskraft am ehemaligen Lehrstuhl für Wirtschaftsinformatik der Universität Hohenheim (Prof. Dr.H. Krcmar)
03/2002 – 05/2005	Gesellschaftender Geschäftsführer Beratung und Strategie der Werbe- und Kommunikationsagentur „freieWildbahn“, Stuttgart
04/2001 – 10/2001	Trainee bei RTS Riegerteam, Werbeagentur, Stuttgart