

A COMPARATIVE AND CUMULATIVE META-
ANALYSIS OF ADVERTISING APPEALS

by

J. Hornik*
G. Miniero**

Working Paper No 7/2010

August 2010

Research No. 02300100

* Faculty of Management, The Leon Recanati Graduate School of Business
Administration, Tel Aviv University, Tel Aviv, Israel.
Telephone: 011-972-3-6408098. Email: hornik@post.tau.ac.il

** Bocconi Business School, Milan Italy.

This paper was partially financed by the Henry Crown Institute of Business Research in Israel.

The Institute's working papers are intended for preliminary circulation of tentative research results. Comments are welcome and should be addressed directly to the authors.

The opinions and conclusions of the authors of this study do not necessarily state or reflect those of The Faculty of Management, Tel Aviv University, or the Henry Crown Institute of Business Research in Israel.

ABSTRACT

Marketing managers and scholars have focused substantial attention on the role of advertising message appeal in the persuasion process. We conducted a comprehensive quantitative and qualitative literature review of more than a thousand reports, not previously systematically reviewed, to study whether and by what size advertisements with appeals are more persuasive and better liked than ads with no structured appeals. We compared the most common appeals (fear, humor, sex, comparative, gain/loss frame, two or one sided, and metaphor) to determine their relative effectiveness. We also performed separate analyses for each response variable to assess the contribution of several moderating effects on the results. A test of heterogeneity indicated the presence of moderators on observed relationships. Moreover, we employed a cumulative meta-analysis (CMA), at every point during the history of a research domain to test for stability of results. To supplement the quantitative analyses and to analyze qualitative papers and case studies, we employed a qualitative comparative analysis (QCA) on an additional 306 qualitative reports. The meta-analysis included 703 articles containing 1631 independent samples and 2547 effect sizes representing 142,443 participants. Results show that the overall appeal effect between conditions is of small to moderate size and that profound differences exist between appeals while some moderating variables have significant effects on effect sizes in appeal studies. All methods provided the opportunity to obtain results of theoretical and practical interest. We discuss explanations for these results, suggest managerial applications, and identify directions for future research.

In a stunning bid to show consumers that behind every fur coat was a living animal beaten to death, the actor Pershia White held the bloody body of a skinned fox in a graphic ad for PETA titled, "Here's the Rest of Your Fur Coat." PETA officials explained, "This image is shocking and make[s] people feel sick and shaky. We are trying to make a statement of fear and anxiety." In

March 2007, PETA replaced this campaign with the I'd Rather Go Naked Than Wear Fur campaign. Some of Hollywood's leading ladies posed nude for PETA's antifur campaign. When later asked, PETA vice president Tracy Rieman, PETA VP responded to the question of which campaign was more successful, "We usually used symbols of violence and fear and decided to move to physical sensations because there are many equally successful ways to make a statement." (Advertising Age, May 1, 2009).

One of the most prominent and enduring questions facing advertising academicians and managers centers on the determinants of ad campaign performance. Because of potential differences in the persuasive power of different messages for the same product, "it is intriguing that there is little published research indicating the relative effectiveness of different types of appeals" (Williams, Lee, and Haugtvedt 2005, p.5). A shortcoming in the ad appeal literature is a lack of clear, consistent construct definitions and conceptualizations, both on the predictor and on the criterion sides of ad appeal effects (Hornikx and O'Keefe 2009). As a result, inconsistencies exist among studies, and empirical findings are not always in line with theoretical predictions. Therefore, our primary motivation with this study was to synthesize the existing research over a wide variety of disciplines and studies. Most published studies have compared single-message appeals with nonappeal ads. So far, while generalizations have been sought in specific areas such as health communications (Keller and Lehmann, 2008), no overall estimates of appeal performance have been reported in the literature. Therefore, this research investigates whether there is a general appeal effect and the conditions apt to produce it.

The seven most commonly used ad appeals—comparison, humor, fear, sex, two sided, gain framed, and metaphor—account for two-thirds of all advertisements (Allen and Preiss 2007; Dahlen, Lange, and Smith 2010) and close to 85% of television commercials (Pechmann and Stewart 1990). Therefore we use a comparative meta-analytical technique to determine the relative effectiveness of the seven major appeals. We also explore potential moderating factors. The literature has put forward several variables as possible moderators of message appeal effects both on the combined (pooled) appeal data and on each individual appeal. We supplement a quantitative meta-analysis with a qualitative comparative analysis (QCA), which enabled us to introduce a new analytical tool to marketing researchers and to compare findings from the two distinct research approaches. Specifically, because the role of advertising appeals has been studied since the early 1960 with notable differences in reported results, we also

employed a cumulative meta-analysis (CMA), or the procedure of performing a (new) meta-analysis at every point in the history of a research domain. This allows us to examine whether the basic pattern of results is changing over time by examining two distinct facets of cumulative knowledge: sufficiency (“Are additional studies needed to establish the existence of the phenomenon?”) and stability (“Will additional studies change the aggregate picture of the phenomenon?”).

We show which appeals, and moderating variables, influence response behavior. For example, we find that emotional appeals have much greater effects than rational messages. The overall comparative effect sizes we found should provide ad executives with a benchmark for designing and evaluating ad campaigns.

The article proceeds as follows. In the following section, we describe the literature streams on which we build our predictions and provide short descriptions of moderators and dependent variables used in our analyses. We next describe our literature review procedures and discuss methodological issues regarding our comparative meta-analysis and QCA. We then present the quantitative and qualitative results. Finally, we discuss our findings, study limitations, and their implications for advertising practice, concluding with some brief suggestions for further research.

Previous Research

The basis of scientific knowledge is the accumulation of research. To make sense of a body of research, some form of synthesis is needed such as “Optimal strategic marketing decisions depend on the magnitude (size and explained variance) of an effect, for example, the impact of advertising on market share or profitability” (Farley et al. 2004, p.145). Detecting and explaining systematic patterns of research findings such as those identified in Hanssens (2009) greatly aids progress in the investigated area (Tellis 2009). Norms can be applied to refine hypotheses to be tested in further research, as with the conclusion that an appropriate null hypothesis for the value of an advertising elasticity should be about .2 rather than 0 (Farley, Lehmann, and Sawyer 1995) or that the persuasiveness of a certain ad appeal can be expected to be in the range of 20% to 30%. The development of norms is especially helpful when the topic is

controversial or when the relevant studies appear in a broad mix of outlets and have a wide range of reported results.

Most studies in the field of ad appeals have used experimental methods to test whether a certain ad appeal influences consumer response. In a typical experiment, researchers show subjects a series of magazine or television advertisements that employ different appeals for products with equivalent attribute information (the experimental conditions) or images that are considered neutral (as a control condition). Following the experimental manipulation, respondents are asked to complete assessments of the advertisements or the advertised products. Differences in the dependent variables are then attributed to differences in the appeal (Armstrong and Patnaik 2009). Consequently, researchers have interpreted the results as showing the importance of the specific type of appeal. Although experiment-based media exposure cannot approximate the massive media bombardment that occurs to consumers daily, a particular strength of this method is the potential for causal inferences on the nature of the relationship between the appeal manipulation and recipients' responses. Despite the frequent use of such studies, there have been some inconsistencies among the specific effect sizes reported in the literature (Vakratsas and Ambler 1989; Yang and Smith 2010). This disparity, largely due to the lack of differentiation among the different appeals types and their relative effects on response behavior, has proved a major issue for practitioners when planning their campaign or ad pretest research (Williams, Lee, and Haugtvedt 2005).

Studies that measured appeal and other message effects have reported relatively small effect sizes. This suggests that there is no single variable (yet identified) that exerts an extraordinarily powerful effect on persuasive outcomes. On the contrary, the variables that have been systematically examined thus far all have produced comparatively small effects, and the differences among them also are small. Past meta-analyses designed to pool results across multiple studies have also found weak effects. For example, the mean difference in persuasive effects between one-sided messages and refutational two-sided messages is (measured as the absolute value of an n -weighted mean r , computed using the $r-z-r$ transformation procedure and the individual effect sizes reported in each meta-analysis) is .07 (number of studies, $k = 42$; O'Keefe 1999), and that between one-sided messages and nonrefutational two-sided

messages is .03 ($k = 65$; O'Keefe 1999). In addition, the mean effect of forewarning on persuasive outcomes is .17 ($k = 12$; Benoit 1998), and that of rhetorical questions is .05 ($k = 1$, Gayle, Preiss, and Allen 1998) while the mean effect of message framing on decision making is .02 (Kuhberger 1998). Thus it appears that one should not expect to find overall mean effect sizes much greater than a correlation of about .20. Nonetheless, one might appropriately wonder whether stronger effects would be observed after taking into account possible moderating factors. That is, even if the mean effect of a given variable is, on average, relatively small, perhaps under specifiable conditions the observed mean effect would be greater.

Researchers have long debated the influence of advertising message appeals on consumer responses. Only during the past two decades have investigators begun to use meta-analysis to determine how effectively certain single appeals are (Table 1). Despite the recency of these meta-analyses, an updated review that reflects the rapidly growing number of articles that have been accumulating during the past few years and simultaneously examines multiple appeals and their covariates is still worth pursuing. Using a preliminary analysis of previous meta-analysis on each single appeal, we suggest in Figure 1 a conceptual framework for the design of our investigations (for a complete review of the seven appeals, see the appendix online).

<Table 1 About here>

<Figure 1 About here>

Rational versus Emotional Appeals

Ad appeals have commonly been classified as rational or emotional. Abernethy and Franke (1995) used that classification to conduct a meta-analysis on the influence of ad content on consumers' response behavior. Rational appeals focus on the practical, functional, or utilitarian benefits derived from the use of the product. Similarly, the definition adopted in marketing communications has used *emotion* to signify any stimulation of feelings. Studies have documented the wide array of emotions that advertisements can evoke (e.g., Lee, Amir, and Ariely 2009; Yang and Smith 2009). Indeed, the avoidance versus approach model can explain, for example, differences between fear and humor. Nielsen, Shapiro, and Mason

(forthcoming) proposed that high levels of emotionality demand attentional preference, and showed a direct effect of emotionality on attention. In contrast, some empirical evidence has shown that consumers prefer rational appeals that provide information that explains clearly the differences between the advertised brand and its competitors (e.g., Stewart and Koslow 1989; Tuan Pham 2009). A meta-analysis of eight advertising studies conducted by the Center of Emotional Marketing shows that advertising that provokes a strong emotional response without providing sufficient product information often breaks through the clutter but is unlikely to persuade (Picot-Zane 2006). To compare the two constructs in this study, we aggregated three emotional appeals—fear, humor, and sex—and compared their impact with that of rational appeals such as comparison, gain/loss, and one or two sided (Hornikx and O’Keefe 2009).

Dependent Variables

The two types of dependent variables most commonly used to assess the effects of advertising appeals have been persuasion and ad liking (e.g., Hornikx and O’Keefe 2009; Tellis 2009). We assessed persuasion through measures such as attitude toward product, attitude toward brand, purchase intention, and product choice. Whenever a study had reported multiple measures for persuasion, we computed an effect size estimate for each measure and averaged these into an overall persuasion measure (Armstrong and Patnaik 2009). The other dependent variable was ad liking, commonly referred to as attitude toward the ad (Aad). This variable denotes a consumer’s feelings and overall attitude toward the ad format. Indeed, Calder and Malthouse (2008) have advanced the engagement hypothesis with evidence that strong feelings toward the ad are able to engage consumers in the ad itself, as well as the brand’s positioning. Most but not all ad response studies collected both dependent variables of interest. Indices of persuasiveness and ad liking have sometimes been empirically positively associated, but ad liking and persuasion can also diverge, especially with more familiar products. Hence, ad persuasiveness and ad liking are treated as different outcomes.

Research Questions

On the basis of the foregoing conceptualization and literature review on the impact of message appeals in advertising, we have formed the following research questions:

RQ1: *Overall, are ads that use the seven basic appeals more persuasive than ads without such appeals? What is the effect of different ad message appeals on persuasion?*

RQ2: *Overall, are ads with the seven basic appeals better liked (greater Aad) than corresponding ads without these appeals? What is the overall effect of ad message appeal on Aad?*

RQ3: *Do emotional appeals yield greater overall response behavior (persuasion and Aad) than rational appeals?*

RQ4: *On a comparative basis, do the seven appeals (i.e., comparison, humor, fear, sex, two sided, gain/loss, and metaphor) yield similar results on the two outcome variables?*

In addition to determining overall effect, we also examine conditions that might enhance or diminish the magnitude of the effects. Specifically, we investigate the following research questions:

RQ5: *Are there moderators that explain variability in the appeal's impact?*

RQ6: *Are there moderators that explain variability in each of the seven appeals' test results?*

Given the 40 years of research on message appeal, and the increasing proliferation of media, we also proposed the following question:

RQ7: *Has there been a trend with respect to appeals effectiveness over the past five decades?*

Methodologically we used quantitative and qualitative meta-analysis as well as QCA to investigate the following:

RQ8: *Do quantitative meta-analysis and qualitative comparative analysis produce similar results?*

Potential Moderating Variables

Meta-analysis is a useful tool for integrating research findings. One major advantage of it is that it allows for an examination of the extent to which variation in findings is due to either random sampling error or to systematic variations (heterogeneity) among studies (Hunter and Schmidt 2004). Heterogeneity indicates the presence of moderators of observed relationships, and is valuable for identifying both the

existence and the nature of moderators. In the absence of a strong theory of message appeal, our identification of potential moderator variables was largely inductive. We used three criteria to select from the numerous potential variables moderating the relationship between advertising appeals and consumer response. First we included variables that were found to be important in an exploratory analysis, provided sufficient effect sizes to test the relationship, and are theoretically important in this particular area of research. These variables included product type, media, method (survey or experiment), respondents (students vs. nonstudents), and gender. Several additional variables also had the potential to explain variations in consumer response but were used in some meta-analyses. We included several of these variables in this meta-analysis as well. In terms of methodology, prior meta-analyses have suggested that multi-item scales produce different results from scales with single-item measures (Henard and Szymanski 2001). In addition, reverse coding versus no reverse coding of items may impact responses. Finally, we also included articles' journal quality, using the classification of Kirca, Jayachandran, and Bearden (2005).

This meta-analysis builds on and extends prior meta-analyses in several ways. First, we quantitatively summarized more than 1000 articles in various disciplines, a number almost three times that of the 360 articles in the previous reviews, resulting in a database of more than 142,000 observations. This enables us to examine more possible determinants of appeals effects with greater statistical power. Second, most previous meta-analyses have treated all studies alike, taking their authors' results at face value. In contrast, we corrected for the influence of statistical artifacts such as sampling error and measurement error. Third, we give extensive attention to moderating effects, which most previous studies have discussed only briefly. This allows for quantitative predictions of the likely effect under different conditions. Fourth, we also employed a comparative meta-analytical approach (Vittengl et al. 2007), which, to the best of our knowledge, has not been used before in the marketing literature.

Method

A major difference between the previous meta-analyses and ours is that the earlier studies concentrated on statistical significance and variance explained, whereas we focus on the magnitude of the effects. In other

words, most studies in this area have relied on the vote-counting method, which consists of tabulating significant and nonsignificant findings. This technique is essentially a dichotomous accept–reject method. Our approach incorporates degree of support because collectively five p -values of .06 constitutes stronger evidence of a relationship than five p -values of .45. Second, while most studies tested one variable at a time, we analyzed the data in a multivariate model. This enabled us to examine partial and interaction effects and to control for the effects of other variables and reduce omitted variable bias (at the price of having to deal with collinearity) (Keller and Lehmann 2008).

With respect to specific meta-analytic procedure, the present analysis followed the Glassian advices of coding of study characteristics (for different methods of meta-analysis, see Farely et al. 2004; Sutton and Higgins 2008). We largely based the formal analysis on Hunter and Schmidt (2004) using the computer program Comprehensive Meta-Analysis (version 2.2.021; Biostat 2007), which was developed for meta-analysis support. We corrected the means and variances for sampling error and attenuation due to measurement error in the predictor and the criteria (i.e., reliability). We used artifact distributions to correct for measurement error in the criterion variables because most but not all studies included reliability estimates for each measure. In the following sections, we describe the five major steps of our meta-analysis: (1) locating and searching the literature, (2) selecting a common metric, (3) computing an average effect size and testing for variance homogeneity across studies, (4) identifying moderators, and (5) computing a fail-safe N . Statistically, this meta-analysis involved two basic procedures: averaging zero-order correlation coefficients of studies weighted by their respective sample sizes and determining whether there is a significant difference in effect size across studies. We also used three supplementary methods. The first was second-order meta-analysis used to synthesize previous meta-analyses. Second, we used cumulative meta-analysis to investigate ad appeal influence over time. Third, we used QCA to explore the contributions of nonempirical papers on the subject. The meta-analysis model consisted of the steps summarized in Figure 2.

<Figure 2 about Here>

Literature Search

We used four phases of data collection to identify studies (articles, book, chapters, and unpublished reports) as input for our meta-analysis. First, we examined the following computerized database retrieval systems: ScienceDirect, Communication Abstracts ABI/INFORM Global, Dissertation Abstracts, PsycINFO, PapersFirst (1993–2009), Article1st (1990–2009), Social Science Abstracts (1960–2009), EconLit.3, and Google Scholar. We searched the databases using the following key words either singly or in combination: *meta-analysis*, *appeal*, *persuasion*, *message framing*, *fear*, *humor*, *comparative*, *gain-framed*, *two-sided*, *one-sided*, and *metaphor*. In addition, we performed a supplementary search including studies with attitude toward advertisements or advertising tactics. Second, manual searches involved 19 major journals of widely accepted scholarly value from 1960 through November 2009. We identified relevant research reports through personal knowledge of the literature and examination of reviews, reference lists of located reports, and relevant conference proceedings (e.g., American Academy of Advertising, Association for Business Communication, Association for Consumer Research, Society for Consumer Psychology). In the third phase, we consulted the reference sections of the articles retrieved in the second phase and the reference sections of all major previous reviews of ad message appeals to identify any studies that we might have overlooked. Mindful of the possibility of publication bias and the attendant possibility that research appearing in the published research literature might yield an inflated estimate of the average effect size (Farely et al. 2004), in the fourth stage we sought both published and unpublished research reports. To that end, we contacted 78 scholars to request unpublished data and other information that might provide additional cross-category data. In some cases, scholars supplied theses, dissertations, citations, and unpublished papers (e.g., Poels, Dewitte, and Vyncke 2009; Strick and van Baaren 2009). We pursued the leads until either we found additional data or received no reply after repeated requests. Last, we included effects from data collected in our own recent research.

We grouped reports into five main types: (1) primary empirical studies that quantitatively assessed the relationships of ad message appeal to outcome variables, (2) secondary reports consisting of reviews or summaries of the primary empirical studies, (3) theoretical reports that provided arguments for

why ad appeal may or may not be effective, (4) primary reports of case studies of successful advertising using appeals and (5) opinions of advertising effectiveness with no empirical support. We used groups 1 and 2 in the quantitative comparative meta-analysis and groups 3 and 4 in the QCA; group (5) was excluded from both the analyses. For inclusion in the QCA, we attempted to include practitioners' reports and single-source reports from some non-English-speaking countries such as Italy, the Netherlands, and Israel if they met the following criteria: (1) quality, objectivity, and consistency and (2) scope, parsimony, usefulness, and linkage with theory.

Inclusion Criteria

We included studies in the meta-analysis if and only if they fulfilled three criteria. First the studies had to examine either appeal main effects or interactions between one appeal and either a different appeal or other variables. We did not consider studies dealing with advertising appeals that did not provide such comparisons (e.g., Unger and Stearns 1983). When two studies used the same data (e.g., an unpublished dissertation and a journal publication), we used Wood's (2008) procedure to detect duplicate effects. We excluded 11 studies because they reported on the same data set. We required that appropriate quantitative information was available to compute effect sizes for a dependent variable of interest, namely persuasion or ad liking. This criterion excluded studies with other dependent variables, such as ad interpretation (Callow 2000). Whenever the reports did not provide sufficient quantitative information, we made efforts to obtain it from the authors. Finally, we required that the studies report sample sizes and the requisite outcome statistics (e.g., r , univariate F , t), to calculate effect sizes. With one exception, the inclusion rules for the meta-analysis of nonexperimental studies were identical to the criteria for experimental studies. This criterion excluded studies in which it was not clear which appeal was manipulated (e.g., Picot-Zane 2006). We also excluded studies that presented only one ad appeal to different audiences or that confounded the manipulation with other manipulations (e.g., brand and/or product) from the meta-analysis. While limiting the number of studies included, this eliminates possible confounds which could not be easily accommodated in the analysis. For studies that manipulated the appeal which we have

measured values (e.g., fear, strength), we converted the manipulated binary variable (e.g., high fear) into scale values. To do this, we used both a logical extreme value (.9) and the 95th percentile value of measured values on the variable (e.g., .81) (Keller and Lehmann 2008). Because the results do not vary substantially depending on which was used, we report results based on that which was easier to implement. [i.e. the logical extreme]

This extensive search of the literature yielded more than 1000 reports, most of which were published. Of the reports, 703 were empirical studies and the rest were secondary studies, consisting of theoretical reports, books, cases, opinions, or reviews of the empirical studies. The 703 studies span 1966 to the current date, 142,443 participants, 1631 independent samples, and yielded 2547 effect sizes. They also cover a variety of countries and contexts, focus on various independent and dependent variables, and use a rich variety of research methods to arrive at their conclusions.

Coding: Once collected, we coded all studies for analysis using a method derived from prior research and theory (Borenstein 2009). Five people coded the data at the study level independently after jointly developing and reviewing coding instructions. They resolved discrepancies in discussion after reviewing the article. Numerical variables consisted of year published, control sample size, control group response behavior, appeal group sample size, and appeal group response behavior. If the variable was manipulated (e.g., case information provided) at the same level for all respondents, it (case information) was coded as a 1 (or 0). To enhance comparability for scales with different ranges, values were converted to 0 to 1 scales. For example, a 4 on a 5-point scale, the most common scale encountered, was set to $(4 - 1)/(5 - 1) = .75$. (Linear recoding of an interval-scaled variable does not affect its correlation with another variable). Coding all variables on a 0–1 scale made it easier to compare and interpret the size of unstandardized coefficients (Sutton and Higgins 2008). Intercoder reliability was checked midway through the coding process and again at the end, with good results. Average intercoder agreement was high (89.8%), as was reliability (kappa) corrected for chance agreements (.903).

Analysis

We first conducted a separate meta-analysis for each of the seven major types of appeals, then one for the combined rational and emotional appeals, and last one for the pooled data. Studies on message appeal effect typically report the difference between appeal and no appeal. While empirical research often presents test statistics and their respective levels of significance, “a test of significance without an effect size estimate fails to tell the whole story” (Rosenthal and Rosnow 1991, p. 45). Fortunately, in most cases it is possible to compute effect sizes from significance tests. For studies that provide t 's or F 's, we converted those to r 's using Wolf's (1986) t -to- r conversion formula:

$$r = \sqrt{\frac{t^2}{t^2 + df}}, \quad (1)$$

where df is the degrees of freedom for the t test. Similarly we used Equation 2 for the F -to- r conversion:

$$r = \sqrt{\frac{F}{F + df}}, \quad (2)$$

where F is the value of an F statistic and df is the degrees of freedom for the F test.

Effect Size

After we converted study results to a common metric, we computed the average effect size across studies. According to Hunter and Schmidt (2004), if the population correlation is assumed to be constant over studies, then the best estimate of that correlation across K studies is a weighted average in which each correlation is weighted by the number of persons in that study. Therefore we computed the sample size weighted mean correlation across studies:

$$r_{ave} = \frac{\sum (N_i r_i)}{\sum N_i}, \quad (3)$$

where N_i is the sample size in study i and r_i is the correlation in study i .

Before undertaking a moderator analysis, we analyzed the data to determine whether the variance in effect size across studies was greater than could be expected by chance, thus warranting a moderator analysis.

We performed a test of homogeneity using a chi-square (χ^2) statistic with

$K - 1$ degrees of freedom (Hunter and Schmidt 2004):

$$\chi^2_{K-1} = \frac{\sum N_i}{(1 - r_{Ave}^2)^2} s_r^2, \quad (4)$$

where r_{ave} is the mean correlation coefficient across all studies, and s_r^2 is sample size weighted variance of the mean correlation coefficient obtained from Equation 5:

$$s_r^2 = \frac{\sum [N_i (r_i - r_{ave})^2]}{\sum N_i}, \quad (5)$$

To calculate pooled mean effect sizes, we used Comprehensive Meta-Analysis software (Version 2.2.021; Biostat 2007). Both random- and fixed-effects models were used to determine whether there were any differences. The fixed-effects model assumes that all studies in the meta-analysis are replications of each other while the random-effects model assumes that the included studies are a sample drawn from a population of studies and that each primary study introduces its own amount of heterogeneity into the meta-analysis.

Test for heterogeneity: We investigated moderating variables to determine whether there was a statistically significant difference (at the $p < .05$ level) among average effect sizes (in terms of r) and whether the mean variance of effect sizes was less than that of the overall mean effect size. Should both criteria be satisfied, the variable is said to be a significant moderator. Regarding the second condition, Hunter and Schmidt (1990) suggested correcting the cross-study variance for sampling error. They posited that, because two factors confound observed variance (actual variation in sample correlations and variation in sample correlations produced by sampling error), an estimate of variance in population correlations should be obtained by correcting the observed variance s^2 for sampling error. Therefore, we computed the observed variance using Equation 5 and then calculated the sampling error using Equation 6:

$$\sigma_e^2 = \frac{(1 - r_{Ave}^2)^2 K}{\sum N_i}, \quad (6)$$

Once the sampling error was computed, we subtracted it from s_r^2 :

$$\text{corrected } s_r^2 = s_r^2 - \sigma_e^2 \quad (7)$$

Fail-safe N: Given the likelihood that we did not obtain every study that examined the effects of message appeal on consumers' response, we calculated the number of additional studies with effect sizes less than a set criterion value that must include in the meta-analysis to make the effect non-significant. We used .10 as the criterion, which is considerably lower than the effect size of .2 that Cohen (1988) suggested.

For p of .05, the formula reduces to the following (Rosenthal 1984):

$$x = \frac{K(KZ_{Ave}^2 - 2.706)}{2.706}, \quad (7)$$

where x is the number of studies needed to reduce the effect size to non-significance, K is the number of initially summarized studies, and Z_{Ave} is the average Z obtained for the K studies.

Temporal Stability

Changes in the magnitude of effects over time are a general phenomenon in marketing. Significant changes in the magnitude and even the direction of research findings with time have recently been reported in several advertising fields (e.g., Hornik and Aner 2010). Such trends have been attributed to paradigm shifts, scientific fads, changes in methodological approaches, and biases in the choice of study systems. Because we investigated ad message appeal studies over almost 50 years, and because appeal affects might have changed over time (Beard 2008), we also conducted a cumulative meta-analysis (CMA), using METAWIN (version 2.0). In the first analysis, we sorted both data sets chronologically. We entered the earliest available study into the analysis first. At each step of the cumulative meta-analysis, we added another study to the analysis and recalculated the mean effect size and 95% confidence interval (CI) (Borenstien et al. 2009; Rosenberg et al. 2000). We analyzed the data using a fixed-effects model because the estimate of the pooled variance was less than or equal to zero. As studies are added to the analysis, the cumulative effect size typically initially changes substantially from one analysis to the next, but gradually stabilizes around the mean value for the whole set of studies unless the effect size depends on, for example, publication year.

Qualitative Comparative Analysis

In addition to the quantitative meta-analysis, we performed a qualitative comparative analysis (QCA).

The QCA is an approach for synthesizing findings across studies on a particular topic using methods consistent with qualitative research (Greckhamer et al. 2008). The aim of the meta-synthesis is to produce findings that offer new insights and deeper conceptualizations than can be achieved through individual investigations and quantitative meta-analysis. QCA, a nonstatistical technique, can supplement a quantitative meta-analysis to examine cases or qualitative publications (Rihoux and Ragin 2009). Basically QCA looks for commonalities across binary outcomes (e.g., fear worked/didn't work) in a set-theory or truth table style analysis. (See Appendix for more details)

The QCA we used partly followed the qualitative classification of Vakratsas and Ambler (1999), who reviewed more than 250 journal articles and books on how advertising effects consumers' responses. Their taxonomy identified three major advertising inputs: (1) message, (2) media, and (3) repetition. Our review suggested another input factor, budget (expenditure), may also be relevant (e.g. Tellis et al, 2005). Therefore we formulated the following model:

MESSAGE = an efficient appeal, i.e. is one of the seven types analyzed
MEDIA = proper mix and scheduling
REPETITION = number of inserts
BUDGET = level of expenditure

Possible findings include:

1. Message appeal → high response
2. Appeal and media → high response
3. Appeal or repetition → high response
4. Appeal and media or repetition and budget → high response.

Note that the four cases have important differences in interpretation:

1. Appeal is necessary and sufficient
2. Appeal is necessary but not sufficient
3. Appeal is sufficient but not necessary
4. Appeal is neither necessary nor sufficient

Results

We first examined the impact of each appeal separately. Table 2 shows the basic results for both Aad and persuasion. Because of space limitations, we discuss only the major results.

<Table 2 About here>

Impact of the Seven Appeals

RQ1 and RQ2 predicted that use of the seven specific ad appeals enhances ad liking and persuasiveness in comparison to neutral or control conditions. Overall, as Table 2 shows, the correlation between message appeal and consumers' response was positive. The overall reliability corrected and sample weighted correlation between appeals and response behavior is .22 and .19 (the uncorrected correlations are .23 and .20) for Aad and persuasion, respectively (with a 95% confidence interval from .13 to .27). The combined probability of the results was statistically significant (Stouffer's $Z = 5.21$, $p < .001$), although the effect size is small (Cohen, 1988).

Impact of Each Appeal (RQ3)

Separate meta-analysis of the seven major types of appeals clearly point in the same direction as the pooled analyses. All appeals had positive, statistically significant effects on the two combined dependent variables. Figure 3 provides the effect sizes of the seven appeals. The effect of sex appeal ($r = .43$) is greatest followed by humor ($r = .37$), and comparative appeals ($r = .18$). The lowest effects come from two-sided messages ($r = .10$), whereas effect sizes for fear and metaphor appeals had a small effect, slightly greater than that for gain-framed appeals.

<Figure 3 About here>

Results for Emotional vs. Rational Appeals (RQ4)

The comparison of the effect size, of the three emotional appeals versus the three rational appeals strongly supports RQ4: $r = .31$ and $r = .33$ compared to $r = .14$ and $r = .17$ for persuasion and Aad, respectively. This implies that, all else equal, consumers respond to emotional appeals more favorably than to rational appeals.

Moderator Effects

Cochran's Q test for homogeneity revealed the presence of considerable heterogeneity (Table 2). Therefore, we performed a moderator analysis for all meta-analyses. In exploring variability in the pooled

and separate appeals (g), we conducted weighted multiple regression (WMR) with the five key moderators as predictors. We begin with an overall analysis, followed by a more detailed, albeit more speculative, description of the particular moderators that account for the more general findings.

Product: Interestingly, product categories did not appear to moderate the impact of having one of the seven appeals. Categories positively interacted with only two appeal factors: sex with persuasion ($Q = 24.67, p < .05$) and humor with persuasion ($Q = 19.87, p < .05$). Sex and humor appeals have a greater impact on nondurable consumer products.

Medium: The most important influence on both dependent variables is the medium in which the ad appeared. Advertising medium is statistically significant in all models of pooled appeals. Television advertising influences ad liking more than magazines, newspapers, or radio in all significant models ($Q = 132.55, p < .01$). Outdoor advertising had a strong positive moderating effect for sex appeals ($Q = 63.60, p < .05$) and humor appeals ($Q = 45.08, p < .05$). The most striking finding was related to emotional appeals, which were more than twice as effective in television advertisements than in other media ($Q = 37.42, p < .01$). Response to humor was stronger for advertising in broadcast media than in print media, for both Aad ($Q = 28.73, p < .05$) and persuasion ($Q = 33.12, p < .01$).

Respondents: Although we found no moderating influence between college students and non-college students' respondents in the pooled model, the type of respondents had a significant positive effect on three appeals: sex ($Q = 46.85; 44.33, p < .01$), humor ($Q = 22.46; 24.59, p < .01$), and comparison ($Q = 16.88; 17.57, p < .05$). Specifically, students had significantly greater responses on the two outcome variables than did the general public in random samples. In addition, analyses of the moderator variables applied to the correlations between humor and the two response variables revealed a statistically significant positive effect for the student sample and for broadcast media.

Gender: We found no gender effects in the pooled models. However, we did find a notable gender effect for sex appeal ($Q = 122.61, p < .01$). Overall, Aad for men increased significantly greater than for women.

Other Variables: Interestingly the effect sizes of published studies are somewhat less than those of unpublished studies ($p < .10$), opposite that suggested by the file drawer problem (Duval and Tweedie 2000). This finding should be interpreted with considerable caution, however, since none of the other control variables had a significant moderating impact on the response variables in either the pooled or individual appeals models.

Outlier analysis: Because the heterogeneity in our data is considerable, identification of outliers may be helpful. Following the work of Borenstein et al. (2009), we defined outliers as those cases that most reduced the homogeneity statistic. We excluded the greatest 10% of outliers, which reduced the database from 703 to 632 effect sizes (i.e., we excluded 71 outliers). The mean combined weighted effect size for the remaining 632 cases was $r = .24$, and the mean unweighted effect size was $r = .26$, which were not significantly different from full pooled results. All other outliers were more or less similar to overall results.

Variation over Time

The QMA revealed only one significant association between magnitude of effect size and publication year in the pooled meta-analysis and in most other analyses. Specifically, more recent studies obtained greater positive persuasion and Aad due to sex appeal than did studies conducted three or four decades ago. Overall, the correlation of .23 between year of study and manipulation effect implies either improved or greater responsiveness to sex appeal in latter periods. Also, the positive–negative framing effect has decreased significantly, ($Z = -2.46, p < .05$).

Qualitative Results

The key tool for analyzing causal complexity in QCA is a truth table (Glaesser et al. 2009; Kent 2008) which lists the logically possible combinations of causal conditions and outcomes. Table 4 is the truth table used to identify explicit connections between combinations of causal conditions and outcomes. The notation in the models follows QCA conventions: A capital letter denotes presence of an input factor, and a lowercase letter indicates the absence of an input. The first column of Table 3 presents five models (combinations of input factors) observed in the literature and tested here. (All other observed cases can be

reduced to these factors by eliminating redundant aspects of the input combinations.) For example, the simplest factor is media. The lowercase letters indicate that the essential trait of this input factor is the absence of media in the qualitative report. The lack of media may occur with any number of other characteristics (or their absence), but for all the additional characteristics or input characteristics, at least one case manifests the additional trait and another case manifests its absence, which makes the trait unnecessary for describing the underlying factor.

<Table 3 About here>

The five input factors reported in Table 3 summarize the unique constellations of the ad effectiveness manifested in the 306 cases analyzed. To understand what the factors of inputs reveal about the observations, it is important to discuss both the prevalence of the different factors and the relationship of the factors to the two dependent variables. Table 3 lists the factors, starting with the most common. The two most commonly occurring input combinations are APPEAL•MEDIA• and REPETITION. The second appeal factor, •MEDIA•REPETITION, describes cases in which appeal was absent (note the lowercase letters used to describe the components of the factor). In other words, this factor depicts cases in which ad message did not describe campaign success. However, other less frequently occurring factors show that message appeal does not appear as an input characteristic. The third and fourth columns of Table 3 report the number of cases in which each factor was associated with ad success or failure. The fifth column reports the ratio of cases with the factors leading to success to those with no success.

The last row reports the summary statistics for all cases combined and thus serves as a baseline and indicates that appeal input first and media second are much more common than are other inputs. Column 5 shows the ratios of positive to negative outcomes for cases holding the input combinations relative to cases that do not. Overall Table 3 reveals that almost three of the four conditions ($QCA > .65$) include some type of association between appeal performance and market response.

Conclusions

According to the body of extant empirical research, the effectiveness of many appeals is equivocal. So, the broad question we pursued here was, what are the persuasive effects of those ad appeals that have

been sufficiently studied so as to give some confidence in conclusions about the average and comparative size of effect associated with single and pooled appeal sizes? The current analyses bring to bear an unusual set of data representing sustained in-depth research and observations of ad message appeal and appeal relations of a sort that are rarely used in marketing research. Results, perhaps unsurprisingly, confirm that advertisements carrying appeals are significantly more persuasive and better liked than are advertisements with no appeals. The advantage of meta-analytic and QCA reviews, of course, is the diversity of evidence on which such conclusions are based. The studies reviewed here used appeals with various values, advertisements for various kinds of products, different media, and participants from many different populations and countries. We obtained an overall effect size (weighted by sample sizes) of .20 linking appeals to responses. According to Cohen's (1988) rule of thumb, this link can be considered a relatively small effect size. However, our qualitative descriptions of effects show a relatively high appeal effect (QCA >.65), which reveals major differences between the two methodologies. These suggest that most qualitative reports that do not rely on data tend to overestimate the role of message appeal in the advertising process.

Appeal research is relatively young; however, one could argue that only studies that succeed in replicating an effect (and thus showing type I error, if the effect did not exist) tend to be published. This file-drawer problem may lead to an overestimation of the true size of effects. On the basis of our calculations, more than 13,000 studies finding null results would have to exist before the overall results could reasonably be ascribed to sampling bias. This is not plausible.

Our study also answered the question of whether all ad appeals are equally affective. On the one hand, we found significant differences among some of the major appeals, such as emotional versus rational appeals, and this may support the claim that emotional appeals create more positive Aad and greater persuasion than do rational appeals (e.g., Tellis 2004, 23-24, 147-174; Nielsen, Shapiro, and Mason 2010). On the other hand, we did not find significant differences among some appeals, especially among rational appeals. We found statistically significant comparative results for two appeals—sex appeal and humor. This implies that, indeed, sex sells.

We conducted a moderator analysis to locate the source of variation, because all tests of homogeneity were statistically significant. The most important moderator seemed to be media type. Indeed, two consequences of advertising success from cases and qualitative reports are message appeal, followed by media mix and scheduling. Both have been widely identified as crucial for high market response in the QCA models, which implies that message appeal is a necessary but not sufficient input for campaign success. Both our quantitative and qualitative analyses strongly suggest substantial interdependence among message appeal, media, repetition, and campaign appropriation in determining campaign performance. The results imply that two or more different combinations of factors can be sufficient for attaining the same outcome, provided inclusion of the appeal attribute, and that any particular factor may have different or opposite effects depending on the presence or absence of other factors.

The finding that in some appeal designs student participants generated significantly greater effects than did samples of the general population does not necessarily mean that use of such designs in appeals research is incorrect or artifactual. It means that researchers must be aware that student subjects usually result in greater results; thus, researchers should consider this when designing their studies and developing their conclusions.

Theoretical Implications

The relatively low mean effect size of message appeal on persuasion and Aad in our meta-analysis is typical of that found in persuasive effects research (e.g., Kuhberger 1998). A possible explanation for the small effect sizes is that experimental advertisements were in some way uncharacteristic of actual consumer advertising and were poorly designed with respect to engaging the relevant consumer's segment or target markets for the advertised product. It is a commonplace of effective persuasion that a message should be adapted to one's audience (Keller and Lehmann 2008). There are, of course, many different bases on which appeals might be adapted to audiences, but the audience's values provide an obvious basis for adapting an appeal (e.g., Hetsroni 2000). Appeals that suggest that the advertiser's view

is connected to relatively important audience values presumably are more persuasive. Indeed, Hornikx and O'keefe (2009) conducted a meta-analytic review of the research concerning the adaptation of consumer advertising appeals to the audience's cultural values. Culturally adapted value appeals were significantly more persuasive than their unadapted counterparts; the mean effect (in a random-effects analysis with 67 cases) corresponds to $r = .07$, $d = .14$. However, this were again quit generalized studied failing to focus on more specific target markets. It remains to be seen whether other kinds of persuasive message, such as mixed-mode appeals, might yield greater effects. Naturally, advertising appeals invoking emotions, especially those of mixed mode, such as sex and humor, should be considerably more persuasive than rational appeals, other emotional appeals, and each of the two appeals in isolation.

The recent popularity and the capabilities of the QCA approach provide a valuable addition, not yet employed in the marketing literature, to marketing researchers' methodological toolbox. The QCA approach to conceptualizing and analyzing causality decisively differs from statistical analyses based on linear algebra. The latter seek to estimate the separate contribution of each cause (independent variable) in explaining variation in the outcome (dependent variable) in an attempt to understand the causality underlying a particular type of outcome. In other words, instead of trying to isolate which appeal, media, repetition, budget, and so on, make the greatest relative contribution to explaining the variance in campaign performance, a QCA approach can examine which specific message appeal, media mix, and level of repetition—and any and all combinations of these inputs—commonly occur across the investigated cases to achieve superior or inferior performance. This opens a new avenue for further research in marketing.

Managerial Implications

The results of the meta-analysis should be of value to ad campaign managers as they establish goals for behavior change in future ad campaigns. Overall, the modest changes that ad appeals cause could have an important impact on private and public campaigns. In general, the effectiveness of appeals is substantially less than many managers tend to believe (Calder and Malthuse 2008). The effects ranged from .09 to .46, and in percentage terms, campaigns changed response behavior of about 20% of the population Medium

percentage changes may affect many people in a community, state, or national campaign. A 20% change among 100,000 targeted adults would yield 20,000 more people engaging in the target behavior. Results show that use of advertisements with a common appeal makes advertisements more persuasive and better liked but also demonstrates that the effects are not especially great. The implication is that advertisers should consider carefully the cost of appeals to ensure appropriate return on investment. If we consider that most studies used in our analyses employed either students or general samples, not specific target markets, we can conclude that the response measures are underestimated compared to those of the same studies conducted among the appropriate segments.

Caveats and Limitations

As with any meta-analysis, the evidence in hand necessarily circumscribes our conclusions. The most obvious limitation is that other uncoded or unknown appeals, such as nostalgia or rhetoric, are likely to confound the analyses of the investigated appeals (Gayle, Priess, and Allen 1998). Also, we found few studies examining potential mixed-mode appeals, such as sex and humor in combination. Further study of appeals among selected groups is needed to increase knowledge of the role of the various appeals in different groups, such as ethnic groups (Hetsroni 2000). Therefore, generalizations derived from our data and applied to advertising campaigns must be made with some caution. Our analysis shows that there is no uniform appeal effect. The effect sizes in the database are heterogeneous. We identified some of the factors that contribute to heterogeneity; others may have gone unnoticed. These problems notwithstanding, we uncovered some notable and potentially important results. Ad message appeal research is not a homogeneous endeavor. Diverse operational, methodological, and task-specific features make this a major topic for future research.

Future Research

The nature and theoretical mechanism of advertising effect is more complicated than is commonly supposed and awaits more scholarly attention. There are many areas for further research on ad appeals. In addition to exploring some of the issues raised in the “Caveats and Limitations” section, further studies should address more relevant issues. For example, much more information is needed on, for example,

how people process fear appeals as opposed to sex appeal or metaphor. Comparative studies of various appeals are scarce (De Barnier et al. 2005) and deserve much theoretical and empirical attention. In addition, most appeal studies were conducted in laboratory settings or in experiments that forced study participants to process appeals. The significant differences between our meta-analyses and QCA may be because most of our appeal results come from contrived, artificial settings. Further research should examine effects of appeals in more realistic, natural settings.

By identifying gaps in previous research, this meta-analysis will help guide the next wave of research in a promising direction. Research integration is an essential step of knowledge accumulation and refinement in science, linking past research with future scientific endeavors and providing empirical generalizations that are useful for practical marketing decisions. More meta-syntheses work on the effects of appeal types in, for example, different media or for different segments (Tellis 2004, P.103-104; Keller and Lehmann 2008) would help both researchers and practitioners. In summary, our meta-analysis has identified important needs in the study of ad appeals across segments, products, media, and countries. This analysis has been a first attempt to identify systematically features that may be relevant to message appeal. We hope that we have identified most of the important features, though some may still be missing. Further important features of appeal tasks that await research are perceptual and cognitive in nature. However, further research should no longer try to demonstrate the reference point effect but should try to clarify when and why it is present.

References

- Abernethy, Avery and George Franke (1996), "The Information Content of Advertising: A Meta-Analysis," *Journal of Advertising*, 25 (2), 1–17.
- Allen, Mike and Raymond Preiss (2007), "Media, Messages, and Meta-Analysis," *Mass Media Effect Research: Advances Through Meta-Analysis*, 15–30. Mahwah, NJ: Lawrence Erlbaum Publishers.
- Benoit, William L. (1998), "Forewarning and Persuasion." In M. Allen and R. Preiss (Eds.), *Persuasion: Advances Through Meta-Analysis*, 139–54. Cresskill, NJ: Hampton Press.
- Armstrong, Scott and Sandeep Patnaik (2009), "Using Quasi-Experimental Data to Develop Empirical Generalizations for Persuasive Advertising," *Journal of Advertising Research*, 49 (2), 170–79.
- Beard, Fred (2008), *Humor in the Advertising Business*. Browman & Littlefield Publishers.
- Borenstien, Michael, Larry Hedges, Julian Higgins, and Hannah Rothstein (2009), *Introduction to Meta Analysis N.Y.*. John Wiley Publishers.
- Calder, Bobby and Edward Malthuse (2008), "Media Engagement and Advertising Effectiveness," in Bobby Calder (ed.), *Kellogg on Advertising and Media*, 1–36. Evanston, IL: Kellogg School of Management.
- Callow, Michal A. (2000), "Do You See What I See? A Cross-Cultural Analysis of the Social Identity Metaphor in Visual Print Advertisements," Ph.D. dissertation, City University of New York, *Dissertation Abstracts International*, 61 (04), 1512A. (UMI No. 9969681).
- Cohen, Jacob (1988), *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Dahl, Darren, Jaideep Sengupta, and Kathleen Vohs (2009), "Sex in Advertising: Gender Differences and the Role of Relationship Commitment," *Journal of Consumer Research*, 36 (3), 215–16.
- Dahlen, Michael, Fredrik Lange, and Terry Smith (2010), *Marketing Communications.N.y.*, John Wiley and Sons.

- De Barnier, Virginie, Virginie Maille, Pierre Valette-Florence, and Karine Gallopel (2005), "A Cross-Cultural Study of the Persuasive Effect of Sexual and Fear Appealing Messages," *Asia Pacific Advances in Consumer Research*, 6, 140–50.
- Duval, Susan J. and R. J. Tweedie (2000), "Trim and Fill: A Simple Funnel Plot-Based Method of Testing and Adjusting for Publication Bias in Meta-Analysis," *Biometrics*, 56 (2), 276–84.
- Eisend, Martin (2006), "Two-Sided Advertising: A Meta-Analysis," *International Journal of Research in Marketing*, 23 (2), 187–98.
- Eisend, Martin (2008), "A Meta-Analysis of Humor in Advertising," *Journal of the Academy of Marketing Science*, 37 (2), 191–203.
- Farely, John, Scott Hoening, Donald Lehmann, and David Szymanski (2004), "Assessing the Impact of Marketing Strategy Using Meta Analysis," in Christine Moorman and Donald Lehmann (eds.), *Assessing Marketing Strategy Performance*. Cambridge, MA: Marketing Science Institute.
- Farley, John, Donald Lehmann, and Allen Sawyer (1995), "Empirical Marketing Generalizations Using Meta Analysis," *Marketing Science*, 14 (3), G36–G40.
- Gayle, Barbara, Raymind Priess, and Mike Allen (1998), "Another Look at the Use of Rhetorical Questions." In M. Allen and R. Preiss (Eds.), *Persuasion: Advances Through Meta-Analysis*, 189–201, Cresskill, NJ: Hampton Press.
- Glaesser, Judith, Richard Gott, Ros Roberts, and Barry Cooper (2009), "The Roles of Substantive and Procedural Understanding in Open-Ended Science Investigations: Using Fuzzy Set Qualitative Comparative Analysis to Compare Two Different Tasks," *Research in Science Education*, 39 (4), 595–624.
- Greckhamer Thomas, Vilmos Misangyi, Heather Elms, and Rodney Lacey (2008), "Using Qualitative Comparative Analysis in Strategic Management Research: An Examination of Combinations of Industry, Corporate, and Business-Unit Effects", *Organizational Research Methods*, 11 (4), 695–726.

- Grewal, Dhruv, Sukumar Kavanoor, Edwaed Fern, Carlyne Costley, and James Barnes (1997), "Comparative Versus Noncomparative Advertising: A Meta-Analysis," *Journal of Marketing*, 61, 1–15.
- Gulas, Charles S. and Mark Weinberger (2006), *Humor in Advertising. A Comprehensive Analysis*. Armonk, NY: Sharpe.
- Hanssens, Dominique M. (2009), "*Empirical Generalizations about Marketing Impact*," Cambridge, MA: Marketing Science Institute
- Heiman, Amir, Jacob Hornik, and Oded Lowengard (2010), "The Effect of Message Appeal for Risk-Reduction Campaign on Sales During Health Crisis," Working paper, Recanati Graduate School of Business, Tel-Aviv University.
- Henard, David and David Szymanski (2001), "Customer Satisfaction: A Meta-Analysis of the Empirical Evidence," *Journal of the Academy of Marketing Science*, 29 (1), 16–35.
- Hetsroni, Amir (2000), "The Relationship Between Values and Appeals in Israeli Advertising: A Smallest-Space Analysis," *Journal of Advertising*, 29 (3), 55–68.
- Higgins, Julian, Simon Thompson, Jonathan Deek, and Douglas Altman (2003), "Measuring in Consistency in Meta-Analysis," *British Medical Journal*, 327 (6), 557–60.
- Hornik, Jacob and Aner Tal (2010), "The Effect of Synchronizing Consumers' Diurnal Preferences with Time of Response on Data Reliability," *Marketing Letters*, 21 (1), 1–15.
- Hornikx, Jos and Daniel O'Keefe (2009), "Adapting Consumer Advertising Appeals to Cultural Values: A Meta-Analytic Review of Effects on Persuasiveness and Ad Liking," *Communication Yearbook*, 33, 39–72.
- Hunter, John E. and Frank L. Schmidt (2004), *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings* (2nd ed.). Newbury Park, CA: Sage.
- Keller, Punam Anand and Donald R. Lehmann (2008), "Designing Effective Health Communications: A Meta-Analysis," *Journal of Public Policy and Marketing*, 27 (2), 117–30.

- Kent, Ray (2008), "Using fsQCA: A Brief Guide and Workshop for Fuzzy-Set Qualitative Comparative Analysis," Working Paper No. 3. Manchester: Cathie Marsh Centre, University of Manchester.
- Kirca, Ahmet, Satish Jayachandran, and William O. Bearden (2005), "Market Orientation: A Meta-Analytic Review and Assessment of Its Antecedents and Impact on Performance," *Journal of Marketing*, 69 (2), 24–41.
- Kuhberger, Anton (1988), "The Influence of Framing on Risky Decisions: A Meta-Analysis," *Organizational Behavior and Human Decision Processes*, 75 (1), 23–55.
- Lee, Leonard, On Amir, and Dan Ariely (2009), "In Search of *Homo economicus*: Preference Consistency, Emotions, and Cognition," *Journal of Consumer Research*, 36 (2), 173–87.
- Nielsen Jesper, Stewart Shapiro, and Charlotte Mason (forthcoming), "Emotionality and Semantic Onsets: Exploring Orienting Attention Responses in Advertising," *Journal of Marketing Research*, in press.
- O’Keefe, Daniel (1999), "Variability of Persuasive Message Effects: Meta-Analytic Evidence and Implications," *Document Design*, 1 (2), 87–97.
- O’Keefe, Daniel and Jakob Jensen (2009), "The Relative Persuasiveness of Gain-Framed and Loss-Framed Messages for Encouraging Disease Detection Behaviors: A Meta-Analytic Review," *Journal of Communication*, 59 (2), 296–316.
- Pechmann, Cornelia and David W. Stewart (1990), "The Effects of Comparative Advertisement on Attention, Memory and Purchase Intentions," *Journal of Consumer Research*, 17 (3), 180–91.
- Picot-Zane, Leslie (2006), "Is Advertising Too Emotional?" *Brandweek*, January 9.
- Poels, Karilien, Siegfried Dewitt, and Patrick Vynake (2009), "The Dual Impact of Gender Identity on Preferences for Sex Specific Advertising Stimuli," Working paper, Department of Communication, Ghent University, Belgium.
- Rihoux, Benoit and Carles Ragin (2009), *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques (Applied Social Research Methods)*. Thousand Oaks, CA: Sage.

- Rosenthal, Robert (1984), *Meta-Analytic Procedures for Social Research*. Beverly Hills, CA: Sage Publications.
- Rosenthal, Robert and Ralf Rosnow (1991), *Essentials of Behavioral Research: Methods and Data Analysis* (2nd ed.). New York: McGraw-Hill.
- Sopory, Pradeep and James Dillard (2002), "The Persuasive Effect of Metaphor: A Meta-Analysis," *Human Communication Research*, 28 (3), 382–419.
- Stewart, David and Scott Koslow (1989), "Executional Factors and Advertising Effectiveness," *Journal of Advertising*, 18 (3), 21–32.
- Strick, Madelijn and Rick van Baaren (2009), "Why Humor Breaks Resistance to Influence?" Working paper, Behavioral Science Institute, Radboud University, the Netherlands.
- Sutton, Alexander and Julian Higgins (2008), "Recent Developments in Meta-Analysis," *Statistics in Medicine*, 27 (4), 625–650.
- Tellis, Gerald J. (2004), *Effective Advertising: Understanding When, How, and Why Advertising Works*, Thousand Oaks: Sage Publications
- Tellis, Gerald J. (2009), "Generalizations About Advertising Effectiveness in Markets," *Journal of Advertising Research*, 49(2), 17–26.
- Tellis Gerald J., Rajesh Chandy, Deborah MacInnis and Pattana Thaivanich (2005), "Modeling the Microeffects of Television Advertising: Which Ad Works, When, Where, for How Long, and Why?" *Marketing Science*, 24(3), 351-366.
- Tuan Pham, Michel (2009), "Emotion and Rationality: A critical Review and Interpretation of Empirical Evidence" *Review of General Psychology*, 11(2), 155-178.
- Unger, Lynette S. and James M. S-20 tears (1983), "The Use of Fear and Guilt Messages in Television Advertising: Issues and Evidence," in 1983 *American Marketing Association Educators' Proceedings*, ed. Patrick E. Murphy et al., 16–20. Chicago: American Marketing Association.
- Vakratsas, Demetrios and Tim Ambler (1999), "How Advertising Works: What Do We Really Know?" *Journal of Marketing*, 63 (1), 26–43.

- Vittengl Jeffery, Lee Clark, Todd Dunn, and Robin Jarrett (2007), "A Comparative Meta-Analysis of Cognitive-Behavioral Therapy's Effects," *Journal of Counseling and Clinical Psychology*, 75 (3), 475–488.
- Williams, Jerome D., Wai-Na Lee, and Curt Haugtvedt (Eds.) (2005), *Diversity in Advertising*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Witte, Kim and Mike Allen (2000), "A Meta-Analysis of Fear Appeals: Implications for Effective Public Health Programs," *Health Education and Behavior*, 27, 591–615.
- Wolf, Fredric M. (1986), *Meta-Analysis: Quantitative Methods for Research Synthesis*. Beverly Hills, CA: Sage.
- Wood, John A. (2008), "Methodology for Dealing with Duplicate Study of Effects in Meta-Analysis," *Organizational Research Methods*, 11, 79–95.
- Yang, Xiaojing and Robert Smith (2009), "Beyond Attention Effects: Modeling the Persuasive and Emotional Effects of Advertising Creativity," *Marketing Science*, 28 (5), 935–49.
- Zaltman, Gery and Lindsay Zaltman (2008), *Marketing Metaphoria*. Boston: Harvard Business School Publications.

Figure 1: **Conceptual Framework for the Meta-Analyses**

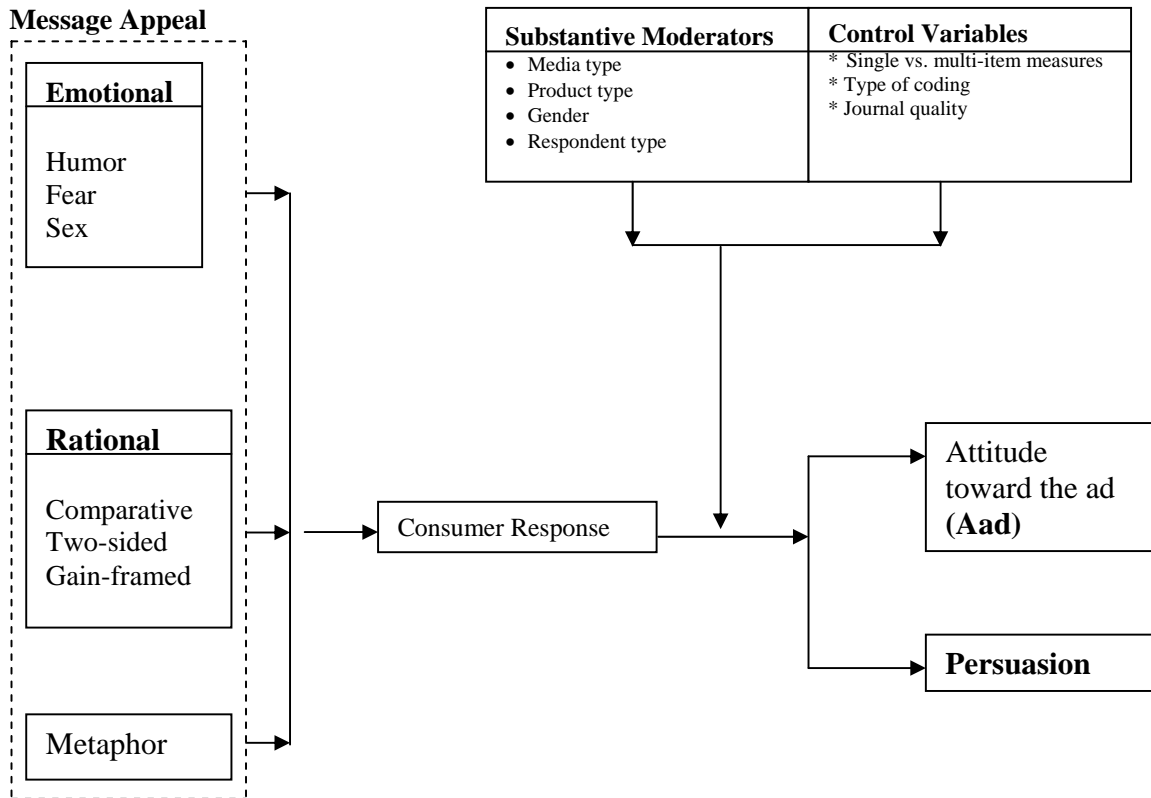


Figure 2: Plan of Statistical Analysis

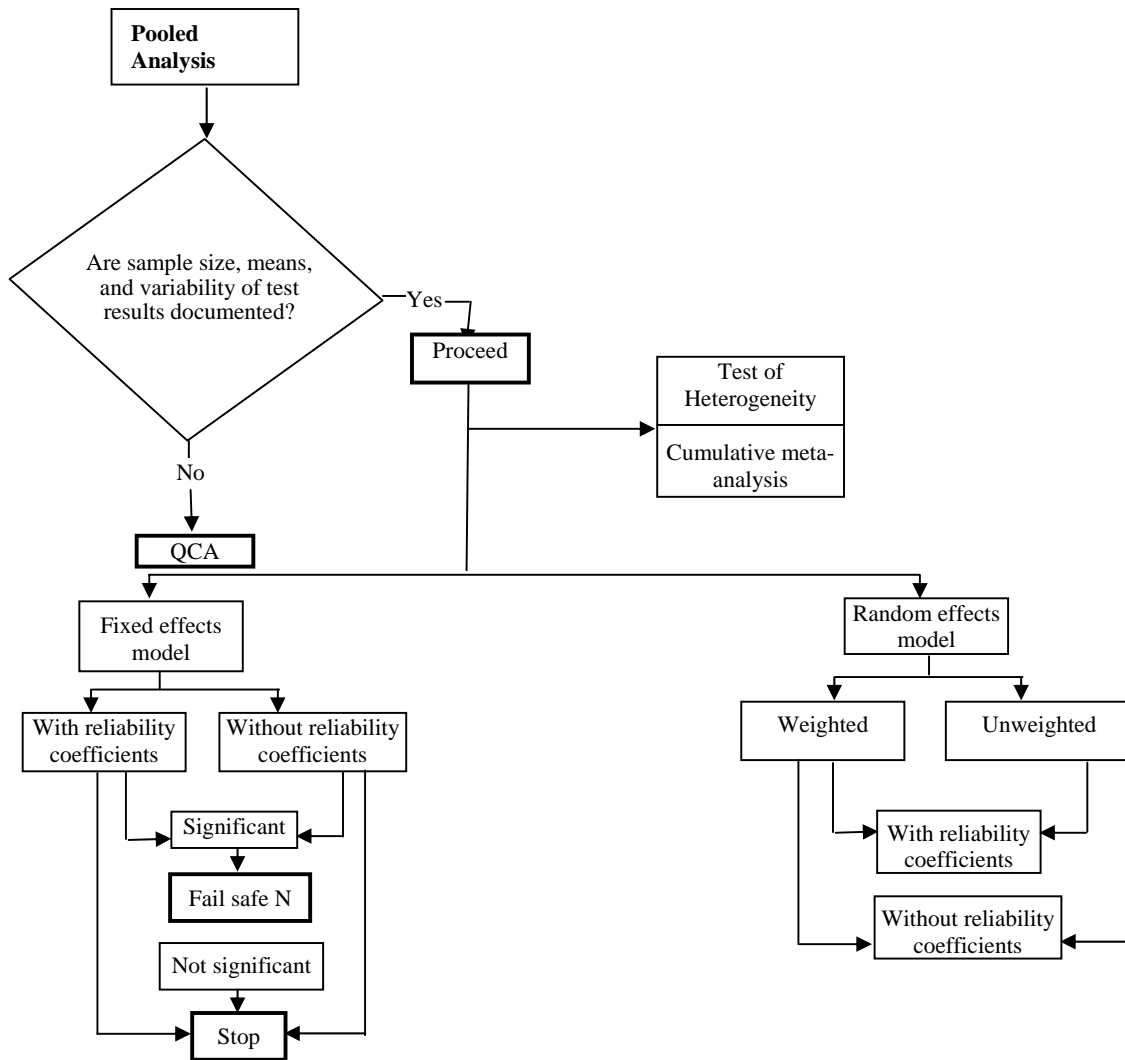


Figure 3: Combined Effect Sizes by Message Appeal

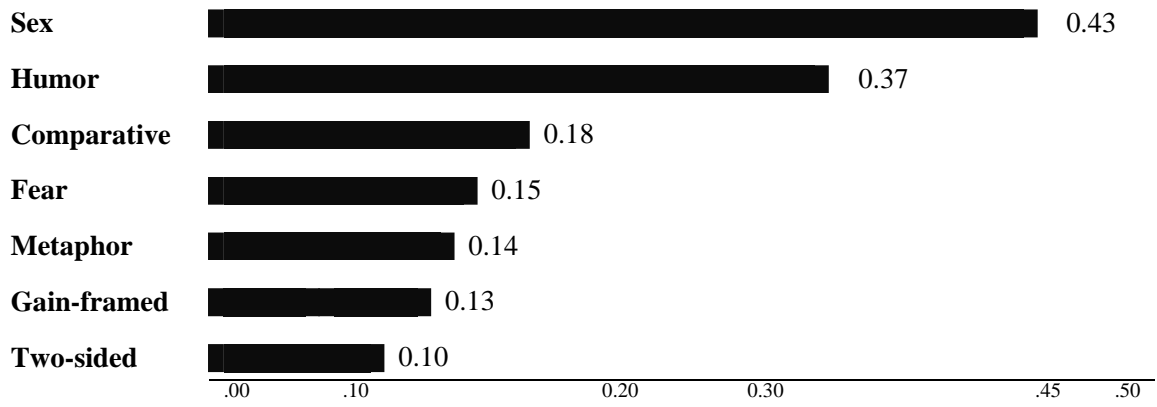


Table 1: **Results for Previous Meta-Analyses**

Appeal	Study	K Number of Studies	R Number of Correlations	N Number of Participants	r Correlations
Fear	Witt and Allen 2000	98	427	12,735	0.12-0.15
Humor	Eisend, 2008	43	369	5,663	0.37-0.42
Sex	Reichert, 2002 ^a	45	316	6,287	0.41-0.53
Comparative	Grewal et al. 1997	77	297	10,309	0.10-0.23
Gain Framed	O'Keefe & Jensen 2008	42	196	6,378	0.06-0.17
Two Sided	Eisend, 2006	26	217	3,305	0.07-0.13
Metaphor	Spory & Pillard, 2002	29	136	3,945	0.10-0.16
		360	1958	48,622	

^a We have conducted a meta-analysis for this review paper.

Table 2: Meta-Analysis Summary Statistics

Dependent Variables	K Number of studies	S Number of effect sizes	N Sample size	r	95% CI ^(a)	File-Safe M	Heterogeneity Q _w
					Lower Upper		
Overall: Aad	254	348	47,00695	0.22 ⁽²⁾	0.14-0.27	3,342	147.52 ⁽²⁾
Persuasion	449	1283	,437	0.19 ⁽²⁾	0.13-0.25	10,057	170.51 ⁽²⁾
Fear: Aad	71	93	9,676	0.13 ⁽²⁾	0.11-.014	1,725	114.34 ⁽²⁾
Persuasion	115	303	23,729	0.16 ⁽²⁾	0.13-0.18	3,239	140.11 ⁽²⁾
Humor: Aad	39	57	7,407	0.38 ⁽²⁾	0.32-0.47	516	59.25 ⁽²⁾
Persuasion	61	211	12,621	0.35 ⁽²⁾	0.30-0.44	1,522	103.84 ⁽²⁾
Sex: Aad	43	68	9,976	0.39 ⁽²⁾	0.31-0.47	533	62.26 ⁽²⁾
Persuasion	64	223	19,352	0.46 ⁽²⁾	0.35-0.52	1,235	112.18 ⁽²⁾
Comparative: Aad	42	67	6,802	0.15 ⁽²⁾	-0.30-0.20	476	60.19 ⁽²⁾
Persuasion	95	305	14,437	0.19 ⁽²⁾	0.02-0.24	1,934	122.14 ⁽²⁾
Gain Framed: Aad	26	33	5,409	0.11 ⁽¹⁾	0.06-0.16	311	32.33 ⁽¹⁾
Persuasion	49	123	9,011	0.14 ⁽²⁾	0.07-0.18	695	67.95 ⁽²⁾
Two-sided: Aad	19	22	4,278	0.09 ⁽¹⁾	0.05-0.15	286	28.26 ⁽¹⁾
Persuasion	37	97	7,321	0.11 ⁽¹⁾	0.07-0.15	421	59.62 ⁽²⁾
Metaphor: Aad	16	20	3,139	0.15 ⁽²⁾	0.12-0.18	197	24.67 ⁽¹⁾
Persuasion	36	118	6,285	0.13 ⁽²⁾	0.11-0.16	403	49.28 ⁽²⁾
Total	703	1631	142,443				

⁽¹⁾p < 0.05; ⁽²⁾p < 0.01

^(a)The 95 percent confidence interval (CI) is based on standard errors computed with the formula for artifact distribution meta-analysis (Hunter and Schmidt 2004: 207).

Table3: **Truth Table for Input Factors Related to Response Outcome (N=306)**

Input Factors	Total Cases	# of Failures	# of Successes	Percent Successes
APPEAL · MEDIA · REPETITION	127	26	101	80.2
APPEAL · repetition · budget	91	26	65	71.4
appeal · MEDIA · REPETITION · BUDGET	19	12	7	36.8
appeal	16	9	7	43.7
appeal · repetition	15	8	7	46.7
MEDIA · appeal · repetition	38	16	22	57.8
Total Cases	306	97	209	68.3

*p<.05; **p<.001 (one tailed test) denotes statistically significant differences between input mean and mean of cases not yet captured by the input factor.

NOTE: Configuration elements in capital letters indicate the presence of the characteristic. Elements in lowercase letters indicate the absence of the characteristic.

Appendix

QCA, through Boolean algebra minimization algorithms, allows for the determination of commonalities among structures and models that lead to the outcome and to generate logical statements that describe those commonalities, in turn allowing for the logical reduction of statements that are used to produce raw data tables in which each case displays a specific combination of conditions (with 0 or 1 values) and an outcome (with 0 or 1 values), and on examining the minimum combination of variables that may result in either the presence or the absence of the outcome. The QCA uses the Quine–McCluskey algorithm, a common algorithm for simplifying set-theoretic statements that is implemented in software packages, such as the fuzzy set QCA (fs/QCA) (Glaesser et al. 2009), which we used here. Using the subset principle, the fuzzy-set method enabled us to establish whether conditions were necessary or sufficient for the outcome. Thus, the method allowed us to assess whether the appeal condition was necessary and/or sufficient to the outcome, to help determine the importance of message appeal to the response measures.

In assessing causal sufficiency, QCA employs the probabilistic concept of quasi sufficiency, wherein sufficiency is assessed on the basis of certain benchmarks. A causal condition can be almost always sufficient (significantly passing a benchmark of 0.8), usually sufficient (significantly passing a benchmark of 0.65), or sufficient more often than not (significantly passing a benchmark of 0.50) in causing the outcome.

As a technique, QCA displays three further qualities. First, it can be used for at least five different purposes (De Meur and Rihoux, 2002: 78–80). The most basic use is simply to summarize data, i.e. to describe cases in a synthetic way by producing a truth table, as a tool for data exploration and typology-building. This use is basic in the sense that it does not rely on a more elaborate, step-wise design of typology-building, such as recently developed by George and Bennett (2005). It can also be used to check coherence within the data: the detection of contradictions allows one to learn more about the individual cases. The third use is to test existing theories or assumptions, to corroborate or refute these theories or assumptions. QCA is hence a particularly powerful tool for theory-testing (e.g. Sager, 2004; Goertz and Mahoney, 2005). Fourth, it can be used to test some new ideas or assumptions formulated by the researcher, and not embodied in an existing theory; this can also be useful for data exploration. Finally, QCA allows one to elaborate new assumptions or theories: the minimal formula ultimately obtained can be interpreted – i.e. confronted with the cases examined – and lead the researcher to formulate new segments of theory. This is probably why QCA is sometimes referred to as a kind of analytic induction (e.g. Hicks, 1994). QCA

is indeed inductive, to the extent that it allows the researcher to discover more through a dialogue with the data. However, there is also a significant input of theory in QCA. For instance, the selection of variables that will be used in the analysis, and the way each variable is operationalized, must be theoretically informed (De Meur and Rihoux, 2002). Arguably, though, a more inductive use of QCA raises more methodological difficulties than a simple, deductive theory-testing (Ebbinghaus, 2005).

Second, QCA is a particularly transparent technique, insofar as it forces the researcher not only to make choices on his or her own (that is, the researcher decides, not the computer), but also to justify these choices, from a theoretical and/or empirical perspective. In the course of the procedure, at several stages, the researcher is confronted with choices. For instance, he or she must decide whether or not he or she wants to obtain the shortest solution possible, to achieve a maximal level of parsimony. If this choice is made, some cases that exist logically, but that have not been observed in the data, will be included in the Boolean minimization. In practice, the software will attribute a [0] or [1] outcome value to these logical cases, thus making 'simplifying assumptions' about these cases. The researcher may reject this option, privileging complexity over parsimony. One also has to make clear choices as to the way each variable is dichotomized, and as far as the choice of variables is concerned.

Third and not least, QCA allows one to consider phenomena that vary both qualitatively and quantitatively. Both of these phenomena can be operationalized in the conditions and outcome variables used for software treatment (De Meur and Rihoux, 2002). Ragin uses the term 'qualitative' to indicate that QCA enables the researcher to analyse phenomena that vary in nature, that are present or absent, and not only in degree (Ragin, 2002), that each case is considered as a complex and specific combination of features (Ragin et al., 1996), and that QCA allows examination of constellations, configurations and conjunctures (Ragin, 1987).
Rihoux *Qualitative Comparative Analysis (QCA)*

Far from being a push-button-type technique, the use of QCA is an iterative and creative process. The researcher must first gain enough familiarity with each of the cases examined, and then produce a good-quality truth table – that is, a table devoid of contradictory configurations. These are configurations whose outcome is, in some cases, equal to [1] and in some cases equal to [0], while displaying the same values on the conditions. Such contradictions must thus be resolved before moving ahead with the analysis. This involves frequent returns to the cases, to the initial qualitative or quantitative data. This must also be done at the end of the analysis, when one finally obtains the minimal formula: the researcher must then make sense out of the solution, interpret it by reinterrogating the cases. This involves going back to the cases and examining each case as a whole. This also means that QCA can be a very labour-intensive technique – which, from a case-oriented perspective, should rather be seen as a strength.

In all following analyses, we uniformly assessed quasi sufficiency as usually sufficient (benchmark proportion = .65, [= .05]) (for more details on QCA, see Greckhamer et al. 2008; Rihoux and Ragin 2009).