Marketers' intuitions about the sales effectiveness of advertisements

Advertisements vary enormously in their sales effectiveness, so choosing the more effective creative executions to air is an important marketing task. Such decisions are often made intuitively. This study assesses the intuitive predictions of a global sample of marketers regarding which television ads are more or less sales effective. The findings show that marketers' predictions were correct no more often than random chance. Multivariate analysis suggests that those with category experience and those in marketing or consumer insights roles make slightly better predictions. Aside from *who* makes better predictions, further research is needed on *how* to improve advertising decisions, including use of evidence-based decision support systems and team decision-making.

Making better marketing decisions than one's competitors is important for business success. Marketers must often decide which ads to air and/or what ads to support with a substantial portion of their media budgets. These decisions concern one of the largest expenses for consumer brand companies, with media spend typically equivalent in value to profits (Sharp 2010). In this paper, we examine the ability of marketers to choose appropriately. Specifically, we ask a large group of marketers and advertising agency executives to predict which ads they believe are better or worse in driving product sales and compare these predictions to in-market results. Their predictions are made based on intuition, which is commonly relied upon when selecting ads to air (Rossiter and Bellman 2005).

The importance of choosing quality ads

Choosing ads is both an opportunity and risk for brands because there are large differences in the resulting effectiveness of campaigns. Jones (1995), for example, found only 20 percent of campaigns achieved strong short-term sales effects. Wood (2009) more recently found some ads could be 10 to 20 times more sales effective than others for the same brand. The challenge for advertisers is to consistently identify and air more effective ads.

Advertising design is a complex process given the tremendous number of creative devices available (Burke et al. 1990). Stewart and Furse (1986) identified over 150 creative devices that still characterize current advertising practice (Hartnett et al. 2016). However, few advertisers document or access empirical evidence regarding what creative devices work more often and under what conditions (Armstrong 2010; Ewing and West 2000). Some marketers even preclude the possibility of generalizable knowledge, leading to the common "mine is different" mantra (Binet and Carter 2013).

Several studies have documented how the advertising development process is managed and who is involved (e.g. Hill and Johnson 2004; Michell 1988; Na et al. 2009;

Turnbull and Wheeler 2015). These studies indicate the types of decisions made but provide no evidence of when better ads are produced. Understanding of the quality of advertising design decisions and how those decisions are made (intuitively or analytically) is sorely lacking.

How good are marketers at predicting more effective ads?

The limited research in this area is surprising when one of the enduring concerns of marketers is how to make great advertising (Tharpe 2013). Published empirical evidence of marketers' judgments related to advertising effectiveness is rather dated. One study asked 83 advertising managers to predict the performance of 24 newspaper ads (Bogart et al. 1970). Managers predicted readership quite well (r=0.64 for recall, r=0.55 for recognition) but not sales performance (r=0.06). Another study asked 270 attendees at a direct mail conference to predict the more effective ads for 10 split-run mailing tests (McCorkell 1985). No individual predicted the outcome for more than seven tests. These studies, however, are both decades old, related only to print advertising, and the latter is more qualitative. Today's marketers may be better decision makers. They have access to sophisticated measures of advertising effects (e.g. marketing mix models, single-source analysis), so have more opportunity to learn what works, hence, further research in this space is thus overdue.

Our study is a large-scale test of how well marketing decision makers' predict the sales effectiveness of television ads. We compare professional marketers' judgments for several pairs of ads with known in-market sales effects. We compare groups based on work experiences (e.g. employer, role, tenure) to determine whether certain types of experience lead to better decisions. We also compare marketers' predictions to those of novices to examine whether *any* experience leads to better judgment. But first, it is useful to discuss decision-making and how it applies to advertising.

Intuition and decision-making

Decision-making involves two parallel cognitive systems: the intuitive and the analytical. We focus on the intuitive, considered the predominant basis for advertising decisions. Intuition is a form of reasoning developed from years of experience and learning that allows a person to identify and apply meaningful patterns to support present decisions (Matzler et al. 2007). Those with good intuition can encode complex information faster because of their highly organized, domain-specific knowledge, which is presumed to result in better decisions (Blattberg and Hoch 1990).

Researchers have studied judgment and decision-making to map the conditions that separate genuine intuitive skill from overconfident impressions (Kahneman and Klein 2009). Intuitive predictions often outperform analytical ones (Andersen 2000; Harteis and Gruber 2008; Khatri and Ng 2000; Sadler-Smith 2004), though there is also much evidence of intuitive predictions being highly inaccurate, irrespective of experience or expertise (Armstrong 1991; Armstrong 1980; Goldberg 1959; Hoch 1988; Karelaia and Hogarth 2008; Tetlock 2005), especially when compared to mechanical or statistical predictions (Franses 2009; Goldberg 1970; Grove et al. 2000; Meehl 1954; Sawyer 1966). Evidence of the superiority of statistical predictions comes mostly from research in medicine, clinical psychology, emergency response, financial and legal services, and politics. It cannot be assumed, however, that findings from non-business domains generalize to the advertising context, given that each field has its own unique issues (Wierenga 2011). Empirical testing is necessary to understand how accurate intuitions are for advertising decisions.

Intuition and decision-making in advertising

The advertising decision maker's environment likely affects how they might make (better) judgments. Hammond's Cognitive Continuum theory provides evidence that

cognitive modes range from pure intuition to pure analysis and that characteristics of the environment stimulate and interact with these cognitive modes (Hammond 2000; Hammond 1980; Hammond et al. 1987). There are several cognitive and task properties that determine a person's location on either continuum. Given the characteristics of the advertising environment, it appears to be predominantly intuition inducing. When evaluating ads, which often incorporate many different audiovisual creative devices, decision makers must utilize many information cues that are presented simultaneously and pictorially (with some exceptions, e.g. radio ads). The interpretation of these cues in regards to perceived creativity (originality, appropriateness, etc.) is largely subjective (Nyilasy and Reid 2009b). Academics continue to debate the most appropriate models of how advertising works (e.g. Ehrenberg et al. 2002; Heath and Feldwick 2008; Jones 1997; Vakratsas and Ambler 1999) and marketers' beliefs do not really resemble prominent theories (Nyilasy and Reid 2009a). The effects of advertising, particularly on sales, can be hard to isolate and predict because of numerous additional influences, including price, promotions, distribution, competitor response, economic conditions, etc.

Given such complexity, it is not surprising that advertisers prefer to use intuition and personal preferences (Rossiter and Bellman 2005). Qualitative research has found advertising agency staff prioritize skills acquisition and intuition as two mental models (both precluding a more rational, evidence-based approach) (Nyilasy et al. 2013). Skills acquisition employs experiential knowledge, where performance improves through trial and error. The intuition model, which follows skills acquisition, reflects expertise accumulated from experiential learning, often described as gut reactions that "this is the right thing to do" without further explanation (Nyilasy et al. 2013). Though popular, using management's intuitive judgment as the basis for choosing which ads to air remains contentious. Ogilvy (1983) originally spoke out against the industry's abhorrence of research and rules; *"Why should our clients be*

expected to bet millions of dollars on your intuition? […] I cannot think of any other profession that gets by on such a small corpus of knowledge. […] What is the reason for this failure to codify experience? […] Is it that any kind of scientific method is beyond the grasp of 'creative' people?" (p. 21), but such views have had limited impact on industry practice.

Cognitive Continuum Theory further predicts that the effectiveness of people's judgement is contingent on matching the task's properties and the cognitive mode used. Using intuition when the task calls for it should lead to more accurate predictions, which supports advertising creatives' enthusiasm for intuition before analysis. Copy testing, for example, is disliked because it is believed to compromise creativity (Hackley and Kover 2007). Analysis of case studies submitted to the UK Institute of Practitioners in Advertising (IPA) Effectiveness Awards found that copy tested campaigns won awards less often (40 versus 74 percent) and reported "very large business effects" less often (43 versus 68 percent) (Binet and Field 2007). Other research has found when people are forced to use rational decision-making when judging advertising, decisions get worse. Explicit reasoning degraded predictions of how likeable ads were (intuitive task) but enhanced predictions of annual cigarette consumption (analytic task) (McMackin and Slovic 2000). These findings endorse the positive effect of matching task properties and intuitive thinking in advertising; however, more tests of the effectiveness of professional marketers' intuitive predictions are still needed.

Another consideration is the ability to learn good intuition. Hogarth's (2001) conceptualization of learning structures suggests intuition is more accurate when environments are kind (i.e. feedback that links judgments to outcomes is plentiful) and less so when environments are wicked (i.e. feedback is noisy, delayed or uncertain). Feedback in wicked environments does not help us learn. In advertising, feedback on the effectiveness of ads is often lacking or difficult to interpret, suggesting a rather wicked context for decision-making. Measuring advertising effects is standard practice for large advertisers but is not

universal, with many measures also arguably flawed (e.g. Kennedy and Northover forthcoming). Because ads use creative devices in different combinations it is difficult to discern what combinations are driving effects and inconsistent use of different measures (e.g., ad recognition, brand recall, emotional response, etc.) complicates issues further. Additional complexity arises when relationships change across markets (e.g. product category, brand size, market age, culture) (Armstrong 2010).

Product category is one condition that can be consistent for a time, allowing marketers to develop more accurate intuitions. Marketers that work for a biscuit brand, for example, who see lots of ads with many similar creative devices linked to some measure of advertising effect might be better able to predict outcomes for biscuit ads than other ads. Testing is needed to know for sure.

Research objectives

We sought to document how well marketers make intuitive advertising decisions. If marketers make effective intuitive decisions then there is little need to introduce time consuming and costly analytical or evidence-based approaches. However, if intuitive advertising decisions are poor, this finding justifies the adoption of evidence-based support systems to help marketers make better decisions. The objectives of this empirical study are: (1) to investigate how accurate marketers are at identifying sales effective ads when relying on their intuition; and (2) to determine *who* makes better decisions, by comparing decisions across different groups based on their experience, for example, their employer (manufacturer or advertising agency), job role (marketing, insights), tenure, and category-specific experience, whilst also comparing to a sample of novices as a baseline.

Marketers have long focused on intermediate measures of advertising effects to assess campaign effectiveness, such as how creative or memorable an ad is. The problem is that

intermediate measures are poor predictors of actual behavior and in-market sales (e.g. Bogart et al. 1970; Haley and Baldinger 1991; Kuse 1991; Lodish et al. 1995). Because businesses seek financial accountability from marketing (Stewart 2009), advertising outcomes linked to profit and sales-based measures are the most appropriate basis for assessing advertising decisions (Eisenstein and Lodish 2002; Van Bruggen et al. 1998). It is a harder test to predict sales, as discovered by Bogart et al. (1970), but given the preference for financial measures the target variable used in this study is short-term sales effectiveness.

Method

The approach had marketers watch pairs of ads and judge which ad in the pair was more sales effective. Using ad pairs creates a quasi-experiment because many contextual variables (e.g. product, brand size, target market, etc.) are controlled for, even though the creative devices used by the ads vary (Armstrong 2010). Quasi-experiments are seldom used to analyze advertising effects but are a valid approach to developing knowledge about how advertising works (Armstrong and Patnaik 2009) or, in this case, testing marketers' intuition.

Selecting the ad stimuli

A global consumer packaged goods (CPG) manufacturer provided eight pairs of matched television ads (16 ads in total) from four product categories, including two staple and two impulse categories (two pairs apiece). Ads in each pair were from the same country; seven were from the UK and one the US. The four categories were well established in these countries and all ads were for established brands. Ads in the pairs were for the same brand for five of eight pairs (i.e. two ads for Brand A, aired at different times). For the other three pairs, the ads were for competing brands that had similar market shares and positioning.

The sales effect for each ad was measured in-market with longitudinal individual-level single-source data. Single-source data continuously records product purchases and advertising

exposures for households. From such data a brand's share of category purchases preceded by advertising exposure(s) can be compared to the share of purchases not preceded by advertising exposure(s) (i.e. two separate groups of households), effectively extracting an advertising experiment (for the history of single-source advertising research see Jones 1995; McDonald 2000, or for examples of how such data can be used to answer advertising questions see Hartnett et al. 2016; Newstead et al. 2009; Taylor et al. 2013; Taylor et al. 2009). Sales effectiveness of the ads was determined by the manufacturer using proprietary analysis that accounts for other variables including frequency of advertising exposure in the weeks prior to purchase and purchases made on discount. Resulting sales indexes therefore isolate brand sales changes driven by the specific advertising execution alone. The relative sales effectiveness of the ads in all pairs was substantively different. One ad was well above the average sales effectiveness index for the product category and country (i.e. relative to the sales effects all other ads in the period) and the other was well below.

Sample of marketers and agency executives

Senior management from the manufacturer contacted all marketers in its employ globally as well as personnel from two affiliated advertising agencies. These are people who make decisions related to their brand portfolio. They were sent a link to the online survey, which was positioned as an opportunity for marketers to test their skills and receive feedback. After two weeks of data collection 616 marketers and 82 agency personnel had participated (response rates were 35 and 100 percent respectively).

Respondents watched five randomly assigned pairs of ads to keep the test short and encourage completion. They were only informed of the brand name(s), ad titles, and the country in which the ads were aired. Respondents were asked if they already knew how one or both of the ads performed in-market immediately after watching each pair. If they did, they were skipped to the next pair to ensure the integrity of the prediction data. If not, they were

asked which of the two ads they thought was more effective at driving sales. The online approach meant that the respondents made their judgments alone and because they were forced to make decisions for ads they were not familiar with, with limited information, and did so reasonably quickly, we can infer that the predictions are based on intuition.

Results

Marketers and agency executives

The accuracy of predictions across the different professional groups is shown in Table 1. The 698 professionals made 2159 predictions, which were almost uniformly distributed between the eight pairs of ads, ranging from 262 to 285 responses per pair. Of all predictions, 51 percent were correct. In aggregate, judgments were no more accurate than guessing with a 50 percent probability of being right, t (2158) = 1.01, p = .31.

Combining independent predictions is a well-established practice in forecasting, often leading to more accurate predictions (see Armstrong et al. 2015). We combined our marketers' predictions to evaluate if this would produce better predictions. To explain, predictions were combined to create a modal prediction for each pair, where the group's collective judgments were used to classify one ad as more effective than the other, equivalent to a voting majority. If, for example, 60 percent of people indicated Ad A was more effective than Ad B, then the modal prediction would select Ad A as the winner. This approach reduces the 2159 individual predictions down to eight modal predictions. Results for modal predictions are also in Table 1.

Managers' modal predictions identified the more sales effective ads significantly better than chance for three of eight pairs (p< .05, i.e. Pairs 5, 6, and 8 highlighted in Table 1). If we put statistical significance aside, modal predictions were correct (above 50%) for a

further two pairs (Pairs 1 and 2). The results indicate managers are not very good at judging ads and there is no "wisdom of the crowd" effect, as group prediction is not much better.

[Table 1 here]

Comparing the types of managers, chi-square tests revealed some differences. Managers with experience working for the advertised category made correct predictions significantly more often than managers without category experience, χ^2 (1) = 10.58, p< .001. Managers working in marketing or consumer insights roles made correct predictions significantly more often than managers working in marketing operations roles, χ^2 (1) = 6.93, p< .01, and 'other' roles, χ^2 (1) = 4.97, p< .05 (note, as these 'other' managers have varied experiences, it is difficult to draw a meaningful comparison here).

Prediction accuracy was similar across other managerial experience, including managers from the manufacturer versus the advertising agencies, χ^2 (1) = 1.71, p = .19, and managers that were responsible for approving ads to air versus those that were not, χ^2 (1) = 0.36, p = .55. One-way ANOVA showed no significant differences between six tenure groups, F (5,2148) = 0.44, p = 0.82. It cannot be assumed, however, that managers who have worked for the manufacturer for less time are universally less experienced. It is possible that managers with shorter tenure may have had long careers at other companies, though they should be much more familiar with advertising for the manufacturer.

The chi-square analysis provides insight into the characteristics of more successful advertising decision makers. A multivariate approach allows us to examine whether any characteristic dominates or interacts with others. Two binary logistic regression models were estimated, with the prediction acting as the dependent variable (correct or incorrect). Two models had to be run because not all experience variables were collected from all respondents. Model 1 includes predictions made by marketers employed by the manufacturer and executives employed by the advertising agencies. These two groups are compared, also

considering their tenure and category experience. Model 2 includes predictions made by marketers employed by the manufacturer only, considering their specific marketing role and responsibility for approving ads (both of which were irrelevant to agency executives), as well as tenure and category experience. The results must be interpreted cautiously, as sample sizes and considerable risk of heterogeneity complicate interpretation. There was particular concern for sample sizes across the tenure groups, so tenure was collapsed to a dichotomous variable of less tenure versus more tenure (up to 7 years versus 8 or more years). The cut point was based conservatively on the expectation that it takes at least 10,000 hours of practice to become an expert (Chase and Simon 1973; Ericsson 2006; Gladwell 2008), which equates to about 5 to 6 years tenure (i.e. at 40 hours per week, 46 weeks per year).

[Table 2a and 2b here]

The binary logistic regressions are in line with the chi-square analysis. In Model 1 (Table 2a) it was found having prior experience working within the product category was associated with better predictions (B = .15, p = .06). The results confirm the result that tenure was not associated with prediction accuracy, nor was whether the employee worked for the manufacturer or agency (p> .05). Nor did any of the interactions between these variables (p> .05). In Model 2 (Table 2b) it was found that the role within the organization was associated with prediction accuracy, with those working in marketing and consumer insights exceeding other roles, and significantly exceeding the mixed 'other' group (B = .288, p = .05). Consistent with Model 1, experience within the product category led to better predictions also (B = .21, p= .02). Surprisingly, tenure with the organization and whether the employee had the responsibility to approve advertising was not associated with the accuracy of marketers' predictions (p> .05). This may be partly due to the exclusion of predictions of those who already knew the performance of ads in the pairs.

Comparing to Novices

A novice sample was collected as a baseline to compare managers' predictions to. The same survey was given to undergraduate students enrolled in an advertising course. The survey was run at the beginning of the first lecture, before students were taught any advertising theories or principles. Students were given the incentive of course credit for participation; 191 of 257 students participated (74 percent). Ten students were excluded because they claimed to have experience in a marketing or advertising role, leaving 181 respondents. Even though marketers' intuitive predictions were generally no better than chance they were better than novices. Only 44 percent of students' predictions were correct, which is worse than chance, $t\ (1447) = -4.50$, $p < .001$, and worse than the managers, $\chi^2\ (1) = 16.81$, $p < .001$. This result suggests marketers' experience marginally improves their judgment over novices. We suspect that the poorer performance by students could relate to their level of engagement with the task, their lack of knowledge about how advertising works, and/or insufficient understanding of broader category purchase mechanisms. Given that some of the brands were also not available where the students lived, it may relate to their brand specific knowledge. However, we are only speculating as to what matters. Investigation beyond this single study is needed to understand what matters, including checking a variety of stimuli and conditions to understand if this is a robust pattern.

Discussion

This study examines the ability of marketers to intuitively predict which ads they should air for established brands. Contrary to what advertising professionals might believe, and what a lot of research in advertising creativity suggests, intuitive decision-making did not perform well. The results, split a variety of ways, suggest that once someone achieves a minimum threshold of experience, further experience or expertise count for less than might be

expected. This finding in this context is consistent with much of the literature about the inaccuracies of intuition and human judgment more broadly.

When predictions were combined, some of the pairs appeared easier for marketers to predict, collectively, than others. Reflecting on why some pairs were easier to judge is much like guess work given that we are limited by a very small sample of ads, which included many varied tactics within them. However, one pattern we note is that the three pairs predicted correctly (significantly better than chance) were all from staple product categories, suggesting sales resulting from ads for impulse products might be harder for managers to predict. There is a substantive research agenda ahead to further investigate conditions where intuition may be better or worse.

These findings should warn advertisers and their agencies not to over-estimate even experienced managers' intuition for making sound judgments about advertising. Relying solely on intuition appears to be high risk for selecting what ads to air. The findings indicate a relatively low base from which advertisers can improve when it comes to judging the potential sales effectiveness of television ads.

Our study shows that research on unaided judgmental forecasts about complex, uncertain situations (e.g. Tetlock, 2005) generalizes to the intuitive predictions of marketers making advertising decisions. The results suggest advertisers' problems in picking effective ads aren't so different from forecasting problems in other fields, and that formal systems may be needed to replace or augment intuition. It justifies the push for research and development of decision systems to help marketers make better advertising decisions. Drawing on the growing literature on decision support systems is likely to provide useful insights for marketers willing to take this journey (Van Bruggen and Wierenga 2010; Wierenga and Van Bruggen 2000). Marketing expert systems may also help to alleviate the shortage of marketing expertise by incorporating knowledge in a medium that allows it to be distributed

easily and inexpensively (Burke et al. 1990). Similarly, the use of evidenced-based principles have been proposed to improve advertising creativity and persuasiveness (Armstrong 2010). The literature does, however, indicate substantial barriers to the adoption of such decision support systems in practice, primarily because of issues related to usability and managers' reluctance to concede that they need such tools (Eisenstein and Lodish 2002; Weiss and Shanteau 2014). With more evidence of the inaccuracy of many advertising decisions, such as this study, advertisers and their advertising agencies should become more realistic about the flaws of intuition and open to exploring decision support systems to facilitate more effective advertising.

To be adopted and subsequently maintained, however, managers will also require evidence that decision aids do indeed improve decision-making. We cannot neglect that feeding success back to managers is a critical component to reinforcing their use of such tools. Lack of validation evidence is considered one of the major reasons for the downfall of many marketing decision support systems (Eisenstein and Lodish 2002). In the field of advertising, further evidence that decision aids do not prevent or dampen creativity is also likely to be necessary to get broad adoption.

Other potential ways to improve marketers decisions include educating marketers intuitions in kinder environments, with specific guidelines available to support this (e.g. seek feedback, acknowledge emotions and make scientific method intuitive) (Hogarth 2003). For many this will require spending more on advertising testing (e.g. Tharpe (2013) reported half of very senior marketing executives spend less than 1% of their budgets on measurement), as well as implementing centralized systems documenting what is advertised and what response campaigns get in-market. Ideally, a culture of conducting natural experiments, online or in-market, would also improve learning.

It may also be beneficial to bring together people with different skills or experience to facilitate collective expertise. In domains such as politics, team decisions can outperform individual decisions as managers collaborate and comment on each other's thoughts (Ungar et al. 2012). Our managers were forced to make decisions alone, but advertising design is often a collaborative process between the advertiser and advertising agency. Advertisers might be best served by forming teams to liaise with advertising agencies. Testing of who should be in such teams is needed, such as including internal and/or external experts, how many people and from what roles, or what traits or skills do particular individuals need.

Our findings suggest that potentially valuable members of advertising decision-making teams may include people with experience working for the product category or that work in marketing and consumer insights roles. Our findings show these particular types of marketers were better judges, with their prior experience giving them a small but significant intuitive edge. What may be driving their superior intuition is continued exposure to similar ads for the same products, or familiarity with when, how, and what motivates people buy from the category, or their greater exposure to research methods and findings. Although the differences between these professionals and their counterparts was small, where one group outperformed the other by a margin of 10 percent or less, these small differences would add up to large financial gains over time in regards to sales driven by advertising.

Other research, however, suggests that assessments should be made individually and independently and then pooled to extract the "wisdom of the crowd" (Surowiecki 2004), akin to the grouping method tested in this study. Advertisers are unlikely to ever have crowds or panels of expert judges as large as ours to select what ads to run. There is much work to be done to discover when and where this approach works best and how it might be optimally implemented (see Graefe and Armstrong 2011). In the growing digital advertising world, split cable live tests may provide crowd feedback.

Intuition will always contribute to advertising design decisions but as advertising becomes more accountable, intuition should no longer be sufficient to fully justify those decisions. Our study contributes evidence to the growing body of literature about how professionals approach the advertising process and also establishes a baseline from which to assess how much decision support tools might improve advertising decision-making.

References

Andersen, Jon Aarum (2000), "Intuition in Managers: Are Intuitive Managers More Effective?," *Journal of Managerial Psychology*, 15 (1), 46-63.

Armstrong, J. Scott (2010), *Persuasive Advertising: Evidence-Based Principles*. Basingstoke: Palgrave Macmillan.

---- (1991), "Prediction of Consumer Behavior by Experts and Novices," *Journal of Consumer Reserach*, 18 (2), 251-56.

---- (1980), "The Seer-Sucker Theory: The Value of Experts in Forecasting," *Technology Review*, 82 (7), 16-24.

Armstrong, J. Scott, Kesten C. Green, and Andreas Graefe (2015), "Golden Rule of Forecasting: Be Conservative," *Journal of Business Research*, 68 (8), 1717-31.

Armstrong, J. Scott and Sandeep Patnaik (2009), "Using Quasi-Experimental Data to Develop Empirical Generalizations for Persuasive Print Advertising," *Journal of Advertising Research*, 49 (2), 170-75.

Binet, Les and Sarah Carter (2013), "Mythbuster: The Mantra That 'Mine Is Different'," in Admap Vol. November. Online: Warc.

Binet, Les and Peter Field (2007), *Marketing in the Era of Accountability*: World Advertising Research Centre.

Blattberg, Robert C. and Stephen J. Hoch (1990), "Database Models and Managerial Intuition: 50% Model+ 50% Manager," *Management Science*, 36 (8), 887-99.

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

Burke, Raymond R., Arvind Rangaswamy, Jerry Wind, and Jehoshua Eliashberg (1990), "A Knowledge-Based System for Advertising Design," *Marketing Science*, 9 (3), 212-29.

Chase, William G. and Herbert A. Simon (1973), "The Mind's Eye in Chess," in *Visual Information Processing*, William G. Chase, Ed. New York: Academic Press.

Ehrenberg, Andrew, Neil Barnard, Rachel Kennedy, and Helen Bloom (2002), "Brand Advertising as Creative Publicity," *Journal of Advertising Research*, 42 (4), 7-18.

Eisenstein, Eric M. and Leonard Lodish (2002), "Marketing Decision Support and Intelligent Systems: Precisely Worthwhile or Vaguely Worthless?," in *Handbook of Marketing*. London: Sage Publications.

Ericsson, K. Anders (2006), "The Influence of Experience and Deliberate Practice on the Development of Superior Expert Performance," in *The Cambridge Handbook of Expertise and Expert Performance*, K. Anders Ericsson and Neil Charness and Paul J. Feltovich and Robert R. Hoffman, Eds. New York, US: Cambridge University Press.

Ewing, Michael T. and Douglas West (2000), "Advertising Knowledge Management: Strategies and Implications," *International Journal of Advertising*, 19 (2), 225-43.

Franses, Philip Hans (2009), "Can Managers Judgmental Forecasts Be Made Scientifically?," *Foresight: The International Journal of Applied Forecasting* (15), 32-36.

Gladwell, Malcolm (2008), *Outliers: The Story of Success*. New York, US: Little Brown and Company.

Goldberg, Lewis R. (1959), "The Effectiveness of Clinicians' Judgments: The Diagnosis of Organic Brain Damage from the Bender-Gestalt Test," *Journal of Consulting Psychology*, 23 (1), 25-33.

---- (1970), "Man Versus Model of Man: A Rationale, Plus Some Evidence, for a Method of Improving on Clinical Inferences," *Psychological Bulletin*, 73 (6), 422-32.

Graefe, A. and J. Scott Armstrong (2011), "Comparing Face-to-Face Meetings, Nominal Groups, Delphi and Prediction Markets on an Estimation Task," *International Journal of Forecasting*, 27 (1), 183-95.

Grove, William M., David H. Zald, Boyd S. Lebow, Beth E. Snitz, and Chad Nelson (2000), "Clinical Versus Mechanical Prediction: A Meta-Analysis," *Psychological Assessment*, 12 (1), 19-30.

Hackley, Chris and Arthur J. Kover (2007), "The Trouble with Creatives: Negotiating Creative Identity in Advertising Agencies," *International Journal of Advertising*, 26 (1), 63-78.

Haley, Russell and Allan Baldinger (1991), "The Arf Copy Research Validity Project," *Journal of Advertising Research*, 31 (2), 11-32.

Hammond, Kenneth R. (2000), "The Cognitive Continuum Theory of Judgment," in *Judgments under Stress*. New York, US: Oxford University Press.

---- (1980), "The Integration of Research in Judgment and Decision Theory," in Technical Report No. 226. Boulder: University of Colorado, Center for Research on Judgment and Policy.

Hammond, Kenneth R., Robert M. Hamm, Janet Grassia, and Tamra Pearson (1987), "Direct Comparison of the Efficacy of Intuitive and Analytical Cognition in Expert Judgment," *IEEE Transactions on Systems, Man, and Cybernetics*, 17 (5), 753-70.

Harteis, Christian and Hans Gruber (2008), "Intuition and Professional Competence: Intuitive Versus Rational Forecasting of the Stock Market," *Vocations and Learning*, 1 (1), 71-85.

Hartnett, Nicole, Rachel Kennedy, Byron Sharp, and Luke Greenacre (2016), "Creative That Sells: How Advertising Execution Affects Sales," *Journal of Advertising*, 45 (1) 102-112.

Heath, Robert and Paul Feldwick (2008), "Fifty Years Using the Wrong Model of Advertising," *International Journal of Market Research*, 50 (1), 29-59.

Hill, Railton and Lester W. Johnson (2004), "Understanding Creative Service: A Qualitative Study of the Advertising Problem Delineation, Communication and Response (Apdcr) Process," *International Journal of Advertising*, 23 (3), 285-307.

Hoch, Stephen J. (1988), "Who Do We Know: Predicting the Interests and Opinions of the American Consumer," *Journal of Consumer Research*, 15 (3), 315-24.

Hogarth, Robin M. (2001), *Educating Intuition*. Chicago, Illinois: University of Chicago Press.

---- (2003), "Educating Intuition: A Challenge for the 21st Century," Vol. 13. Barcelona: Universitat Pompeu Fabra.

Jones, John Philip (1997), "Is Advertising Still Salesmanship?," *Journal of Advertising Research*, 37 (3), 9-15.

---- (1995), *When Ads Work: New Proof That Advertising Triggers Sales* (1 ed.). New York: Lexington Books.

Kahneman, Daniel and Gary Klein (2009), "Conditions for Intuitive Expertise," *American Psychologist*, 64 (6), 515-26.

Karelaia, Natalia and Robin M. Hogarth (2008), "Determinants of Linear Judgment: A Meta-Analysis of Lens Model Studies," *Psychological Bulletin*, 134 (3), 404-26.

Kennedy, Rachel and Haydn Northover (Forthcoming), "Using Neuro Measures for Better Advertising Decisions: Issues for Practitioners and Research Priorities for Vendors and Scholars," *Journal of Advertising Research*.

Khatri, Naresh and H. Alvin Ng (2000), "The Role of Intuition in Strategic Decision Making," *Human Relations*, 53 (1), 57-86.

Kuse, Allan R. (1991), "Measurement Tools for Ads That Sell," in Copy Research The New Evidence - Eighth Annual ARF Copy Research Workshop. New York: Advertising Research Foundation.

Lodish, Leonard M., Magid Abraham, Stuart Kalmenson, Jeanne Livelsberger, Beth Lubetkin, Bruce Richardson, and Mary Ellen Stevens (1995), "How Tv Advertising Works: A Meta-Analysis of 389 Real World Split Cable Tv Advertising Experiments," *Journal of Marketing Research*, 32 (2), 125-39.

Matzler, Kurt, Franz Bailom, and Todd A. Mooradian (2007), "Intuitive Decision Making," *MIT Sloan Management Review*, 49 (1), 13-15.

Mccorkell, Graeme (1985), "When Experts Can Get It Wrong." London: MSW Rapp and Collins.

Mcdonald, Colin (2000), "How Single-Source Data Research First Developed," in *International Advertising - Realities and Myths*, John Philip Jones, Ed.: Sage Publications.

Mcmackin, John and Paul Slovic (2000), "When Does Explicit Justification Impair Decision Making?," *Applied Cognitive Psychology*, 14 (6), 527-41.

Meehl, Paul E. (1954), *Clinical Versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence*. Minneapolis, MN: University of Minnesota Press.

Michell, Paul (1988), "Where Advertising Decisions Are Really Made," *European Journal of Marketing*, 22 (7), 5-18.

Na, Woonbong, Roger Marshall, and Arch G. Woodside (2009), "Decision System Analysis of Advertising Agency Decisions," *Qualitative Market Research: An International Journal*, 12 (2), 153-70.

Newstead, Kate, Jennifer Taylor, Rachel Kennedy, and Byron Sharp (2009), "The Long-Term Sales Effects of Advertising: Lessons from Single Source," *Journal of Advertising Research*, 49 (2), 207-10.

Nyilasy, Gergely, Robin Canniford, and Peggy J. Kreshel (2013), "Ad Agency Professionals' Mental Models of Advertising Creativity," *European Journal of Marketing*, 47 (10), 1691-710.

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

---- (2009b), "Agency Practitioners' Meta-Theories of Advertising," *International Journal of Advertising*, 28 (4), 639-68.

Ogilvy, David (1983), *Ogilvy on Advertising*. London: Pan Books Ltd.

Rossiter, John R. and Steven Bellman (2005), *Marketing Communications: Theory and Applications*. Frenchs Forest: Pearson Education.

Sadler-Smith, Eugene (2004), "Cognitive Style and the Management of Small and Medium-Sized Enterprises," *Organization Studies*, 25 (2), 155-81.

Sawyer, Jack (1966), "Measurement and Prediction, Clinical and Statistical," *Psychological Bulletin*, 66 (3), 178-200.

Sharp, Byron (2010), *How Brands Grow*. Melbourne: Oxford University Press.

Stewart, David W. (2009), "Marketing Accountability: Linking Marketing Actions to Financial Results," *Journal of Business Research*, 62 (6), 636-43.

Stewart, David W. and David H. Furse (1986), *Effective Television Advertising: A Study of 1000 Commercials*. Lexington, MA: Lexington Books.

Surowiecki, J. (2004), *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*: Doubleday.

Taylor, Jennifer, Rachel Kennedy, Colin Mcdonald, Laurent Larguinat, Yassine El Ouarzazi, and Nassim Haddad (2013), "Is the Multi-Platform Whole More Powerful Than Its Separate Parts? Measuring the Sales Effects of Cross-Media Advertising," *Journal of Advertising Research*, 53 (2), 200-11.

Taylor, Jennifer, Rachel Kennedy, and Byron Sharp (2009), "Making Generalizations About Advertising's Convex Sales Response Function: Is Once Really Enough?," *Journal of Advertising Research*, 49 (2), 198-200.

Tetlock, Philip E. (2005), *Expert Political Judgment: How Good Is It? How Can We Know?* Princeton, NJ: Princeton University Press.

Tharpe, Arthur (2013), "Top Concerns: Recent Ana Survey Findings (Marketing & Procurement)," in MASB Summer Summit. Boston: ANA.

Turnbull, Sarah and C. Wheeler (2015), "The Advertising Creative Process: A Study of Uk Agencies," *Journal of Marketing Communications* (In Press).

Ungar, Lyle, Barb Mellors, Ville Satoppa, Jon Baron, Phil Tetlock, Jaime Ramos, and Sam Swift (2012), "The Good Judgment Project: A Large Scale Test of Different Methods of Combining Expert Predictions," in AAAI Technical Report- Machine Aggregation of Human Judgment. Philadelphia: University of Pennsylvania.

Vakratsas, Demetrios and Tim Ambler (1999), "How Advertising Works: What Do We Really Know?," *Journal of Marketing*, 63 (1), 26-43.

Van Bruggen, Gerrit H., Ale Smidts, and Berend Wierenga (1998), "Improving Decision Making by Means of a Marketing Decision Support System," *Management Science*, 44 (5), 645-58.

Van Bruggen, Gerrit Harm and Berend Wierenga (2010), "Marketing Decision Making and Decision Support: Challenges and Perspectives for Successful Marketing Management Support Systems," *Foundations and Trends in Marketing*, 4 (4), 209-332.

Weiss, David J. and James Shanteau (2014), "Who's the Best? A Relativistic View of Expertise," *Applied Cognitive Psychology*, 28 (4), 447-57.

Wierenga, Berend (2011), "Managerial Decision Making in Marketing: The Next Research Frontier," *International Journal of Research in Marketing*, 28 (2), 89-101.

Wierenga, Berend and Gerrit Harm Van Bruggen (2000), *Marketing Management Support Systems: Principles, Tools, and Implementation*. Netherlands: Springer Science & Business Media.

Wood, Leslie (2009), "Short-Term Effects of Advertising: Some Well-Established Empirical Law-Like Patterns," *Journal of Advertising Research*, 49 (2), 186-92.

Table 1: Accuracy of intuitive decisions

Group	Subgroup	Sample size (n)	Predic-tions (n)	% Correct across individuals	% Correct predictions for each pair of ads (Modal decision is correct if result exceeds 50% for a pair)							
				All pairs	Pair 1	Pair 2	Pair 3	Pair 4	Pair 5	Pair 6	Pair 7	Pair 8
All	n/a	698	2159	51 (49-53)^	52 (46-58)	51 (45-57)	36 (30-42)	43 (37-49)	64 (58-70)	70 (64-75)	32 (27-38)	61 (55-66)
Employer	Manufacturer	616	1909	52 (49-54)	54 (48-60)	50 (44-56)	37 (31-43)	43 (37-50)	64 (58-70)	72 (66-78)	32 (26-38)	59 (53-65)
	Advertising agency	82	250	47 (41-53)	41 (24-59)	56 (38-74)	30 (11-48)	38 (20-55)	61 (43-78)	48 (27-69)	33 (17-50)	71 (54-88)
Role	Marketing operations	484	1507	51 (48-53)	53 (47-60)	48 (41-56)	33 (26-39)	45 (38-53)	64 (57-71)	72 (66-79)	31 (25-38)	57 (50-65)
	Marketing insights	62	189	61 (54-68)	45 (23-68)	66 (47-84)	63 (43-82)	33 (9-57)	70 (49-90)	82 (67-97)	42 (18-67)	70 (49-90)
	Other	70	213	50 (43-57)	64 (44-84)	43 (22-65)	41 (19-63)	37 (18-57)	61 (43-78)	56 (35-77)	31 (13-49)	62 (43-81)
Category*	Experience	n/a	744	56 (52-59)	54 (45-63)	50 (41-60)	30 (18-42)	34 (22-47)	70 (63-78)	72 (64-79)	38 (25-52)	64 (52-76)
	No experience	n/a	1415	49 (46-51)	51 (43-59)	51 (43-59)	38 (31-44)	45 (38-52)	57 (49-66)	67 (59-76)	31 (25-37)	60 (53-66)
Approve ads	Experience	334	1063	52 (49-55)	51 (43-59)	55 (46-64)	36 (27-44)	49 (40-57)	64 (55-72)	74 (67-82)	33 (24-41)	60 (52-69)
	No experience	282	846	51 (47-54)	58 (49-68)	44 (35-54)	39 (29-48)	36 (27-46)	65 (56-74)	69 (60-78)	32 (22-42)	58 (48-68)

*Managers could have experience in multiple categories and relevant category experience differed across pairs, which were from different categories

^Confidence intervals (at 95%)

Tables 2a and 2b: Logistic regression results[1]

	B	S.E.	df	p	Exp(B)	Lower CI	Upper CI
Table 2a: Model 1 (Manufacturer and Agency employees)							
EmployerManufacturer[2]	-.04	.15	1	.77	.96	.72	1.28
TenureMore	.19	.15	1	.19	1.21	.91	1.61
CategoryExperience	.15	.08	1	.06	1.16	1.00	1.36
EmployerManufacturer * TenureMore	-.19	.15	1	.19	.83	.62	1.10
EmployerManufacturer * Cat.Exp.	-.00	.07	1	.98	1.00	.87	1.15
TenureMore * Cat.Exp.	-.01	.05	1	.82	.99	.89	1.10
Table 2b: Model 2 (Manufacturer employees only)							
RoleInsights[3]	.288	.15	1	.05	1.33	1.00	1.79
RoleOperations[3]	.015	.10	1	.88	1.02	.84	1.23
ApproveAds	.147	.12	1	.21	1.16	.92	1.45
TenureMore	-.058	.09	1	.53	.94	.79	1.13
CategoryExperience	.210	.09	1	.02	1.23	1.03	1.47
RoleInsights * ApproveAds	.096	.14	1	.50	1.10	.84	1.45
RoleOperations * ApproveAds	.060	.10	1	.53	1.06	.88	1.28
RoleInsights * TenureMore	-.159	.11	1	.15	.85	.69	1.06
RoleOperations * TenureMore	-.137	.08	1	.10	.87	.74	1.03
RoleInsights * Cat.Exp.	.155	.11	1	.15	1.17	.95	1.44
RoleOperations * Cat.Exp.	.110	.08	1	.18	1.12	.95	1.31
TenureMore * ApproveAds	.034	.06	1	.57	1.04	.92	1.16
TenureMore * Cat.Exp.	-.020	.06	1	.71	.98	.88	1.09
ApproveAds * Cat.Exp.	.043	.06	1	.44	1.04	.94	1.16

[1]All variables were contrast coded; [2]Reference value is EmployerAgency; [3]Reference value is RoleOther