

# Do Emotional and Storytelling Boost Engagement? A Simple Decision Model Using Advertisements for the Super Bowl on YouTube

Minkyung Kim

*Berkeley College New York City, Midtown Campus, 3 E 43rd St, New York, NY 10017, United States*

## ABSTRACT

Storytelling and emotional appeals are often recommended advertising strategies. However, it remains unclear about whether they reliably improve audience response during high-attention events like the Super Bowl. This study used a simple but transparent decision rule to pick the creative style with the higher expected engagement per view on advertisements. This study particularly seeks to ask whether emotional and storytelling advertisements on the Super Bowl generate higher engagement on YouTube than non-emotional or product-first advertisements. This study hypothesizes that emotional and storytelling advertisements would generate more per-view outcomes on likes-per-views and comments-per-views. Assembling a publicly available dataset of 219 Super Bowl commercials and merging it with basic YouTube metadata (views, likes, and comments) with independent creative labels, this study normalized outcomes by views to compare advertisements. Compared group means were calculated by using Welch's unequal-variance t-test. Across both outcomes, differences between emotional vs. non-emotional and storytelling vs. product-first advertisements turned out to be small and not statistically significant (all  $p > 0.05$ ). Confidence intervals overlapped. Therefore, hypothesis in this study was not supported. In a highly optimized setting, high-level creative labels alone did not much increase observable social engagement.

**Keywords:** Super Bowl advertising; emotional appeals; storytelling; YouTube engagement; advertising effectiveness; decision-making model

## INTRODUCTION

Marketers usually face a practical question as to whether they need to build an advertisement around emotion or around story when their goal is to be

remembered by people and to spark observable audience engagement. Emotional appeals are often known to be ubiquitous and powerful means for advertisement. However, storytelling advertisements appear to capture attention and generate culturally vibrant atmosphere. This paper aims to propose a simple yet decision-oriented model that compares "emotional" versus "storytelling" advertisement based on two outcomes: consumer recall and engagement metrics (e.g., views, likes, and comments) that are important to both academic research and practice. The model

---

**Corresponding author:** Minkyung Kim, E-mail: maeveek@gmail.com.

**Copyright:** © 2025 Minkyung Kim. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Accepted** September 18, 2025

<https://doi.org/10.70251/HYJR2348.35376382>

was designed with publicly available secondary data of coded creative characteristics with platform analytics.

According to dual-process persuasion theory, advertisements are known to be processed by either elaborative routes or heuristic and affect-driven routes (1, 2). Emotional appeals often relate to the latter through the activation of the “affect heuristic” that rapid judgments are given by feelings (3). In addition, according to neuroscientific work, arousal may improve attention, while consolidating memory through amygdala-hippocampal mechanisms (7-9). Even if message details fade, positive affect linked to brands may persist (10). Even under low attention, strong emotional cues may stay effective (11). Meta-analyses showed that attention, attitudes, and recall were enhanced by humor (12). However, sexual appeals reliably attracted attention, while shaping evaluations but not always improving memory for brand information (13). In contrast, storytelling emphasizes narrative means to audiences to decrease counter-arguing, while strengthening the connections to brands and protagonists (4-6, 14). Stories tend to indirectly communicate with audiences through character arcs, visuals and episodes (15), improving the authenticity of a brand and attachment (16). Putting all of these together, emotional appeals tend to be effective for arousal and quick affective responses. On the other hand, storytelling emphasizes organized schemas that embed the meaning of a brand into memory.

Engagement metrics, such as views, watch time, likes, comments, and shares on digital video platforms play a role of behavioral traces of attention and involvement. In social media, cognitive, emotional, and behavioral investments in brand-related interactions are captured by consumer brand engagement (CBE) (17). Both message qualities and user motivation are reflected by the engagement according to the cross-platform works (18). High-arousal effect with emotions, including awe or anger, tends to increase sharing propensity (19). However, prior studies indicated that emotionality shapes diffusion in real time (20). For video advertisements, emotional responses of viewers predict attention and completion, connecting emotions to particular viewing behavior (21). Consequently, engagement metrics can be compared across creative types when exposure opportunities are similar.

Despite these theoretical advances, there are only a few studies that directly compared emotional and storytelling advertisements using large-scale and real-world digital engagement data. Most of the prior studies

focused on lab-based stimuli or self-reported outcomes from small-sized samples. This has limited replicability of results. Meta-analyses focused on pooling disparate categories of emotions, such as humor or sexual cues without analyzing narrative form, making it unclear whether effects were from emotionality or a structure of story (12, 13). On the other hand, research on social media diffusion indicated that sharing and commenting were predicted by emotions (19, 20). However, these studies rarely classified advertisements through narrative structure. Therefore, there remains a literature gap about whether non-narrative emotional appeals or storytelling better predict observable engagement traces, including comments or likes in naturalistic conditions.

The model proposed in this study applies creative approach as a key predictor with storytelling (narrative structure present) and non-narrative emotional appeals as two mutually exclusive categories. For recall, organization and distinctiveness are increased by storytelling through plot arcs and character goals, enhancing encoding and retrieval of memory (4-6). Non-narrative, emotional advertisements tend to rely on arousal-driven memories, making the recall depend on how much affect is associated with brand cues (8, 10), (11). For engagement, both storytelling and emotional advertisements would work. Storytelling may prompt comments and shares through identification (6), while emotional advertisements (humor or awe) promote quick likes and shares through arousal (19, 21). The model in this study acknowledged moderators, such as category, celebrity presence, or human/sex cues, that meta-analyses have proven to matter (12, 13).

The objective of this study is to compare storytelling and emotional appeals in Super Bowl advertisements by using engagement metrics in YouTube. This study specifically seeks to answer if storytelling advertisements outperform non-narrative, emotional advertisements on recall from consumers and digital engagement metrics when assessed with publicly available datasets. This study hypothesized that storytelling (narrative) advertisements appeal more positively to higher recall from consumers and also higher digital engagement metrics (views, likes, and comments) related to non-narrative, emotional appeals. This hypothesis aligned with transportation theory (4, 6) and also the evidence from structured narratives aiding consumer recall and also promoting social interaction (5). However, there existed a caveat that high-arousal emotional advertisements may exceed storytelling on immediate engagement when tested empirically (19, 21).

## METHODS AND MATERIALS

In this study, two publicly available datasets: a “Super Bowl Ads + Engagement” file containing per-video YouTube analytics (view counts, like counts, dislike counts, comment counts, and published titles, etc.) and “Creative Labels” with binary creative features (danger, animals, funny, celebrities, show\_product\_quickly, and etc.) were used. As archival, these records span various brands and years across Super Bowls. The advertisements span Super Bowls from 2010 to 2020. Political or issue-advocacy advertisements were excluded to particularly focus on commercial brands.

For the meaningful comparison across ads with different levels of exposure, raw accounts were normalized into rates. LikeRate was calculated as like counts/view counts. CommentRate was calculated as comment counts/view counts. Rows with missing values or zero view counts were excluded from the calculation of these rates. With common practice for proportion outcomes with different denominators, Welch’s unequal-variance t-test was used for the comparisons of group means to avoid assuming equal variances or balanced group sizes.

This study adopted two creative approaches. One of them is non-narrative emotional advertisement defined as an advertisement flagged true on at least one of funny, patriotic, danger, animals, and use\_sex. This identified high-arousal or affect-laden executions without considering plot structure. The other is narrative-based storytelling advertisement operationalized where show\_product\_quickly = FALSE as a widely used heuristic for “story first, product later.” These were contrasted with product-first advertisements (show\_product\_quickly = TRUE). As the labels came from the creative file, data classification was transparent and reproducible. The two contrasts tested were (a) emotional vs. non-emotional and (b) storytelling vs. product-first.

Because the creative labels were binary and subjective time to time, potential misclassification would be possible, such as partial narratives or borderline cases between humor and non-humor. These misclassifications may mask observed differences and should be taken into account when interpreting the results.

### Decision Model

A two-part model was proposed in this study by using public data. For engagement decision score, the following formulas were used. In these formulas,  $\mu_L$  and  $\mu_C$  indicate the sample means of LikeRate and CommentRate, and  $s_L$  and  $s_C$  indicate their

corresponding standard deviations.  $w_L$  and  $w_C$  show non-negative weights ( $w_L + w_C = 1$ ), allowing engagement utility to balance comments and likes. For example,  $w_L = 0.6$  and  $w_C = 0.2$  weights likes more heavily than comments.

$$zL(i) = \frac{L_i - \bar{L}}{s_L}, \quad zC(i) = \frac{C_i - \bar{C}}{s_C}$$

where  $\bar{L}$ ,  $s_L$ , and  $\bar{C}$ ,  $s_C$  are the sample mean and standard deviation.

Engagement utility for ad  $i$  was defined as follows.

$$U_i = w_L zL(i) + w_C zC(i)$$

Where  $w_L, w_C \geq 0$ ,  $w_L + w_C = 1$ . A neutral choice for both  $w_L$  and  $w_C$  was 0.5.

For each creative approach  $g \in \{\text{Story, Emotional}\}$ , the expected utility was estimated as follows.

$$\mu_g = \frac{1}{n_g} \sum_{i \in g} U_i$$

The engagement advantage of Story over Emotional was calculated as follows.

$$\Delta_U = \mu_{\text{Story}} - \mu_{\text{Emotional}}$$

A Welch 95% confidence interval was calculated for  $\Delta_U$  by using group means, variances, and sample sizes by applying a decision rule as follows. Storytelling was chosen if the lower bound of the 95% confidence interval for  $\Delta_U$  was greater than zero, and Emotional was chosen if it was less than zero. Otherwise, the approaches were treated as comparable on engagement, while making a decision based on the brand strategy, cost, or recall considerations.

Welch’s t was used for the primary tests to calculate the differences in group means of LikeRate and CommentRate. For each contrast, t, p, and group sizes ( $n_1, n_2$ ) were reported. With the heavy right-tail distribution as a typical engagement measures, it was confirmed that the conclusions drawn in this study were not driven by a few extreme denominators when expressing outcomes as rates instead of raw counts. There was no additional trimming or transformation applied in the data analysis.

## RESULTS

After merging the engagement and creative-label files and removing missing values in the rows, the

analytic sample turned out to vary slightly by outcome because of missing values of like counts or comment counts. For the contrast of emotional vs. non-emotional, sample sizes were n=193 vs. 26 for LikeRate, and 193 vs. 26 for CommentRate (Table 1). For the contrast of Storytelling vs. Product-first, the sample sizes were n=67 vs. 152 for LikeRate, and 67 vs. 152 for CommentRate (Table 2).

**Primary Contrasts**

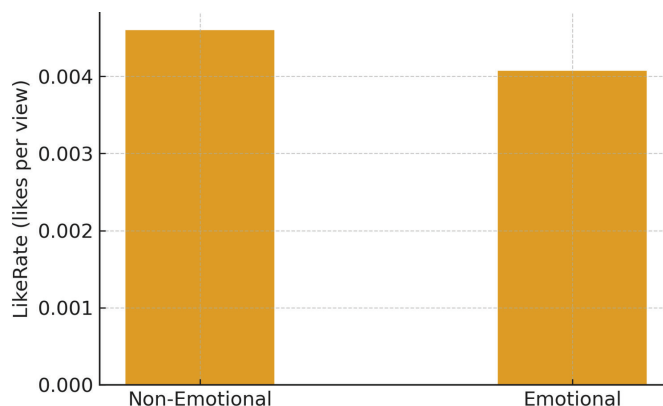
Emotional vs. Non-emotional

For LikeRate, Welch t was calculated to be -0.60 (p=0.552) (n=193 vs. 26) (Table 3). For CommentRate, Welch t was calculated to be -1.52 (p=0.137) (n=193 vs. 26). A statistically significant difference in per-view likes or per-view comments between advertisements tagged as emotional and those not tagged emotional was not detected in the analysis (Figure 1; Table 1). Within this dataset, a reliably higher normalized engagement was not shown from high-level emotional cues (funny, danger, patriotic, animals, sexual cues) when differences in exposure was considered (Figure 2; Table 1). Overall, these results suggest that emotional cues alone did not yield higher engagement per view.

Storytelling vs. Product-First

For LikeRate, Welch t score was calculated to be -0.78 (p=0.434) (n=67 vs. 152) (Figure 3; Table 2; Table

3). For CommentRate, Welch t was calculated to be -1.7718 (p=0.078) (n=67 vs. 152) (Table 3). There was not much difference between storytelling and product-first advertisements (Table 2). For CommentRate, the difference approached to conventional significance (two-sided p ≈0.078), and pointed slightly against storytelling. This indicated that product-first advertisements indicated a marginally higher per-view



**Figure 1.** Mean LikeRate (likes per view) for emotional versus non-emotional Super Bowl advertisements. Bar indicate group means. Error bars indicate 95% confidence intervals. Emotional advertisements were defined to contain humor, danger, animals, patriotic or sexual content.

**Table 1.** Summary of the Emotional vs. Non-Emotional Advertisements

Group	n	Mean LikeRate	SD LikeRate	95% CI (LikeRate)	Mean CommentRate	SD CommentRate	95% CI (CommentRate)
Non-Emotional Ads	26	0.0046	0.0042	0.0029 – 0.0063	0.0007	0.0010	0.0003 – 0.0011
Emotional Ads	193	0.0041	0.0039	0.0035 – 0.0046	0.0004	0.0012	0.0002 – 0.0005

The table reports the number of advertisements (n), group means and standard deviations for LikeRate (likes per view), and CommentRate (comments per view), as well as 95% confidence intervals. “Non-emotional” means advertisements without high-arousal cues (danger, animals, sexual, funny, patriotic), and “Emotional” indicates advertisements with at least one of such cues.

**Table 2.** Summary of the Storytelling vs. Product-First Advertisements

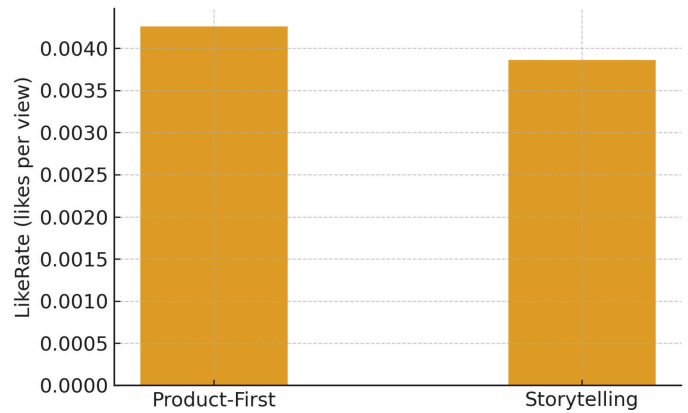
Group	n	Mean LikeRate	SD LikeRate	95% CI (LikeRate)	Mean CommentRate	SD CommentRate	95% CI (CommentRate)
Product-First Ads	152	0.0043	0.0043	0.0036 – 0.0049	0.0005	0.0014	0.0003 – 0.0007
Storytelling Ads	67	0.0039	0.0030	0.0031 – 0.0046	0.0003	0.0003	0.0002 – 0.0003

The table reports sample size (n), group means and standard deviations for LikeRate and CommentRate, as well as their 95% confidence intervals. “Storytelling” is applied as advertisements coded with short\_product\_quickly = FALSE (story first). “Product-First” indicates advertisements with show\_product\_quickly = TRUE.

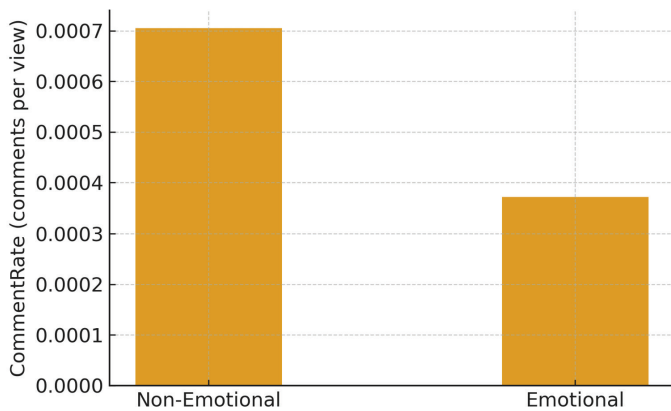
comment rates (Figure 4; Table 2). With the borderline p-value and unequal group sizes, this was interpreted as suggested but not definitive in this study (Table 3). Overall, these results suggest that storytelling did not outperform product-first advertisements. In addition, product-first advertisements even indicated a slight edge in comments.

**Decision-Model Implications**

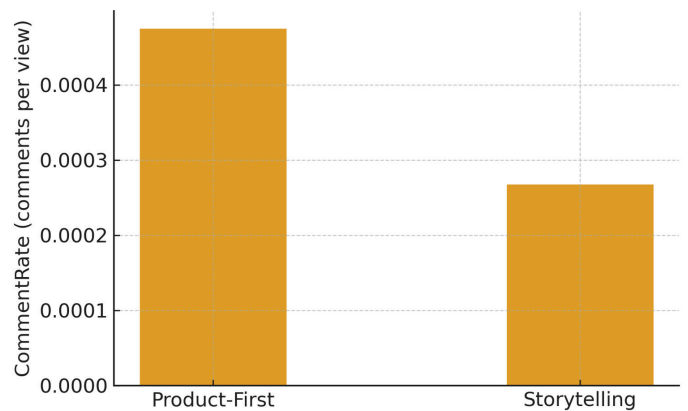
With the engagement-only model, the empirical evidence implied that the expected engagement utilities and were not statistically significant (Table 3). They were translated into the decision rule as follows. For typical weights, such as a neutral choice of 0.5, the 95% confidence interval around included zero. Then, this was interpreted to imply no clear engagement advantage for story over emotional. When a team valued comment more heavily, meaning when  $W_c$  was greater than



**Figure 3.** Mean LikeRate (likes per view) for storytelling versus product-first Super Bowl advertisements. Storytelling advertisements were defined to have delayed product appearance (`show_product_quickly = FALSE`). Bars indicate group means. Error bars indicate 95% confidence intervals.



**Figure 2.** Mean CommentRate (comments per view) for emotional versus non-emotional advertisements. Bars indicate group means. Error bars indicate 95% confidence intervals. CommentRate shows the number of comments in YouTube divided by a total number of views.



**Figure 4.** Mean CommentRate (comments per view) for storytelling versus product-first advertisements. Bars indicate group means. Error bars indicate 95% confidence intervals. CommentRate shows the number of comments in YouTube divided by a total number of views.

**Table 3.** Welch T-Test (by Outcome and Contrast)

Contrast	Outcome	n (Group 1)	n (Group 2)	t-statistic	p-value
Emotional vs. Non-Emotional	LikeRate	193	26	-0.60	0.552
Emotional vs. Non-Emotional	CommentRate	193	26	-1.52	0.137
Storytelling vs. Product-First	LikeRate	67	152	-0.78	0.434
Storytelling vs. Product-First	CommentRate	67	152	-1.77	0.078

This table reports tests comparing mean LikeRate and CommentRate between emotional and non-emotional advertisements and also between storytelling and product-first advertisements. Sample sizes, t-statistics, and two-sided p-values were reported for each test.

0.5), then, borderline finding with favor of product-first on CommentRate was interpreted to indicate extra caution before applying story-led lift in conversational engagement. Since the current calibration in this study did not have a recall proxy, the model focused only on engagement.

On average, emotional cues did not guarantee higher normalized engagement, and this aligned with how emotion grabbed attention but did not always turn into per-view interaction when scaled across many advertisements. In addition, there was no LikeRate difference shown between story-first vs. product-first, while only the borderline disadvantage was shown for Story on CommentRate. When the goal was just quick conversion or comments, a product-first structure may perform well or even slightly better. If the goal was brand building through storytelling, the storytelling route remained justified. However, engagement alone did not show a clear win without a recall proxy. In summary, the decision model showed no obvious engagement advantage for storytelling or emotional strategies, given the tested conditions.

## **DISCUSSION**

This study examined whether two frequently applied creative strategies: emotional appeals and storytelling narratives, were correlated with higher observable engagement for Super Bowl commercials on YouTube. With publicly available and transparent and per-view metrics, including likes per view and comments per view, this study compared emotional versus non-emotional advertisements and also storytelling versus product-first advertisements by using Welch's unequal variance tests.

With both outcomes, it turned out that the group difference was not statistically significant (all  $p > 0.05$ ). Confidence intervals overlapped, while point estimates suggested only small directional differences. Therefore, the findings in this study did not support the hypothesis in this study. Emotional or storytelling advertisements would not provide higher per-view engagement than their counterparts in the datasets used in this study. Therefore, this study confirmed that creative style did not predict engagement as an answer to the research question. Within this Super Bowl sample, likes or comments per view did not increase only by creative style.

The findings in this study are theoretically plausible. As a uniquely high-salience environment, Super Bowl

is exposed to massive reach, strong production values, and also celebrity assets. Under these circumstances, incremental variance attributable to a single creative dimension of either emotion-based or just product-first advertisement would be small relative to other forces. For example, brand values or distribution choices, such as upload timing or cross-platform amplification, would dominate observable engagement. More importantly, per-view likes and comments are regarded as downstream social signals. Because they are public, comparable, and reproducible, these social signals are valuable. However, they capture only a part of response mechanisms. Even when emotion and narrative do not generate social interactions on YouTube, they may still reinforce brand attitude or memory.

The Super Bowl is not a typical media context but more of an extreme example in terms of extreme budgets, cultural attention, and expectations. Creative effects that may be meaningful in ordinary campaigns may be amplified or distorted in the event of the Super Bowl. Therefore, these findings shall not be merely generalized to other digital video advertisements, but care needs to be taken.

## **Limitations**

While this study provides an important insight about emotional appeals vs. storytelling advertisements of a huge public event, there are limitations. This study used coarse creative labeling. This study used binary tags for "emotional" and "storytelling" that collapsed diverse subtypes, such as humor vs. warmth, full narrative vs. vignette montage, etc. Therefore, misclassification seemed to be an issue. Second, confounding factors, such as channel subscriptions, product category, upload timing, or advertisement length, were not considered when modeling, but they may still account for substantial engagement variance. Third, the design was observational. The results may not be causal as creative choices may associate with budget or brand strength. Lastly, the scope of outcome was narrow. Only social interaction was identified with likes-per-view or comments-per-view without considering memory, persuasion, or share rate.

## **CONCLUSION**

Using publicly available and transparent per-view engagement metrics from YouTube, no reliable evidence was found that emotional or storytelling advertisements on the Super Bowl outperform non-emotional or

product-first advertisements. Group differences in likes- or comments-per-view turned out to be small and not statistically significant. Therefore, the hypothesis in this study was not supported. This suggests that there would be other aspects, such as brand fit, or distribution strategy, matter as much as high-level creative labels that have already been optimized. With limitations in this study, it is suggested for future study to apply multilevel models applied with brand or year random effects and category or timing covariates. In addition, it is suggested to use finer creative coding, such as narrative structure or brand integration depth. In addition, it is also recommended for future study to analyze broader outcomes by including watch time, share rate, or branded search lift.

### CONFLICT OF INTEREST

The author declares no conflicts of interest related to this work.

### REFERENCES

- Petty RE, Cacioppo JT. Communication and persuasion: Central and peripheral routes to attitude change. *New York: Springer-Verlag*; 1986. ISBN: 9780387962509.
- Kahneman D. Thinking, fast and slow. *New York: Farrar, Straus and Giroux*; 2011. ISBN: 9780374275631.
- Slovic P, Finucane M, Peters E, MacGregor DG. The affect heuristic. In: Gilovich T, Griffin D, Kahneman D, editors. *Heuristics and biases: The psychology of intuitive judgment*. Cambridge: Cambridge University Press; 2002. p. 397-420. <https://doi.org/10.1017/CBO9780511808098.025>
- Green MC, Brock TC. The role of transportation in the persuasiveness of public narratives. *J Pers Soc Psychol*. 2000; 79 (5): 701-21. <https://doi.org/10.1037/0022-3514.79.5.701>
- Escalas JE. Narrative processing: Building consumer connections to brands. *J Consum Psychol*. 2004; 14 (1-2): 168-80. [https://doi.org/10.1207/s15327663jcp1401&2\\_19](https://doi.org/10.1207/s15327663jcp1401&2_19)
- Van Laer T, De Ruyter K, Visconti LM, Wetzels M. The extended transportation-imagery model: A meta-analysis. *J Consum Res*. 2014; 40 (5): 797-817. <https://doi.org/10.1086/673383>
- Vakratsas D, Ambler T. How advertising works: What do we really know? *J Mark*. 1999; 63 (1): 26-43. <https://doi.org/10.1177/002224299906300103>
- LaBar KS, Cabeza R. Cognitive neuroscience of emotional memory. *Trends Cogn Sci*. 2006; 10 (7): 295-305. <https://doi.org/10.1038/nrn1825>
- Kensinger EA. Remembering the details: Effects of emotion. *Mem Cognit*. 2009; 37 (2): 176-85. <https://doi.org/10.1177/1754073908100432>
- Ambler T, Burne T. The impact of affect on memory of advertising. *J Advert Res*. 1999; 39 (2): 25-34. <https://doi.org/10.1080/00218499.1999.12466598>
- Heath R. Reconsidering the role of attention in advertising effects: A cognitive neuroscience perspective. *Int J Mark Res*. 2006; 48 (4): 369-89.
- Eisend M. A meta-analysis of humor in advertising. *J Acad Mark Sci*. 2009; 37 (2): 191-203. <https://doi.org/10.1007/s11747-008-0096-y>
- Wirtz J, Sparks B, Zimbres TM. The effects of sexual appeals in advertising: A meta-analytic review. *Int J Advert*. 2018; 37 (2): 168-98. <https://doi.org/10.1080/02650487.2017.1334996>
- Escalas JE. Self-referencing and persuasion: Narrative transportation versus analytical elaboration. *J Consum Res*. 2007; 33 (4): 421-9. <https://doi.org/10.1086/510216>
- McQuarrie EF, Mick DG. Visual and verbal rhetorical figures under directed processing. *J Consum Res*. 2003; 29 (4): 579-87. <https://doi.org/10.1086/346252>
- Hamby A, Jacobson RP, Immerwahr G. It's about our values: How founder's stories influence brand authenticity. *J Consum Behav*. 2019; 18 (6): 492-503. <https://doi.org/10.1002/mar.21252>
- Hollebeek LD, Clark MK, Macky K. Consumer brand engagement in social media: Conceptualization, scale development and validation. *J Interact Mark*. 2014; 28 (2): 149-65. <https://doi.org/10.1016/j.intmar.2013.12.002>
- Voorveld HAM, van Noort G, Muntinga DG, Bronner F. Engagement with social media and social media advertising: The differentiating role of platform type. *J Advert*. 2018; 47 (1): 38-54. <https://doi.org/10.1080/00913367.2017.1405754>
- Berger J, Milkman KL. What makes online content viral? *J Mark Res*. 2012; 49 (2): 192-205. <https://doi.org/10.1509/jmr.10.0353>
- Stieglitz S, Dang-Xuan L. Emotions and information diffusion in social media-Sentiment of microblogs and sharing behavior. *Journal of Management Information Systems*. 2013; 29 (4): 217-47. <https://doi.org/10.2753/MIS0742-1222290408>
- Teixeira T, Wedel M, Pieters R. Emotion-induced engagement in Internet video advertisements. *J Mark Res*. 2012; 49 (2): 144-59. <https://doi.org/10.1509/jmr.10.0207>