

Standing Out While Fitting In: Visual Design of Text Overlays in Social Media Communication

Journal of Marketing

1-20

© American Marketing Association 2025

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/00222429251322773

journals.sagepub.com/home/jmx



Stefania Farace , Francisco Villarroel Ordenes, Dennis Herhausen ,
Dhruv Grewal , and Ko de Ruyter

Abstract

The vast amount of content on social media platforms makes it challenging for firms to create posts that get noticed. An increasingly popular approach to increase customer engagement relies on text overlays, where text is placed directly on images. Such practices raise questions of how to balance the visual and text elements for the best impact. Three key factors, commonly used by practitioners, can trigger engagement because of their visual salience: (1) the degree of dynamism or implied motion in the image as well as the (2) size and (3) centrality of the text overlay. Using multiple methods, including field studies, online experiments, and managerial interviews, the authors establish that a text overlay that is too large and centrally placed, combined with a dynamic image, has negative effects on consumer engagement, because these design combinations make the post visually unappealing. The authors leverage these findings to develop an interactive app that can help managers compose more engaging multimodal social media posts.

Keywords

salience, text overlays, dynamic images, visual appeal, social media engagement

Online supplement: <https://doi.org/10.1177/00222429251322773>

Submitted March 27, 2021

The massive preponderance of visual communication through social media—some estimates indicate that users upload an astounding 2 billion visual images every day (Sutarwala 2024)—creates challenges for both consumers and social media managers. Consumers must scroll quickly through the massive amount of information available (Wooley and Sharif 2022), and social media managers must find ways to make their content stand out and attract consumers' attention (Smith and Fischer 2021). Furthermore, reflecting shrinking attention spans overall, consumers increasingly engage with social media content through automatic, visually driven practices (Chan, Chen, and Leung 2023) rather than in-depth text analyses. Such trends suggest the need for reconsiderations of how to include text information in social media without disrupting users' visual experience.

For example, text overlays (TOs) are words or phrases (beyond a brand trademark or logo) that content providers insert deliberately within a delimited area of an image in a social media post (The Content Authority 2023; Velarde 2017). Integrating TOs

within the visual space of the image ensures that consumers see the textual message, even if they only glance at the post. However, TOs can be embedded in myriad ways and in widely divergent images, which raises additional questions regarding when and how to use text information and TOs to increase consumer engagement rather than hindering the visual experience.

Existing evidence pertaining to related combinations of TOs and images tends to be anecdotal, ad hoc, and conflicting. For example, in an extended A/B test, Instagram images without TOs outperformed those with TOs (Ayres 2020). In practice, Facebook and X (formerly Twitter) offer inconsistent suggestions:

Stefania Farace is Assistant Professor of Marketing, University of Bologna, Italy (email: stefania.farace@unibo.it). Francisco Villarroel Ordenes is Professor of Marketing, University of Bologna, Italy (email: francisco.villarroel@unibo.it). Dennis Herhausen is Associate Professor of Marketing, Vrije Universiteit Amsterdam, Netherlands (email: dennis.herhausen@vu.nl). Dhruv Grewal is Professor of Marketing, Babson College, USA (email: dgrewal@babson.edu). Ko de Ruyter is Professor of Marketing, King's Business School, UK (email: ko.deruyter@kcl.ac.uk).

The latter recommends TOs (Twitter 2022), whereas the former cautions that TOs should not occupy more than 20% of the image (Hesterberg 2021). However, we know of no theoretical or empirical marketing research that offers precise, well-founded guidelines for how marketers should combine images and TOs.

In line with Patrick and Hagtvædt (2011), we propose that people intuitively perceive certain visual designs as more aesthetically pleasing. In social media posts, poorly integrated TOs and images might create a sense of visual clutter (Chen et al. 2024; Pieters, Wedel, and Zhang 2007), whereas a well-structured visual arrangement, in which the distribution of visual elements features appropriate levels of salience, is essential to making designs “feel right” (Locher 2003; Locher, Stappers, and Overbeeke 1999). Certain visual characteristics also may be more or less effective for capturing attention and engaging audiences. To help social media managers make strategic decisions about how to present their products or brands (e.g., static portrayal vs. product in use), what size of the text to use, and where to place the text in the image, we examine the effects of three fundamental, salience-inducing characteristics—(1) the degree of dynamism or implied motion in images (Cian, Krishna, and Elder 2014; Hubbard 1995) and the (2) size and (3) centrality of the TO (Pieters and Wedel 2007)—on consumer responses. By combining these elements, we demonstrate that consumers engage less with posts that pair dynamic images with a large, centrally placed TO, because they find such posts visually unappealing.

Using a combination of multiple methods, including field studies, online experiments, and managerial interviews, we establish theoretical contributions for research pertaining to social media engagement, visual saliency, and visual aesthetics. First, we reveal how the composition and interplay of image type and TO affect engagement. A few existing studies consider image–text effects, but only in relation to captions, which fall outside the visual frame of the image (e.g., Li and Xie 2020; Villarroel Ordenes et al. 2019). As we clarify, applying ad hoc design principles can optimize designs in which text and images share the same visual space. In contrast, suboptimal combinations, especially of equally salient images and TOs, hinder positive perceptions of the visual design.

Second, Pieters and Wedel (2004, 2007) address salience in print advertising, but we know of no comprehensive investigations of visual salience in social media. In conventional marketing contexts, such as advertising, packaging, and shelving, experimental and eye-tracking studies establish how the salience of images can affect consumers’ attitudes, purchase intentions, and behaviors (Van der Lans, Pieters, and Wedel 2021). Although such effects might arise in social media platforms too (Appel et al. 2020), empirical evidence is lacking. Because salient properties such as image dynamism, TO size, and TO centrality can trigger attention (Cian, Krishna, and Elder 2014; Fang et al. 2020), we explicitly establish their joint effects on consumer engagement through an effective distribution of salience across visual elements.

Third, we expand on marketing aesthetics research (Grinstein, Hagtvædt, and Kronrod 2019; Patrick and Hagtvædt 2011) and

move beyond considerations of product design (Hagtvædt and Brasel 2017; Hagtvædt and Patrick 2008) and text-based logos (Hagtvædt 2011). In response to calls for research into multimodal aesthetics (Hagtvædt 2022), we determine that appropriate combinations of images and TOs can enhance the visual appeal of a social media post. Poor combinations instead disrupt these perceptions, which then reduces consumer engagement.

For managers, this article thus offers practical recommendations for how to highlight visual elements in social media posts to ensure that they stand out but do not clash. To support design optimization, we introduce an interactive app based on a vision transformer model (Dosovitskiy et al. 2020) that (1) assesses whether an image is dynamic or static, (2) forecasts consumer engagement based on the image and TO characteristics, and (3) recommends the ideal TO size and centrality for each image. With this automated tool, managers can assess an image’s dynamism, before pairing it with a TO of the right size and position, to boost engagement.

Theoretical Background

In the conceptual framework in Figure 1, we specify how TOs and images likely interact to affect consumer engagement. Building on research in social media engagement (Li and Xie 2020), image–text salience (Pieters and Wedel 2004), and visual aesthetics (Patrick and Hagtvædt 2011; Berlyne 1974), we explore how to arrange images and TOs in social media posts. We define optimal arrangements as those that account for the combined salience of images and TOs, so that the resulting social media posts are visually appealing and attractive, due to their effective, formal execution.

What Drives Social Media Engagement?

For our study context, engagement refers to interactions between users and firm-generated content on social media platforms. It can manifest as various cognitive and behavioral responses that signal users’ attention, interest, and involvement with the shared content (Van Doorn et al. 2010). Two popular measures of engagement on social media reflect (1) a behavioral dimension, which includes actions such as likes, comments, shares, and clicks, and (2) a cognitive dimension, which reflects users’ intentions to like or share branded content (e.g., Cascio Rizzo et al. 2024; Wies, Bleier, and Edeling 2023). In addition, we propose that social media posts often include three elements that may affect engagement: the image (pictorial element), the caption, and the TO (textual elements). As the examples in Figure 2 indicate, TOs can be integral to the visual properties of the image and thus likely affect its visual appeal and consumer engagement.

Visual factors that tend to influence engagement include visual concealment (Sevilla and Meyer 2020), the percentage of dominant colors (Li and Xie 2020), image complexity (Overgoor, Rand, and Mazloom 2022), and the presence of humans (Hartmann et al. 2021). When studies focus on visual factors, they often control for the textual properties of the caption, such as its valence (Schweidel and Moe 2014), arousal

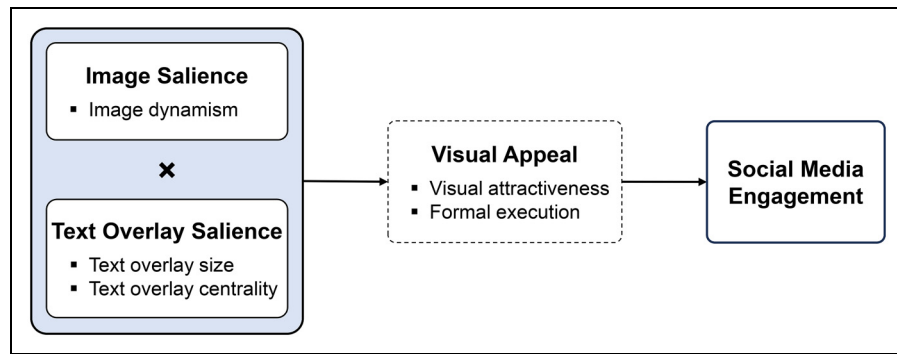


Figure 1. Conceptual Framework.

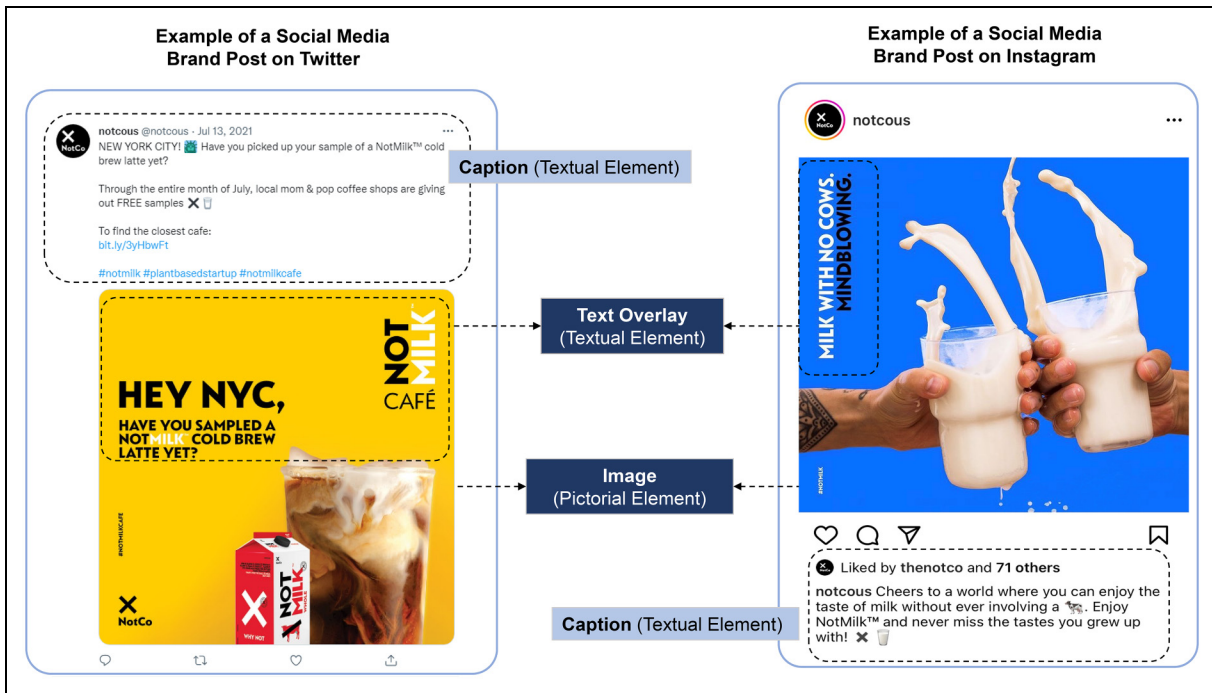


Figure 2. Examples of Social Media Post with Text Overlays.

(Kanuri, Chen, and Sridhar 2018), readability (Pancer et al. 2019), and other semantic constructs (Berger et al. 2020). If researchers instead consider both visual and textual factors, they generally refer to the semantic fit between the image and caption (Li and Xie 2020; Van Rompay, De Vries, and Van Venrooij 2010; Villarroel Ordenes et al. 2019). However, captions appear in a separate space in social media posts; they do not visually overlay the image. Thus, findings pertaining to captions might not generalize to TOs. Instead, text superposed over the image requires considerations of the interplay of these elements from a perceptual, stylistic perspective. As we have implied, if the combination (image overlaid with text) enhances the visual experience—that is, consumers' perception that the composition is visually appealing, well-structured, and intuitively right—it could increase consumer engagement (Affonso and Janiszewski 2023; Locher, Smith, and Smith 2001).

Visual Appeal of Image–TO Combinations

Viewers tend to form positive impressions of effectively arranged visual elements (Locher, Stappers, and Overbeeke 1999). However, viewers' experience can be disrupted by too many focal points or competing elements, which make compositions seem visually cluttered (Pieters, Wedel, and Zhang 2007). Visual clutter occurs if excess or disorganized elements interfere with one another and cause crowding (Balas, Nakano, and Rosenholtz 2009), masking (Breitmeyer 2007), or reduced recognition due to occlusion (Chun and Wolfe 2005). In social media contexts, these consequences may be amplified, because consumers rapidly scroll through feeds and form overall, often fleeting visual impressions (i.e., “doom scrolling”; Salamon 2024). Social media users also tend to process visual designs initially on a perceptual level (Chaiken and Trope 1999; Treisman and Gelade 1980). During an initial, preattentive

stage, human brains form immediate, overall impressions of a social media post's visual structure (Affonso and Janiszewski 2023). If this early impression aligns with viewers' intuitive sense of aesthetically pleasing visuals, they are more likely to engage further with the social media post on a deliberate, cognitive level (Locher, Stappers, and Overbeeke 1999).

Visual appeal also depends on the sizes, shapes, and positions of visual elements, which signal their salience (Arnheim 1954; Berlyne 1974; Hekkert 2006). According to prior research on multiformat communication, when visual and text cues have overlapping functions (i.e., both attract attention), they increase communication costs and require viewers to invest more time, effort, and resources to process the message (Moffett, Garretson Folse, and Palmatier 2021). In this sense, social media posts that combine highly prominent visual and textual cues may be aesthetically unappealing (Pieters and Wedel 2007). In support of this notion, Affonso and Janiszewski (2023) show empirically that visual designs (e.g., font selection, image size, content organization) generate perceptions of (un)structured compositions that influence consumers' responses, moderated by the type of market positioning (utilitarian vs. hedonic). Similarly, Pieters, Wedel, and Zhang (2007) identify the different effects of the size of brand elements, pictorials, text, price information, and promotions in feature advertising on attention selection and attention engagement. Specifically, the size of the pictorial has the strongest effect, followed by price and promotion; the size of the text has the smallest (insignificant) effect on both attention selection (first impressions) and attention engagement (deeper information processing). We consider whether similar effects arise in fast-paced social media contexts, where viewers spend less time analyzing the content than they do in traditional print advertising studies and where the text might overlay the images.

According to Rosenholtz and Jin's (2005) saliency model, human visual systems prioritize "unusual" items that stand out due to contrast, color, orientation, or motion (Wolfe 1998). A composition's effectiveness also depends on its organization; poorly arranged elements can lead to visually cluttered scenes that fail to attract attention. In social media, we predict that balancing the salience of different visual elements is key to maintaining a composition's visual appeal. When a social media post features too many highly salient elements, its visual appeal may diminish, because it overwhelms viewers with multiple, competing focal points (Pieters, Wedel, and Zhang 2007). A positive visual experience instead should arise when consumers intuitively believe the post "feels right," because its elements (and their salience) are effectively distributed. Based on this reasoning, extant research, and 24 interviews with industry experts (Web Appendix A), we derive formal hypotheses next.

Hypotheses

We focus on essential visual properties of the image and TO in our effort to understand their optimal combination in a visually appealing composition (Arnheim 1954; Gershoni and Hochstein 2011) that increases consumer engagement with the

social media post. One such property is image *dynamism*, defined as a portrayal of action, motion, or movement (Cian, Krishna, and Elder 2014; Kress and Van Leeuwen 2006). Dynamism influences a range of attitudinal (Cian, Krishna, and Elder 2014) and behavioral (e.g., Milosavljevic et al. 2012; Poor, Duhachek, and Krishnan 2013) outcomes. When an image is dynamic, viewers perceive a sense of movement (Cian, Krishna, and Elder 2014) that tends to draw their attention, because humans are evolutionarily hardwired to detect and anticipate movement (Henderson, Weeks, and Hollingworth 1999). Image dynamism also goes beyond the simple presence or absence of a visual feature (e.g., color, presence of humans, facial expressions), in that it requires actions performed by entities that engender a sense of movement (Kress and Van Leeuwen 2006). Even still images can evoke dynamism, assuming they portray visual elements that trigger perceived motion (Cian, Krishna, and Elder 2014). Visual perception research in other contexts establishes dynamism as an effective, salient feature that attracts attention (Sample, Hagtvedt, and Brasel 2020). For example, greater dynamism in warning signs results in earlier attention and safer driving (Cian, Krishna, and Elder 2015).

With regard to the salience of the TO, we consider two visual features: its *size*, or the space occupied by the TO in the image region (Pieters and Wedel 2004; Pieters, Wedel, and Batra 2010), and its *centrality*, reflecting the location of the TO in the image (ranging from peripheral to central) (Sample, Hagtvedt, and Brasel 2020; Wilson and Chatterjee 2005). In addition to their impacts on attentional processes (Pieters and Wedel 2007), TO size and centrality influence consumers' visual perceptions (Sample, Hagtvedt, and Brasel 2020). We predict that both features combine with salient dynamic images to exert relevant effects and thus offer independent hypotheses about the interplay of (1) dynamic images and TO size and (2) dynamic images and TO centrality. In managerial interviews, social media managers affirmed the relevance of image dynamism and TO size and centrality¹ (Web Appendix A), which provides an ecologically valid foundation (Blanchard et al. 2022) for our analyses of effective multimodal social media communication.

Interplay of TO Size and Image Dynamism

Generally, the larger the text in print ads, the more attention it draws (Pieters and Wedel 2004). However, a social media post that combines a large TO with a highly dynamic image could feel intuitively "wrong." By showing action rather than still objects, dynamic images inherently demand more attention,

¹ When asked whether they prefer posting static or dynamic images, most informants selected dynamic images (63%, both = 12%, static = 25%), indicating that they "grab more attention," "appeal to the imagination," and "are engaging." However, these preferences are not reflected in the images actually used by companies in Studies 1 and 5, which use 64% and 82% static images, respectively. Furthermore, 83% of these managers confirmed that including TOs is a relevant practice (average score of 6.29 on a scale from 1 = "not important at all for our company" to 7 = "very important for our company"). The managers also provided open answers about their experiences with TOs and how they approach their design (size and centrality), which we summarize in Web Appendix A.

so adding another prominent element, like a large TO, may disrupt the visual experience and lessen the post's effectiveness (Pieters, Wedel, and Zhang 2007). Instead, an appealing composition likely requires a visual balance between the salience of the dynamic image and the size of the TO. Such balance can better match consumers' implicit beliefs about visually pleasing depictions (Hagtvedt, Patrick, and Hagtvedt 2008).

Prior research indicates a curvilinear relationship between stimulus complexity and liking (Overgoor, Rand, and Mazloom 2022), so in turn, we propose that larger TO sizes enhance the effectiveness of dynamic images up to a certain threshold. Beyond this point, excessively large TOs detract from engagement by diverting attention away from the salient image. As the TO size continues to increase, the impact of dynamic images diminishes further (Patrick and Hagtvedt 2011). However, because static images are less salient than dynamic ones (Cian, Krishna, and Elder 2014), increasing the size of the TO over static images should have a linear, positive effect on consumer engagement. A large TO complements the static image and improves its perceived visual appeal by taking the primary focus in the post (Hagtvedt, Patrick, and Hagtvedt 2008). As traditional advertising research shows, increasing the size of the text in static magazine ads does not detract attention from the picture (Pieters and Wedel 2004). Taken together:

H_{1a}: For dynamic images, TO size has an inverted U-shaped effect on consumer engagement, such that a TO occupying greater areas in the image increases engagement up to an optimal point, after which it decreases consumer engagement.

H_{1b}: For static images, TO size has a positive effect on consumer engagement, such that increasing the area occupied by a TO in the image increases consumer engagement.

Interplay of TO Centrality and Image Dynamism

Within any image, TOs could appear in several places, which also might affect consumers' visual perceptions of the social media post. We posit that the location of a salient TO in a dynamic (vs. static) image might determine viewers' engagement (Wilson and Chatterjee 2005). As scene viewing literature documents, people exhibit a central bias in their eye-fixation distributions (Tatler 2007); they look more frequently to the central, rather than peripheral, area of a visual scene (Atalay, Bodur, and Rasolofoarison 2012). Thus, the most optimal combination with salient dynamic images might require a noncentral TO rather than a (more salient) centrally placed TO. A TO placed in the center is likely to compete with a salient, dynamic image and decrease visual appeal, which may be detrimental to consumer engagement.

However, combining dynamic images with a peripheral TO might generate a greater visual appeal, due to this appealing good composition, and increase consumer engagement. According to visual communication studies, locating text at anchoring (cf. central) points in a salient image, such as at intersections created by a rule-of-thirds grid (Firoze et al. 2018; Zhang et al. 2022), can significantly increase message effectiveness

(Kosslyn 2006). The positive effect of shifting text away from the primary image also resonates with research that details the need for sufficient white space around visuals, to ensure that each element stands out clearly (Kwan, Dai, and Wyer 2017; Pracejus, Olsen, and O'Guinn 2006). Conversely, we expect a positive engagement effect when TOs appear centrally in static images (Gardner et al. 2008), which are less salient than dynamic images (Cian, Krishna, and Elder 2014). The salient features that arise due to the central positioning of the TO can complement the static image, rather than clashing with it. This combination should enhance visual appeal of such a social media post and exert a positive effect on engagement. Formally:

H_{2a}: For dynamic images, TO centrality has a negative effect on consumer engagement, such that positioning the TO less centrally increases consumer engagement.

H_{2b}: For static images, TO centrality has a positive effect on consumer engagement, such that positioning the TO more centrally increases consumer engagement.

We test these hypotheses in a field study conducted on Twitter (Study 1) and in experimental studies (Study 2a and 2b). We then present field studies conducted on Facebook (Study 3a and 3b) and experimental studies in which we test the mediating effect of visual appeal (Study 4a and 4b). With another field study (Study 5), conducted on Instagram, we further generalize our findings and develop an automated, practical image classifier tool.

Study 1: Field Study on Twitter²

To start addressing knowledge gaps and the managerial challenges, we examine the effects of multimodal posts collected from Twitter, which hosts 91% of *Fortune* 500 companies (Barnes, Kane, and Maloney 2018) and has substantially transitioned its focus from text to images³ (Kane and Pear 2016; Sheikh 2024). We focused on consumer brands, whose communication, goals, channels, appeals, and measures of success differ from those for messages targeting business clients (Agnihotri et al. 2016). To increase generalizability, the sample covers several industries and both products and services, such as food, manufacturing, retailing, and hospitality. For each sector, we include an industry leader as a representative brand, except for retail, for which retailers' widespread presence in social media requires consideration of multiple subcategories. By investigating the industry leader, we gain insights into a social media strategy that is broadly accepted by consumers and potentially copied by other brands; in contrast, a small brand might opt for an unconventional strategy to garner attention, but such an approach would likely be inappropriate for most firms. Our final selection includes brands from the food (Standard Industrial Classification [SIC] 2000), manufacturing (SIC 3000), retailing trade

² This study was conducted before Twitter's rebrand to X, so we refer to it as "Twitter" throughout.

³ Approximately 97% of users focus on visuals on Twitter (Twitter 2022).

(SIC 5200), and hospitality (SIC 7000) sectors and excludes industries such as mining (SIC 1000) or health care services (SIC 8000). We then used SocialBakers.com to identify brands and industries with a strong social media presence on both Twitter and Instagram, set up specifically to engage end consumers. We chose the following brands: Coca-Cola, Ford, Walmart, Tesco, McDonald's, Amazon, Disney, and Nike. We obtained the data with the selenium package from Python. To ensure the validity of the data collection (Boegershausen et al. 2022), we manually cross-checked a random subsample of 500 tweets and found 100% matching. The collected data set includes 8,275 tweets, posted between May 1, 2016, and May 1, 2017. We exclude text-only tweets (1,125) and tweets with videos (949). We also excluded 602 tweets with more than one image, which resulted in a final sample of 5,589 tweets, each with a single image.

Measures

In line with previous studies, we measure engagement by using the number of retweets (Gavilanes, Flatten, and Brettel 2018; Gong et al. 2017). Human raters annotated the dynamism of the images on a seven-point scale ranging from 1 = "very static" to 7 = "very dynamic" ($\alpha = .91$), TO size as a percentage (1% to 100%; $\alpha = .76$), and centrality on a scale from 1 = "not central at all" to 7 = "very central" ($\alpha = .81$; see Web Appendix B). Reflecting our definition of TO, we only investigate images with one overlay ($N = 615$), which represents a conservative approach. That is, images with two or more overlays might represent an infographic, with a mainly informational purpose. In addition, in the presence of two overlays, it would be difficult to determine which one is central, for our measure of TO centrality. To check the consistency of our findings, we tested a model that includes all TOs ($N = 976$), while controlling for the number of TOs, and find consistent results. Furthermore, we incorporate several covariates that might influence engagement, as detailed in Table 1.

The descriptive statistics and correlations are in Web Appendix B. The correlation of TO size and TO centrality ($r = .86$) implies that larger TOs typically are placed more centrally. In line with Marinova, Singh, and Singh's (2018) approach, we regressed TO centrality on TO size, then used the residual as an instrument for TO centrality in the empirical models. After applying this approach, the variance inflation factors (VIFs) all fall below 1.935. We adopt a negative binomial regression, given the distribution of the number of retweets, which shows an overdispersed count around the mean but no excess of zeros (i.e., less than 1%).

Results

Table 2 contains the results for Study 1. We report incident rate ratios (IRRs) to indicate the probability of engagement due to each variable ($IRR > 1$ indicates a positive effect and more engagement; $IRR < 1$ indicates a negative effect and less engagement). The effect of image dynamism is positive and significant. The IRR of 1.043 ($z = 5.70, p < .001$) indicates that a one-point increase in image dynamism (seven-point scale) results in about 4.3% more engagement. Panels A–F of Figure 3 display the three-dimensional

plots of the interactive effects of TO size and TO centrality, as well as the marginal effects of TO size and TO centrality for static and dynamic images.

The interaction of image dynamism and TO size exhibits an inverted U-shape, such that the simple interaction is positive ($IRR = 1.008, p < .001$), but the quadratic interaction is negative ($IRR = .999, p < .001$), as revealed in Figure 3, Panel A. For dynamic images, the simple slopes indicate a positive main effect ($dy/dx = 3.151, p < .001$) and a negative quadratic effect ($dy/dx = -.073, p < .001$). Figure 3, Panel B, displays the marginal effects of different TO sizes in dynamic images, relative to the marginal effect of dynamic images without TO. Initially, increasing TO size prompts more engagement with a dynamic image, but when the TO size exceeds 24% of the image, increasing it further weakens engagement. When the TO size is greater than 46% of the image, engagement falls lower than it would have been for a dynamic image without TO. For static images, the simple slopes show nonsignificant effects (main: $dy/dx = -.326, p = .18$; quadratic: $dy/dx = .002, p = .66$; Panel C); engagement does not change as the TO gets larger or smaller.

The interaction of image dynamism and TO centrality is negative and significant ($IRR = .929, p < .001$), as revealed in Figure 3, Panel D.⁴ For dynamic images, we find a negative effect of TO centrality: Increasing TO centrality reduces engagement ($dy/dx = -3.067, p < .001$), as revealed in Figure 3, Panel E. Engagement with a typical dynamic image decreases as a TO becomes more central. For static images, the positive effect in Panel F ($dy/dx = .640, p < .001$) indicates that engagement with a typical static image increases as the TO moves more centrally.

Robustness Tests

We report several robustness tests in Web Appendix B. The results do not change when we log-transform the dependent variable or measure engagement using the sum of favorites and comments. To account for potential selection effects on images with TO, we use propensity score matching (PSM) and find robust results with the matched sample. We find consistent results when we use a BERT embedding model (<https://huggingface.co/google-bert/bert-large-uncased>) to assess the meaning (embedding) similarity of the TO with the caption and the image too. As noted previously, the results remain consistent when we operationalize TO as images with one or more TOs, while controlling for the number of TO used. Finally, we test a model with three additional TO properties⁵: font type (serif vs. sans serif), TO alignment (horizontal vs. vertical), and TO color contrast (yes vs. no). Accounting for these properties does not change our results.

⁴ We consider a potential curvilinear interaction of image dynamism and TO centrality, but the effect is nonsignificant, confirming the negative interaction effect.

⁵ We thank an anonymous reviewer for suggesting these properties. Two independent coders annotated each image with a TO and achieved intercoder agreements of .82 (font), .79 (alignment), and .85 (color contrast); all disagreements were resolved through discussion.

Table 1. Measurement in the Field Studies.

Variable	Operationalization	Source	Related Studies
Dependent Variable			
Engagement	- Study 1: Number of retweets, sum of favorites and comments as robustness test - Study 5: Sum of likes and comments	Twitter API, Instagram API	Gong et al. (2017); Herhausen et al. (2019)
Independent Variables			
Image dynamism	- Study 1: Human raters annotated image dynamism on scale ranging from 1 = “very static” to 7 = “very dynamic” - Study 5: Automated classifier for image dynamism (dynamism probability)	Human annotation, deep learning	Hartmann et al. (2021)
TO size	Percentage of overlay relative to the image	Human annotation	—
TO centrality	Extent to which the overlay is placed in the center of the image, where 1 = “not central at all,” and 7 = “very central”	Human annotation	—
Image Controls			
Human presence	- Study 1: Human raters annotated presence of human (1/0) - Study 5: Google Cloud Vision annotation (1/0)	Human annotation, Google Cloud Vision	Li and Xie (2020)
Color concentration	Sum of pixel percentage of top-three colors	Google Cloud Vision	Li and Xie (2020)
Color saturation	Degree of saturation of the dominant color (0 to 100%)	Python script	Li and Xie (2020)
Duplicate image	Image that has been used in a preceding post (1/0)	Python script	Hartmann et al. (2021)
TO Controls			
TO–caption similarity	Jaccard similarity score between TO and caption	Text mining	Li and Xie (2020)
Price	Whether the TO includes price information (1/0)	Human annotation	Jalali and Papatla (2019)
Product	Whether the TO depicts the name of a product (1/0)	Human annotation	Jalali and Papatla (2019)
TO–image similarity	Jaccard similarity between the content of the TO and the one of the image (extracted through Google Cloud Vision objects)	Google Cloud Vision, text mining	—
Caption Controls			
Length	Number of characters in the caption	Text mining	Herhausen et al. (2019)
Emphasis	Number of words in capital letters plus exclamation marks	Text mining	Luangrath, Peck, and Barger (2017)
Valence	Positive emotions minus negative emotions score	LIWC dictionary	Berger and Milkman (2012)
Arousal	Arousal scores from the dictionary of affect in language	DAL dictionary	Berger and Milkman (2012)
Sales promotion	Presence of sales promotion words in the caption (1/0)	Text mining	Jalali and Papatla (2019)
Complexity	Flesh Reading Ease Index (ranges from 0 to 100)	Text mining	Packard and Berger (2023)
Question	Presence of a question mark in the caption	Text mining	Villarroel Ordenes et al. (2019)
Emojis	Number of emojis in caption	Text mining	McShane et al. (2021)
Hashtags	Count of hashtags in caption	Text mining	Walker et al. (2017)
URL	Presence of URL in caption (1/0)	Text mining	Walker et al. (2017)
Mentions	Count of mentions in caption	Text mining	Walker et al. (2017)
Additional Controls			
Weekend	Whether the post appeared on the weekend (1/0)	Twitter API, Instagram API	Herhausen et al. (2019)
Carryover effect	Engagement with the previous post	Twitter API, Instagram API	Villarroel Ordenes et al. (2019)
Fixed Effects			
Brand	Dummy variable capturing the eight brands	Twitter API, Instagram API	Herhausen et al. (2019)
Hour of day	Dummy variable capturing the hour of day	Twitter API	Herhausen et al. (2019)

Notes: LIWC = Linguistic Inquiry and Word Count; DAL = Dictionary of Affect in Language.

Discussion

The Study 1 findings suggest that the effectiveness of a social media post depends on an appropriate combination of salient

features, as defined by image dynamism and TO size and centrality. Social media posts with dynamic images benefit most from TOs that are around 24% of the size of the image and

Table 2. Results of Study 1.

	Model 1			Model 2		
	IRR	Coef.	Z	IRR	Coef.	z
Independent Variables						
Image dynamism	1.051***	.050	6.97	1.043***	.042	5.70
TO size	1.008*	.008	2.29	.986*	-.014	-2.44
TO size ²	1.000***	.000	-3.57	1.000 [†]	.000	1.92
TO centrality	.985	-.015	-.79	1.196***	.179	5.24
Interactions						
Image dynamism × TO size				1.008***	.008	4.48
Image dynamism × TO size ²				.999***	-.000	-4.02
Image dynamism × TO centrality				.929***	-.074	-6.96
Image Controls						
Human presence	.731***	-.313	-9.82	.727***	-.319	-10.05
Color concentration	1.006***	.006	6.18	1.006***	.006	6.09
Color saturation	1.030	.029	.68	1.011	.011	.25
Duplicate image	.777***	-.252	-4.99	.777***	-.252	-5.02
TO Controls						
TO-caption similarity	.419***	-.870	-3.55	.415***	-.878	-3.59
Price	2.307***	.836	7.24	2.358***	.858	7.45
Product	2.162***	.771	5.96	2.257***	.814	6.31
TO-image similarity	1.328	.284	.13	1.677	.517	.23
Caption Controls						
Length	.997***	-.003	-5.51	.996***	-.004	-6.03
Emphasis	1.081***	.078	4.25	1.078***	.075	4.12
Valence	.996 [†]	-.004	-1.79	.995 [†]	-.005	-1.88
Arousal	.894***	-.113	-.85	.894***	-.112	-.85
Sales promotion	1.852***	.616	21.75	1.856***	.618	21.89
Complexity	1.005***	.005	5.99	1.005***	.005	6.00
Question	.737***	-.305	-7.44	.737***	-.305	-7.48
Emojis	1.104***	.099	4.92	1.108***	.102	5.11
Hashtags	1.130***	.122	4.14	1.152***	.141	4.80
URL	.712***	-.340	-7.22	.727***	-.318	-6.79
Mention	1.004***	.004	.12	1.004***	.004	.12
Additional Controls						
Weekend	.903***	-.102	-3.44	.905***	-.100	-3.38
Carryover effect	1.001***	.001	17.54	1.001***	.001	17.34
Fixed Effects						
Brand	Included			Included		
Hour of day	Included			Included		
Significance (Chi ²)	2,981.9***			3,037.9***		
Nagelkerke pseudo-R ²	.413			.419		
N (AIC)	5,581 (57,778)			5,581 (57,728)		

[†] $p < .10$.* $p < .05$.** $p < .01$.*** $p < .001$.

Notes: AIC = Akaike information criterion. An IRR greater than (less than) 1 indicates a positive (negative) effect. The IRRs for TO size are rescaled to 10 percentage points for ease of interpretation. Robustness tests are reported in Web Appendix B.

placed less centrally. Social media posts with static images instead benefit most from TOs that are placed centrally, but changes in their size do not influence their effectiveness. Next, we examine these effects in a more controlled environment.

Study 2: Online Experiments on Text Overlay Effects

Even with multiple control variables and PSM, we cannot rule out all potential sources of endogeneity with nonexperimental field

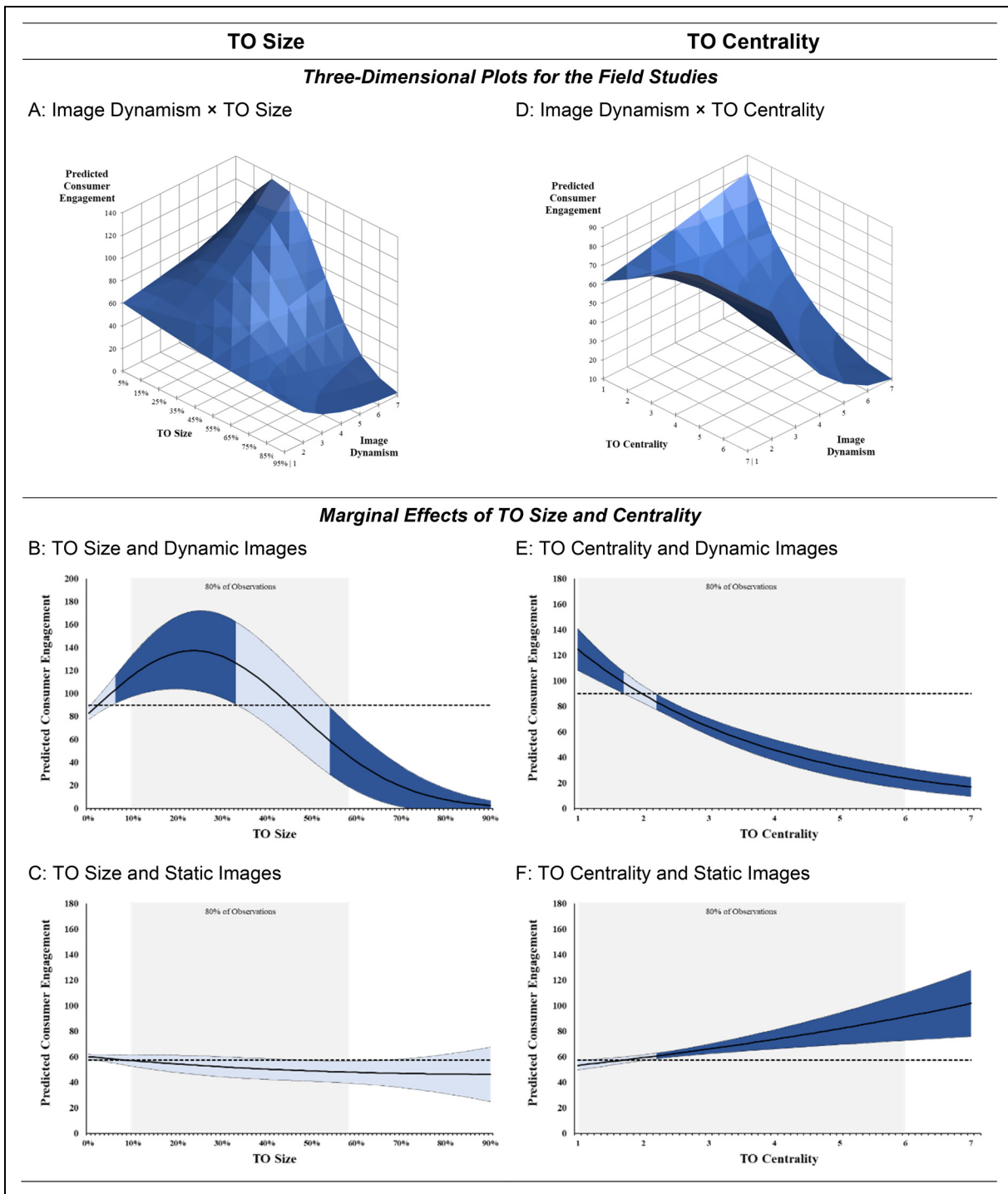


Figure 3. Effects of Text Overlay Size and Centrality in Study 1.

Notes: The three-dimensional plots display the full range of observations. Marginal effects with 95% confidence intervals. The area in gray is where 80% of observations are located. Dotted lines are the predicted means for dynamic (static) images without TO. The dark blue color indicates areas where a TO is beneficial (detrimental) for engagement.

data. Therefore, we conduct two experimental studies, in which we ask respondents to rate their engagement with real social media posts. Both studies were preregistered (Study 2a: <https://aspredicted.org/vj3m-xvzz.pdf>, Study 2b: <https://aspredicted.org/p2xz-hvm8.pdf>). We randomly selected 20 different social media posts (10 static, 10 dynamic) from the Study 1 data set. None of

these posts included TOs, so we added them, manipulating their size and centrality, such that we generated 120 different social media posts (see Web Appendix C for further details).

In both Studies 2a and 2b, each respondent assessed 10 randomly assigned social media posts. We use a mixed-design approach, in which the image (dynamic vs. static) is a

Table 3. Experimental Results Studies 2a and 2b.

	TO Size						TO Centrality					
	Static Images			Dynamic Images			Static Images			Dynamic Images		
	Coef.	SE	t	Coef.	SE	t	Coef.	SE	t	Coef.	SE	t
Image Characteristics												
TO size	-.041	.045	-.91	-.078**	.035	-2.21						
TO size ²	-.052	.082	-.64	-.128**	.065	-1.98						
TO centrality							.161	.111	1.45	-.255**	.102	-2.49
Fixed Effects												
Participant		Included			Included			Included			Included	
Image		Included			Included			Included			Included	
R ²		.111			.075			.117			.089	
N		1,580			1,570			750			790	

* $p < .05$.** $p < .01$.*** $p < .001$.

Notes: SE = Clustered, robust standard error.

between-subjects factor that was crossed with participants and the TO centrality (peripheral, somewhat central, central) or TO size (small, medium, large) is a within-subjects factor that was nested with participants (Judd, Westfall, and Kenny 2012). Thus, respondents saw either dynamic images or static images, but not both, and different text overlay centrality (or text overlay size) for 10 different images. For example, respondents assigned to the dynamic condition would assess 10 dynamic images, each of which varied in text overlay centrality (or size), so they never saw the same dynamic image with a different manipulation of text overlay centrality (or size).

We calculated the desired sample size with a statistical power of .80, an alpha level of .05, and the estimated effect sizes we obtained from the exploratory studies reported in Web Appendix D. The results specified a desired sample size of 323 (153) participants for Study 2a (2b). With the recognition that we excluded about 3% of respondents from our previous studies, we planned to stop the data collection after 333 (157) respondents had completed Study 2a (2b). Due to simultaneous submissions beyond our control, Study 2a (2b) ultimately included 339 (165) respondents, obtained from Prolific. We pre-selected U.S. residents with at least a 95% acceptance rate, who completed at least 500 human intelligence tasks; who are active on Facebook, Instagram, or Twitter; and who are between 18 and 65 years of age. They assessed the 10 randomly assigned social media posts in random order (i.e., “If you saw this tweet online, how likely would you be to ‘like’ or ‘share’ it with others online?”; two separate items, 1 = “extremely unlikely,” and 7 = “extremely likely”; $r_{\text{size}} = .75$, $r_{\text{centrality}} = .77$, both $ps < .001$).

Results Study 2a

In line with the preregistration, we excluded 14 participants who failed attention checks, 4 who did not complete the survey, and 6 who provided inappropriate answers to the open-ended

question, leaving a final sample of 315 participants ($M_{\text{age}} = 38.58$ years; 49.4% women). We use panel regression, including fixed effects for the participants and images to isolate the causal effects of TO size, and we employ effect-coding of the conditions (TO size: small = -1, medium = 0, large = 1). As detailed in Table 3, we find an inverted U-shaped effect of TO size for dynamic images (main: $\beta = -.078$, $p = .03$; squared: $\beta = -.128$, $p = .05$) but no significant effect for static images (main: $\beta = -.041$, $p = .37$; squared: $\beta = -.052$, $p = .52$). The results remain consistent when we consider the images as random effects (Web Appendix C).

Results Study 2b

We excluded 3 participants who failed attention checks, 6 who did not complete the survey, and 2 who provided inappropriate answers to the open-ended question, leaving a final sample of 154 participants ($M_{\text{age}} = 39.47$ years; 46.3% women). We effect-coded the TO conditions (TO centrality: low = -1, medium = 0, high = 1). The results in Table 3 reveal that TO centrality has a negative effect for dynamic images ($\beta = -.255$, $p = .02$) and a nonsignificant effect for static images ($\beta = .161$, $p = .15$). These results hold with images as random effects (Web Appendix C).

Discussion

Our Study 2 findings confirm an inverted U-shaped effect of TO size for dynamic images, such that dynamic images with medium TOs prompt more sharing, as well as a negative effect of TO centrality for dynamic images, such that dynamic images with less central TOs prompt more sharing. However, these experimental studies lack behavioral outcomes. Publicly available social media data also offer only limited evidence of individual-level actions that might indicate engagement beyond sharing, commenting, or liking

(Moe and Schweidel 2017). Thus, we also investigate whether our findings generalize to actual click-through rates (CTRs) of display ads.

Study 3: Manipulating Text Overlays in the Field

Guided by consistent findings of Studies 1 and 2 regarding dynamic images, we focus on this condition and its behavioral outcomes in Study 3. With the assistance of Facebook's Ad Manager (Braun et al. 2024), we compare display ads that feature a dynamic image in combination with TOs of different sizes (Study 3a) and different degrees of centrality (Study 3b), holding all other textual and visual elements constant. We created a fictional business page called "Adventure Tours" to make the sponsored posts look realistic. Web Appendix E provides the stimuli, pretests, descriptive statistics, and robustness tests for both studies.

Results Study 3a

We created three display ads for a ski destination with a dynamic image that differed only in TO size (small, medium, large). We budgeted \$100 for each ad and delivered them for one week, targeting U.S. participants between 18 and 54 years of age with an interest in skiing. We also restricted the ads to desktop applications and the Facebook newsfeed, to minimize variance in delivery channels. The ads prompted a total of 165,617 impressions. The CTRs, measured as the number of clicks on an ad divided by number of times it was displayed, are higher for the medium-sized TO (.33%) than the large TO (.26%; $\chi^2 = 4.57$, $p = .03$) or the small TO (.15%; $\chi^2 = 40.36$, $p < .001$).

Results Study 3b

We created three display ads for a surf destination with a dynamic image and a TO ("Win a Holiday in the Sun!") that appeared peripheral, somewhat central, or central to the image. Again, we budgeted \$100 for each ad, delivered them for one week to U.S. participants between 18 and 54 years of age with an interest in surfing, and restricted the delivery to desktop applications and the Facebook newsfeed. The ads received 93,301 total impressions. The calculated CTR is higher for the peripheral TO (.74%) than the somewhat central TO (.60%; $\chi^2 = 4.78$, $p = .03$) or the central TO (.62%; $\chi^2 = 3.68$, $p = .06$).

Discussion

Studies 3a and 3b replicate the pattern of results from the previous studies. To increase CTRs for posts with dynamic images, the TO should be medium sized, rather than too large or too small, and its positioning should be less central. The CTR increased by 127% (27%) for a medium-sized compared with small (large) TO, and by 24% (20%) for a TO with low centrality compared with medium (high) centrality. Next, we explore the mechanism of these effects.

Study 4: Exploring the Mechanism for Dynamic Images

Thus far, we have established that the most effective social media posts combine dynamic images with TOs that are less salient (i.e., small to medium sized, peripherally placed). In this study, we test our supposition that visual appeal drives these effects. Accordingly, we predict that social media posts that combine dynamic images with less salient TOs are more appealing, aesthetically and technically, than social media posts that combine dynamic images with salient TOs. To test this psychological mechanism, we use two experiments that feature social media ads with dynamic images and TOs of varying sizes (Study 4a) and degrees of centrality (Study 4b).

We examine six potential mediators of engagement—visual appeal, curiosity, creativity, attention, surprise, and elaboration—that we identified through a review of relevant literature pertaining to visual aesthetics, visual salience, and perceptual processing, as well as insights from our interviews with content managers (Web Appendix A). In detail, we predict that the combination of dynamic images with TOs of varying sizes and centrality influence engagement by triggering different levels of curiosity and creativity (Hagtvedt, Patrick, and Hagtvedt 2008). The effects of TO size and centrality also might depend on the heightened attention generated by dynamic images (Zhang, Wedel, and Pieters 2009). If these design combinations evoke surprise, they could lead to increased engagement (Berger and Milkman 2012). Finally, greater engagement could reflect the increased cognitive effort required to assess dynamic images and varying TO characteristics, such that they prompt viewers to engage more deeply to understand the message (Sahni, Wheeler, and Chintagunta 2018).

We measure visual appeal as a combination of visual attractiveness ("Please rate the extent to which the social media post is aesthetically appealing to you"; 1 = "not at all appealing," and 9 = "extremely appealing") and formal execution ("Please rate the extent to which the social media post is well-executed"; 1 = "not at all well-executed," and 9 = "extremely well-executed"). We use single items to measure curiosity ("How would you describe this social media post?"; 1 = "not very interesting," and 9 = "very interesting") and creativity ("How would you describe this social media post?"; 1 = "not very creative," and 9 = "very creative"). The attention measure reflects the time spent on each social media post (in seconds). For surprise, we used the single item, "This social media post is surprising to me" (1 = "not at all," and 9 = "very much"), and for elaboration, we asked, "Please rate the extent to which you would like to read more about the give-away in the social media post" (1 = "not very much," and 9 = "a lot more"). Finally, we measured preference for the social media post with the item, "Please indicate your preference for the social media post you saw" (1 = "do not prefer," and 9 = "prefer a great deal"). Using single-item measures helps limit respondent fatigue, which is pertinent, because we asked them to rate several social media posts. The measures and detailed study results appear in Web Appendix F.

Results Study 4a

We use a three-factor (TO size: small, medium, large) within-subjects design that features three different product images (total of nine social media ads). Each respondent assessed all nine social media posts in random order. We recruited 121 participants from Prolific, who live in the United States, have at least a 95% acceptance rate, and have completed at least 500 human intelligence tasks. After excluding 8 participants who failed an attention check, we retained a final sample of 113 participants ($M_{\text{age}} = 37.09$ years; 46.7% women).

The multilevel analysis includes random intercepts to account for participant-specific effects, fixed effects for the three different products, and effect-coding of the conditions (small TO size = -1 , medium TO size = 0 , large TO size = 1). We identify a positive but nonsignificant main effect ($\beta = .078$, $p = .20$) and negative quadratic effect ($\beta = -.544$, $p < .001$) of TO size on preferences. Simple t -tests confirm greater preferences for images with a medium TO ($M = 3.97$, $SD = 2.18$) compared with a small TO ($M = 3.35$, $SD = 2.17$; $t = 3.70$, $p < .001$) or large TO ($M = 3.49$, $SD = 2.20$; $t = 2.83$, $p < .01$). To test the potential mediators, we specify a multilevel structural equation model (MSEM) (Preacher, Zyphur, and Zhang 2010) that accounts for nested observations and correlations among the competing mediators. Visual appeal is the strongest mediator of the inverted U-shaped effect of TO size on preference ($\beta = -.252$, $p < .001$), stronger than curiosity ($\beta = -.076$, $p = .01$) or elaboration ($\beta = -.063$, $p < .01$). Creativity ($\beta = -.015$, $p = .34$), surprise ($\beta = .007$, $p = .21$), and attention ($\beta = -.007$, $p = .16$) do not mediate this effect.

Results Study 4b

We use a three-factor (TO centrality: peripheral, somewhat central, central) within-subjects design and three different product images (total of nine social media ads). With the same criteria as applied in Study 4a, we recruited 121 different participants through Prolific. After the exclusion of 5 participants who failed an attention check, we retained a final sample of 116 participants ($M_{\text{age}} = 40.52$ years; 50% women). Each respondent assessed all nine configurations in random order. In the multilevel analysis, with random intercepts for participants and fixed effects for the three different products, TO centrality exhibits a negative effect on preferences ($\beta = -.210$, $p < .001$). Simple t -tests indicate greater preferences for images with peripheral TO ($M = 4.03$, $SD = 2.33$) than central TO ($M = 3.63$, $SD = 2.26$, $t = 2.40$, $p = .02$), though no difference relative to somewhat central TO ($M = 4.03$, $SD = 2.27$, $t = .03$, $p = .97$). In the MSEM, we again find that visual appeal is the strongest mediator of the negative effect of TO centrality on preference ($\beta = -.187$, $p < .001$), stronger than curiosity ($\beta = -.049$, $p < .01$), creativity ($\beta = .028$, $p < .01$), and elaboration ($\beta = -.013$, $p = .09$). Attention ($\beta = -.002$, $p = .32$) and surprise ($\beta = -.001$, $p = .58$) do not mediate the effect of TO centrality.

Discussion

Studies 4a and 4b identify visual appeal as the strongest mechanism for explaining the effects of TO size and TO centrality on

preference judgments. Even if they tend to be processed quickly, social media multimodal compositions evidently can evoke consistent visual aesthetics evaluations. Thus, the observed TO effects for dynamic images appear driven primarily by perceptions of visual composition that feel intuitively correct, visually attractive, and formally well-executed, as well as, to a lesser extent, by creativity, curiosity, and elaboration.

Study 5: Field Study on Instagram

To help researchers replicate and extend our findings in the field, and to facilitate measures of dynamism in social media images, we use recent advancement in vision transformers (Dosovitskiy et al. 2020) to build a classifier that can detect the probability of dynamism in an image. The classifier can also be used by social media managers.

Data Collection and Measurement

Noting the importance of the medium for evaluating the effectiveness of image and text characteristics (Grewal, Gupta, and Hamilton 2021) and evidence of platform-specific differences in image effects (Hartmann et al. 2021; Li and Xie 2020), we apply the Study 1 data collection approach to the Instagram posts posted by the eight brands identified in Study 1 between January 2015 and February 2020. Among the initially collected 11,268 Instagram posts, 8,467 featured a single image, so they constitute the sample. The variable operationalizations are the same as in Study 1 (Table 1), except that Instagram does not allow for direct reposts, so we measure engagement with the sum of likes and comments (Cascio Rizzo et al. 2024).⁶ We develop and use an automated classifier for image dynamism, and we rely on Google Cloud Vision to identify TOs automatically, then confirm this assessment by manually annotating their size and centrality.

Development of the Image Classifier

With the annotated Twitter images from Study 1, which we previously had coded on a 7-point static–dynamism scale, we simplified the classification, as static (<4) or dynamic (≥ 4), to create the training data set. For generalizability across platforms, we also coded the dynamism (1–7) of a random 30% sample of Instagram images and added them to the training set. In addition to standard image preprocessing, the images were resized/rescaled to the same resolution (224×224) and normalized across RGB (red, green, and blue) channels. After we removed 384 duplicates to prevent overfitting, the final data set included 8,256 images. We split the annotated data into 90% for training (of which 10% is used for validation) and 10% for testing (Sisodia, Burnap, and Kumar 2024).

By empirically assessing a ResNet50 (He et al. 2016) and vision transformer (Dosovitskiy 2020) algorithm, we chose a

⁶ The results are consistent when we use likes and comments as separate engagement variables.

pretrained vision transformer (ViT) architecture, due to its superior performance when pretrained in large data sets such as ImageNet-21k (Dosovitskiy et al. 2020).⁷ The pretrained deep learning model (base patch 16–224) is designed for image classification tasks; it applies a transformer architecture, originally developed for natural language processing, to sequences of image patches. By dividing an image into fixed-size patches and treating the patches as a sequence, the ViT leverages self-attention mechanisms to capture long-range dependencies across the entire image. This self-attention mechanism allows the model to consider global context throughout the image, rather than primarily relying on local receptive fields, as is the case with ResNet50. This global approach to understanding each image might explain why it is better at identifying patterns of movement, such as dynamism.

To avoid overfitting, we used a dropout layer (i.e., randomly set a fraction of the neurons in the model to 0 during each forward and backward pass). To identify optimal hyperparameters in the validation sample, we followed a grid search strategy with different combinations of batches (16, 32, and 68) and learning rates (.00015, .0001, and .00005), while keeping the number of epochs constant at 4 (beyond this value, no substantial reduction of cross-entropy loss occurred). The best hyperparameters (validation accuracy=.868) were 4 epochs, 64 batches, and a .00005 learning rate. With a model trained on these parameters, we evaluated it with the test set (10% annotated data), which produced an average accuracy of 85.8%. The area under the curve (ROC) is 87.5%. The majority class (static images) achieves an F1 (combination of precision and recall) of .91, and the minority class (dynamic images) F1 is .69. These results indicate good predictive ability when applied to nonannotated pictures.

Using a random sample of 100 Instagram images, we also checked for external validity. The first and second authors annotated a random sample of 100 Instagram images manually (1 = “very static,” 7 = “very dynamic”) and compared it with the results of the automated classifier. The probability of the image being classified as dynamic (0% to 100%) is the independent variable, which is more accurate than using a dummy variable based on probabilities. The comparison reveals a correlation of .87 ($p < .001$).

Results

We provide descriptive statistics and correlations in Web Appendix G.⁸ The VIFs are all below 2.079. As the results in Table 4 reveal, the effect of image dynamism is positive and significant (IRR = 1.139, $p < .001$). A 10% increase in the probability of image dynamism would result in 14% more engagement. Furthermore, in Figure 4,

Panel A, the interaction of image dynamism and TO size follows an inverted U-shape, such that the simple interaction is positive (IRR = 1.029, $p = .02$) and the quadratic interaction is negative (IRR = .999, $p < .01$). For dynamic images, the simple slopes show a nonsignificant main effect ($dy/dx = 1.556$, $p = .25$) and a negative quadratic effect ($dy/dx = -.065$, $p = .04$, Figure 4, Panel B). Initially, increasing TO size prompts more engagement with a dynamic image, but when the TO size expands beyond 11% of the image, increasing it further weakens engagement. If the TO grows so large that it exceeds 44% of the size of the image, engagement decreases to a point even lower than that for a dynamic image without TO. For static images, the simple slopes show a negative main effect ($dy/dx = -1.255$, $p = .004$) and a positive quadratic effect ($dy/dx = .0189$, $p = .006$; Figure 4, Panel C). Thus, engagement with a static image does increase as a TO gets larger.

The interaction of image dynamism and TO centrality is negative and significant (IRR = .824, $p = .02$; Figure 4, Panel D). For dynamic images, simple slopes indicate a negative effect of TO centrality: Increasing TO centrality reduces engagement ($dy/dx = -2.072$, $p = .04$), such that engagement with a typical dynamic image decreases as the TO moves to a more central position. For static images, the simple slopes show a positive but nonsignificant effect of TO centrality ($dy/dx = .306$, $p = .22$). Engagement with a static image does not change as the TO moves to become peripheral. We present these marginal effects in Figure 4, Panels E and F.

Robustness Tests

As in Study 1, we applied PSM as an alternative model specification, added a robustness check accounting for the semantic similarity of the TO with the caption and image according to BERT embeddings, and ran a model with the three additional TO features (font, alignment, and color contrast). We also tested a model that includes only the first 150 characters of the Instagram caption, in accordance with real-world considerations; most Instagram users access the site from their smartphones, on which only the first part of the post is visible before they must click to see “more.” All the analyses produce results consistent with our main analysis (Web Appendix G).

Discussion

Using a dynamism classifier developed for this research, we assess the effects of TO size and centrality on Instagram posts. In line with the empirical results from the preceding studies, we find that the effectiveness of the TO depends on the dynamism of the image with which it is paired. Dynamic images benefit most from TOs that are about 11% the size of the image, placed less centrally. Details on how to use the dynamism classifier are reported in Web Appendix H.⁹

⁷ With Resnet50, we achieved an overall accuracy of .82; the precision and recall values are .65 and .60 for dynamic images and .87 and .89 for static images, respectively.

⁸ We observe 973 missing values for the hour in which the post was made. In these cases, we imputed the average time of the post, depending on the order in which it appeared on a particular day.

⁹ It is possible to access the ViT model directly, using a low-code visual workflow on the KNIME Hub: <https://hub.knime.com/-/spaces/-/~HU6v5648F37uAwcS/current-state/>.

Table 4. Results of Study 5.

	Model 3			Model 4		
	IRR	Coeff.	Z	IRR	Coeff.	z
Independent Variables						
Image dynamism	1.153***	.142	5.22	1.139***	.130	4.62
TO size	.992*	−.008	−2.55	.989**	−.011	−3.07
TO size ²	1.000*	.000	2.06	1.000***	.000	2.98
TO centrality	1.008	.008	.50	1.034	.034	1.63
Interactions						
Image dynamism × TO size				1.029*	.029	2.37
Image dynamism × TO size ²				.999**	−.001	−2.68
Image dynamism × TO centrality				.824*	−.194	−2.38
Image Controls						
Human presence	1.097***	.093	2.06	1.100***	.096	2.13
Color concentration	.997***	−.003	−5.53	.997***	−.003	−5.54
Color saturation	.954***	−.047	−2.20	.953***	−.048	−2.26
Duplicate image	.967	−.034	−.91	.967***	−.034	−.91
TO Controls						
TO–caption similarity	1.394	.332	1.32	1.371	.315	1.25
Sales promotion	1.095	.090	.75	1.074	.072	.60
Product	1.357***	.305	4.54	1.343***	.295	4.33
TO–image similarity	1.356***	.305	.16	1.904	.644	.34
Caption Controls						
Length	1.000***	.000	7.34	1.000***	.000	7.35
Emphasis	.967***	−.034	−5.27	.967***	−.034	−5.29
Valence	.999	−.001	−.86	.999	−.001	−.83
Arousal	.996	−.004	−.07	.993	−.007	−.12
Sales promotion	.945***	−.057	−3.34	.945	−.057	−3.36
Complexity	1.001*	.001	2.53	1.001***	.001	2.44
Question	.954**	−.047	−2.82	.954**	−.047	−2.79
Emojis	.982***	−.019	−3.84	.982***	−.018	−3.83
Hashtags	.873***	−.136	−8.94	.874***	−.134	−8.85
URL	.788***	−.238	−7.25	.789***	−.238	−7.23
Mention	.990***	−.010	−.70	.991***	−.009	−.66
Additional Controls						
Weekend	1.069***	−.178	4.99	1.070***	.067	5.02
Carryover effect	1.000***	.067	23.67	1.000***	.000	23.61
Fixed Effects						
Brand	Included			Included		
Hour of day	Included			Included		
Significance (Chi ²)	22,369***			22,378***		
Nagelkerke pseudo-R ²	.928			.929		
N (AIC)	8,459 (171,323)			8,459 (171,320)		

†*p* < .10.**p* < .05.***p* < .01.****p* < .001.

Notes: AIC = Akaike information criterion. An IRR greater than (less than) 1 indicates a positive (negative) effect. The IRRs for TO size are rescaled to 10 percentage points for ease of interpretation. Robustness tests are reported in Web Appendix G.

General Discussion

The vast visual content available on social media platforms, combined with the common user practice of scrolling quickly through this content, challenges companies to get their social media communication noticed. In this research,

we address the option of adding TOs to the visual space of an image, a popular managerial practice that can help social media posts gain consumers' attention. As we show, overlaying text on images requires a well-informed strategy, to ensure that the overall visual experience is not hindered. With this

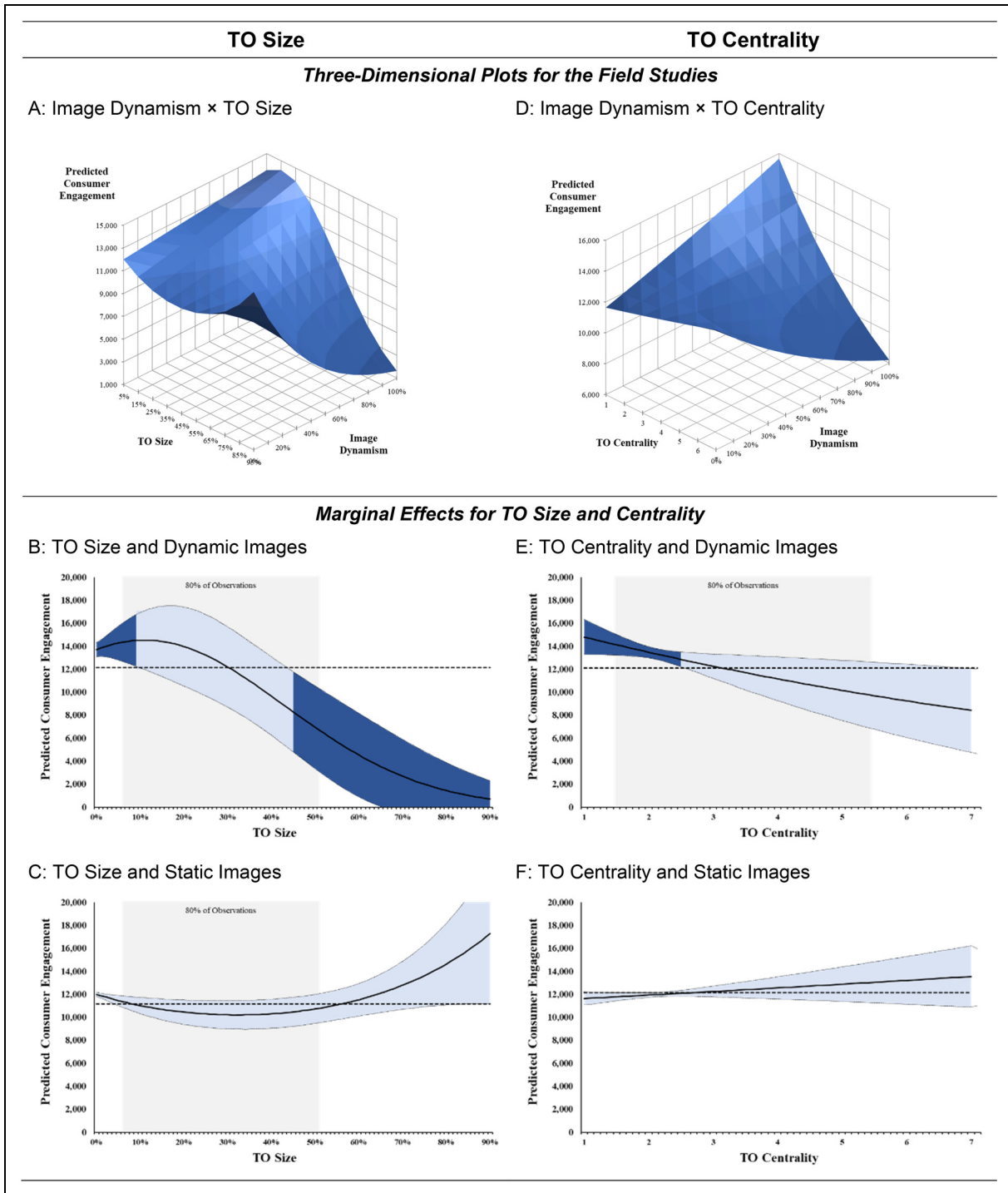


Figure 4. Effects of Text Overlay Size and Centrality in Study 5.

Notes: The three-dimensional plots display the full range of observations. Marginal effects with 95% confidence intervals. The area in gray is where 80% of observations are located. Dotted lines are the predicted means for dynamic (static) images without TO. The dark blue color indicates areas where a TO is beneficial (detrimental) for engagement.

research, we offer theoretically well-founded insights into how to combine images and TOs to keep them from interfering with each other or undermining consumers' social media engagement.

Theoretical Contributions

This research makes three main theoretical contributions. First, as a contribution to the social media engagement literature, we shed light on the joint effects of image type and TO on

consumer engagement. A few existing studies have addressed the combined effects of images and captions (e.g., Li and Xie 2020; Villarroel Ordenes et al. 2019). Relative to captions that appear outside the visual frame of the image, however, superposing TOs exerts unique influences, and their capacity to trigger engagement depends on the salience of the combined features. As we show, increasing the size of a TO on a dynamic image on Twitter or Instagram can increase engagement, but if that TO becomes too large, it hinders engagement. Placing the TO peripherally (vs. centrally) also appears to be the best choice for dynamic images. Notably, some of our findings conflict with recommendations for print and television commercials. Specifically, previous research predicts that increasing the size of the text will not detract attention from the picture (Pieters and Wedel 2004) and placing text centrally does not affect behavioral intentions about fast-forwarding television commercial (Brasel and Gips 2008). But in our social media research context, large TOs that appear centrally on dynamic images lead to diminished consumer engagement. Static images in combination with different levels of TO salience do not exert similar causal effects on consumers' engagement though, in contrast with our predictions. Instead, we find no impact of increasing the size of the TO in static images. The absence of these predicted effects might signify an overuse of static images, such that they no longer evoke much attention (as indicated by generally low engagement in all our studies). We speculate that salient TO features are not enough to compensate for the inability of static images to grab consumers' attention. In this case, our findings for static images do not imply an absence of effects; rather, the nonsignificant results might reflect the specific static images we included in our studies. We address this point further in the limitations and further research section.

Second, for research on visual saliency, we establish that insights pertaining to text salience in print ads (Pieters and Wedel 2004, 2007) do not necessarily apply to TOs in social media communication. When images and TOs appear together, it is critical to balance their salience. In multimodal compositions posted to social media, no specific element should be emphasized at the expense of others. If marketers decide to emphasize TOs, the outcomes could be suboptimal if they fail to consider carefully which *type* of image (static or dynamic) their post also features. Our premise is particularly important for social media content displayed on small devices like mobile phones, because the boundaries of visual elements that share the same space likely become blurred, so consumers, scrolling quickly through their feeds, might struggle to differentiate what counts as text and what constitutes the image (Bateman 2014).

Third, addressing calls for further research into visual and multimodal aesthetics (Hagtvedt 2022), we highlight that the visual appeal of multimodal social media communication is particularly influential, among multiple underlying mechanisms. Even when consumers quickly scroll through their feeds and spend minimal time processing posts, social media managers should prioritize creating visually attractive and well-executed

designs to evoke the positive feelings associated with visual appeal. Engaging consumers in the preattentive stage is crucial, because they form immediate overall impressions of posts' visual structure (Affonso and Janiszewski 2023). If their early impression aligns with viewers' intuitive sense of what looks visually appealing, they are more likely to engage with the social media post. Our work further highlights the importance of effectively combining the size and centrality of TOs with the dynamism of images to create visually appealing multimodal social media compositions (Grewal et al. 2022). Studying the combination of image dynamism with TO size and centrality also reveals insights that would not be evident if we had studied these elements separately. Because suboptimal combinations of image and TOs, which fail to balance the salience of the elements, prevent intuitive perceptions of visual appeal, they decrease the likelihood that viewers will engage with the content.

Our findings further corroborate and advance research in several other marketing domains. In line with investigations of print ads (Cian, Krishna, and Elder 2014; Peracchio and Meyers-Levy 2005), we confirm that images depicting dynamism result in more engagement. Similarity between the caption and TO has a negative effect on engagement on Twitter (nonsignificant on Instagram), which aligns with predictions of complementarity between captions and visual content (Villarroel Ordenes et al. 2019). Although human presence appears negatively associated with engagement on Twitter, it has the opposite effect on Instagram. This discrepancy parallels the platform-specific differences identified by recent research on user-generated images shared on Twitter and Instagram (Hartmann et al. 2021; Li and Xie 2020). We also note that greater color concentration (few dominant colors) has a positive effect on engagement on Twitter, but not on Instagram, likely because Instagram is a more visually colorful platform, on which color diversity relates to users' extraversion (Kim and Kim 2019). Similar to Jalali and Papatla (2019), we find that on Twitter, text referring to sales promotions has positive effects on sharing. We do not find any significant effects of text valence or arousal, possibly because consumers already expect positive, aroused content in brand posts. Borah et al. (2020) offer similar evidence related to valence, suggesting that the effect of arousal might be contingent on the type of emotion and posting time (Kanuri, Chen, and Sridhar 2018).

Managerial Implications

Our interviews indicate that managers recognize the practical relevance of exploring the interplay of TOs and images, but it is less clear to them *how* and *why* combining verbal and visual modes, each of which separately can attract attention, is beneficial or detrimental for consumer engagement. Our findings help in this regard. We clarify that combinations of images and TOs must feel visually appealing to the consumers to attract their attention. Specifically, increasing the TO size for a dynamic image on Twitter (Instagram) can increase engagement relative to an image without a TO, but only to the point

that the TO covers 46% (44%) of the image. Increasing TO size further erases all the positive effect of using a TO, such that consumer engagement might fall below the level achieved by a dynamic image without a TO. For dynamic images, moving the TO from the center to a peripheral region of the image can increase engagement by up to 43%.

Managers should assess the dynamism of the images they use before combining them with TOs of the appropriate size and centrality, to ensure positive and avoid detrimental effects on consumer engagement. To help them do so, we complement our theorizing with a user-friendly app that simulates how engagement increases with different combinations of images and TOs that vary in their salience (see Web Appendix I). Our open-source, interactive app (available at <https://www.knime.com/visual-flow-data-app>) can facilitate automatic image selection, TO design processes, and A/B testing of multimodal social media posts. This app has an intuitive user interface and is designed for managers interested in effectively designing and predicting the engagement of multimodal social media posts on Instagram or Twitter (now X). Its open-source nature makes this app an evolving product; it will continue to be enhanced with improved modeling and additional data. At the current stage of development, it offers multiple, pertinent functionalities for social media marketing managers and practitioners. For example, they can improve content creation processes by leveraging its basic graphic design help (e.g., classifying the image as static or dynamic).

Limitations and Further Research

Although we obtain consistent results across multiple studies and broad support for the proposed effects of image dynamism and TOs, we acknowledge some limitations that we hope will stimulate further research. First, we did not find consistent evidence that more salient TOs increase engagement with static images. Continued research, focused on static images specifically, could delve further into this finding and identify which, if any, additional features might make them more engaging. Second, we prioritize single TOs in our research, because they represent the most common case, according to our field data gathered from Twitter (64%) and Instagram (92%). But there can be more than one TO on an image. Our results remain consistent when we control for posts with more than one TO, but we hope additional research will investigate the intricacies of multiple TOs superposed on the same image. Third, continued research might study the use of TOs in different communication contexts (e.g., websites, shopping ads). For example, brand posts might tell stories with multiple pictures, showcased at the same time or in a specific order (e.g., carousels, albums, Instagram galleries). Extending the research focus from single images to sequences of multiple images, or to videos and GIFs, could reveal new insights on their interplay with TOs (even dynamic TOs).

We also hope our image classifier and interactive app stimulate further research. Both tools reflect the contextual boundaries of our research project. For example, to simplify

the classification, we categorize images as static or dynamic. Additional research could use continuous scores to train a regression rather than a discrete classifier. The vision transformer relies on pretrained weights from the ImageNet-21k data set, which is an object classification task. Pretrained data from alternative resources that capture activities from videos (e.g., ActivityNet) or that use different ViT architectures might be informative as well. Thus far, we have trained the classifier only with images from the eight brands in our samples, so it needs further development to offer predictions in other contexts. This task might require more extensive training data (e.g., annotated images on dynamism) and fine tuning; we plan to undertake such tasks through further app development efforts.

Acknowledgments

The authors are grateful for valuable feedback from Henrik Hagtvedt, Jochen Hartmann, Grant Packard, Isamar Troncoso, Aradhna Krishna, and Juanjuan Zhang, as well as insightful comments from participants at the 2024 Customer Journey Conference, EACR 2023, and ACR 2022. They also thank the KNIME Analytics team for supporting the development and hosting of the dynamism classifier and research app.

The authors further express their deep gratitude to the *JM* review team for their invaluable guidance and constructive suggestions.

The first and second authors collected and analyzed the managerial interview data. The second author collected data for Studies 1 and 5, which were analyzed by the second and third authors. The first and third authors collected and analyzed data for Studies 2a, 2b, 4a, and 4b. The third author collected and analyzed data for Studies 3a and 3b.

Coeditor

Shrihari Sridhar

Associate Editor

Caleb Warren

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Stefania Farace  <https://orcid.org/0000-0002-3107-8105>

Dennis Herhausen  <https://orcid.org/0000-0002-4335-1703>

Dhruv Grewal  <https://orcid.org/0000-0002-7046-6063>

References

- Affonso, Felipe M. and Chris Janiszewski (2023), "Marketing by Design: The Influence of Perceptual Structure on Brand Performance," *Journal of Marketing*, 87 (53), 736–54.
- Agnihotri, Raj, Rebecca Dingus, Michael Y. Hu, and Michael T. Krush (2016), "Social Media: Influencing Customer Satisfaction in B2B Sales," *Industrial Marketing Management*, 53, 172–80.

- Appel, Gil, Lauren Grewal, Rhonda Hadi, and Andrew T. Stephen (2020), "The Future of Social Media in Marketing," *Journal of the Academy of Marketing Science*, 48 (1), 79–95.
- Arnheim, Rudolf (1954), *Art and Visual Perception: A Psychology of the Creative Eye*. University of California Press.
- Atalay, A. Selin, H. Onur Bodur, and Dina Rasolofoarison (2012), "Shining in the Center: Central Gaze Cascade Effect on Product Choice," *Journal of Consumer Research*, 39 (4), 848–66.
- Ayres, Scott (2020), "Should Text Be on Instagram Image Posts: Yes or No?" AgoraPulse (February 4), <https://www.agorapulse.com/social-media-lab/text-instagram-image-posts/>.
- Balas, Benjamin, Lisa Nakano, and Ruth Rosenholtz (2009), "A Summary-Statistic Representation in Peripheral Vision Explains Visual Crowding," *Journal of Vision*, 9 (12), 1–18.
- Barnes, Nora G., Allison Kane, and Kylie Maloney (2018), "The 2018 Fortune 500 Target Millennials and Seek Uncensored Expression," (accessed February 25, 2021), <https://www.conference-board.org/blog/human-capital/Fortune-500-Target-Millennials>.
- Bateman, John A. (2014), *Text and Image. A Critical Introduction to the Visual/Verbal Divide*. Routledge.
- Berger, Jonah, Ashlee Humphreys, Stephan Ludwig, Wendy W. Moe, Oded Netzer, and David A. Schweidel (2020), "Uniting the Tribes: Using Text for Marketing Insight," *Journal of Marketing*, 84 (1), 1–25.
- Berger, Jonah and Katherine L. Milkman (2012), "What Makes Online Content Viral?" *Journal of Marketing Research*, 49 (2), 192–205.
- Berlyne, Daniel E. (1974), *Studies in the New Experimental Aesthetics: Steps Toward an Objective Psychology of Aesthetic Appreciation*. Hemisphere Publishing Corporation.
- Blanchard, Simon J., Jacob Goldenberg, Koen Pauwels, and David A. Schweidel (2022), "Promoting Data Richness in Consumer Research: How to Develop and Evaluate Articles with Multiple Data Sources," *Journal of Consumer Research*, 49 (2), 359–72.
- Boegershausen, Johannes, Hannes Datta, Abhishek Borah, and Andrew T. Stephen (2022), "Fields of Gold: Scraping Web Data for Marketing Insights," *Journal of Marketing*, 86 (5), 1–20.
- Borah, Abhishek, Sourindra Banerjee, Yu-Ting Lin, Apurv Jain, and Andreas B. Eisingerich (2020), "Improvised Marketing Interventions in Social Media," *Journal of Marketing*, 84 (2), 69–91.
- Brasel, S. Adam and James Gips (2008), "Breaking Through Fast-Forwarding: Brand Information and Visual Attention," *Journal of Marketing*, 72 (6), 31–48.
- Braun, Michael, Bart De Langhe, Stefano Puntoni, and Eric M. Schwartz (2024), "Leveraging Digital Advertising Platforms for Consumer Research," *Journal of Consumer Research*, 51 (1), 119–28.
- Breitmeyer, Bruno G. (2007), "Visual Masking: Past Accomplishments, Present Status, Future Developments," *Advances in Cognitive Psychology*, 3 (1–2), 9–20.
- Cascio Rizzo, Giovanni L., Francisco Villarroel Ordenes, Rumen Pozharliev, Matteo De Angelis, and Michele Costabile (2024), "How High-Arousal Language Shapes Micro-Versus Macro-Influencers' Impact," *Journal of Marketing*, 88 (4), 107–28.
- Chaiken, Shelley and Yaacov Trope (1999), *Dual-Process Theories in Social Psychology*. Guilford Press.
- Chan, Irene Chen Chu, Zhaoyu Chen, and Daniel Leung (2023), "The More the Better? Strategizing Visual Elements in Social Media Marketing," *Journal of Hospitality and Tourism Management*, 54 (March), 268–89.
- Chen, Siyun, Veronika Ponomarenko, Linxiang Lv, and David Ahlstrom (2024), "Visual Complexity, Brand Gender, and Ad Effectiveness," *International Journal of Research in Marketing* (published online September 21), <https://doi.org/10.1016/j.ijresmar.2024.09.004>.
- Chun, Marvin M. and Jeremy M. Wolfe (2005), "Visual Attention," in *Blackwell Handbook of Sensation and Perception*, E. Bruce Goldstein, ed. John Wiley & Sons, 272–310.
- Cian, Luca, Aradhna Krishna, and Ryan S. Elder (2014), "This Logo Moves Me: Dynamic Imagery from Static Images," *Journal of Marketing Research*, 51 (2), 184–97.
- Cian, Luca, Aradhna Krishna, and Ryan S. Elder (2015), "A Sign of Things to Come: Behavioral Change Through Dynamic Iconography," *Journal of Consumer Research*, 41 (6), 1426–46.
- The Content Authority (2023), "How To Use 'Overlay' in a Sentence: Usage and Examples," (accessed July 2024), <https://thecontentauthority.com/blog/how-to-use-overlay-in-a-sentence>.
- Dosovitskiy, Alexey, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby (2020), "An Image is Worth 16 (16 Words: Transformers for Image Recognition at Scale)," arXiv, <https://doi.org/10.48550/arXiv.2010.11929>.
- Fang, Ying, Liyu Zhu, Xueni Cao, Liqun Zhang, and Xiaodong Li (2020), "Visual Saliency: How Text Influences," in *Social Computing and Social Media. Design, Ethics, User Behavior, and Social Network Analysis*, G. Meiselwitz, ed. Springer.
- Firoze, Adnan, Tousif Osman, Shahreen S. Psyche, and Rashedur M. Rahman (2018), "Scoring Photographic Rule of Thirds in a Large MIRFLICKR Dataset: A Showdown Between Machine Perception and Human Perception of Image Aesthetics," in *Asian Conference on Intelligent Information and Database Systems*. Springer, 466–75.
- Gardner, Jonathan S., Charless Fowlkes, Christine Nothelfer, and Stephen E. Palmer (2008), "Exploring Aesthetic Principles of Spatial Composition Through Stock Photography," *Journal of Vision*, 8 (6), 337–337.
- Gavilanes, José Manuel, Tessa Christina Flatten, and Malte Brettel (2018), "Content Strategies for Digital Consumer Engagement in Social Networks: Why Advertising Is an Antecedent of Engagement," *Journal of Advertising*, 47 (1), 4–23.
- Gershoni, Sharon and Shaul Hochstein (2011), "Measuring Pictorial Balance Perception at First Glance Using Japanese Calligraphy," *I-perception*, 2 (6), 508–27.
- Gong, Shiyang, Juanjuan Zhang, Ping Zhao, and Xuping Jiang (2017), "Tweeting as a Marketing Tool: Field Experiment in the TV Industry," *Journal of Marketing Research*, 54 (6), 833–50.
- Grewal, Rajdeep, Sachin Gupta, and Rebecca Hamilton (2021), "Marketing Insights from Multimedia Data: Text, Image, Audio, and Video," *Journal of Marketing Research*, 58 (6), 1025–33.
- Grewal, Dhruv, Dennis Herhausen, Stephan Ludwig, and Francisco Villarroel Ordenes (2022), "The Future of Digital Communication Research: Considering Dynamics and Multimodality," *Journal of Retailing*, 98 (2), 224–40.

- Grinstein, Amir, Henrik Hagtvedt, and Ann Kronrod (2019), "Aesthetically (Dis)Pleasing Visuals: A Dual Pathway to Empathy and Prosocial Behavior," *International Journal of Research in Marketing*, 36 (1), 83–99.
- Hagtvedt, Henrik (2011), "The Impact of Incomplete Typeface Logos on Perceptions of the Firm," *Journal of Marketing*, 75 (4), 86–93.
- Hagtvedt, Henrik (2022), "A Brand (New) Experience: Art, Aesthetics, and Sensory Effects," *Journal of the Academy of Marketing Science*, 50 (3), 425–8.
- Hagtvedt, Henrik and S. Adam Brasel (2017), "Color Saturation Increases Perceived Product Size," *Journal of Consumer Research*, 44 (2), 396–413.
- Hagtvedt, Henrik and Vanessa M. Patrick (2008), "Art Infusion: The Influence of Visual Art on the Perception and Evaluation of Consumer Products," *Journal of Marketing Research*, 45 (3), 379–89.
- Hagtvedt, Henrik, Vanessa M. Patrick, and Reidar Hagtvedt (2008), "The Perception and Evaluation of Visual Art," *Empirical Studies of the Arts*, 26 (2), 197–218.
- Hartmann, Jochen, Mark Heitmann, Christina Schamp, and Oded Netzer (2021), "The Power of Brand Selfies in Consumer-Generated Brand Images," *Journal of Marketing Research*, 58 (6), 1159–77.
- He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun (2016), "Deep Residual Learning for Image Recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–78.
- Hekkert, Paul (2006), "Design Aesthetics: Principles of Pleasure in Design," *Psychology Science*, 48 (2), 157–72.
- Henderson, John M., Phillip A. Weeks, Jr., and Andrew Hollingworth (1999), "The Effects of Semantic Consistency on Eye Movements During Complex Scene Viewing," *Journal of Experimental Psychology: Human Perception and Performance*, 25 (1), 210–28.
- Herhausen, Dennis, Stephan Ludwig, Dhruv Grewal, Jochen Wulf, and Marcus Schoegel (2019), "Detecting, Preventing, and Mitigating Online Firestorms in Brand Communities," *Journal of Marketing*, 83 (3), 1–21.
- Hesterberg, Karla (2021), "Facebook Text Overlay: How to Use the 20% Rule to Improve Your Ads," HubSpot (August 25), <https://blog.hubspot.com/marketing/facebook-text-overlay>.
- Hubbard, Timothy L. (1995), "Cognitive Representation of Motion: Evidence for Friction and Gravity Analogues," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21 (1), 241–54.
- Jalali, Nima Y. and Purushottam Papatla (2019), "Composing Tweets to Increase Retweets," *International Journal of Research in Marketing*, 36 (4), 647–68.
- Judd, Charles M., Jacob Westfall, and David A. Kenny (2012), "Treating Stimuli as a Random Factor in Social Psychology: A New and Comprehensive Solution to a Pervasive but Largely Ignored Problem," *Journal of Personality and Social Psychology*, 103 (1), 54–69.
- Kane, Gerald C. and Alexandra Pear (2016), "The Rise of Visual Content Online," *MIT Sloan Management Review* (January 4), <https://sloanreview.mit.edu/article/the-rise-of-visual-content-online/>.
- Kanuri, Vamsi K, Yixing Chen, and Shrihari (Hari) Sridhar (2018), "Scheduling Content on Social Media: Theory, Evidence, and Application," *Journal of Marketing*, 82 (6), 89–108.
- Kim, Jang Hyun and Yunhwan Kim (2019), "Instagram User Characteristics and the Color of Their Photos: Colorfulness, Color Diversity, and Color Harmony," *Information Processing & Management*, 56 (4), 1494–1505.
- Kosslyn, Stephen M. (2006), *Graph Design for the Eye and Mind*. Oxford University Press.
- Kress, Gunther and Theo van Leeuwen (2006), *Reading Images. The Grammar of Visual Design*. Routledge.
- Kwan, Canice M.C., Xianchi Dai, and Robert S. Wyer, Jr. (2017), "Contextual Influences on Message Persuasion: The Effect of Empty Space," *Journal of Consumer Research*, 44 (2), 448–64.
- Li, Yiyi and Ying Xie (2020), "Is a Picture Worth a Thousand Words? An Empirical Study of Image Content and Social Media Engagement," *Journal of Marketing Research*, 57 (1), 1–19.
- Locher, Paul J. (2003), "An Empirical Investigation of the Visual Rightness Theory of Picture Perception," *Acta Psychologica*, 114 (2), 147–64.
- Locher, Paul J., Jeffrey K. Smith, and Lisa F. Smith (2001), "The Influence of Presentation Format and Viewer Training in The Visual Arts on the Perception of Pictorial and Aesthetic Qualities of Paintings," *Perception*, 30 (4), 449–65.
- Locher, Paul J., Pieter Jan Stappers, and Kees Overbeeke (1999), "An Empirical Evaluation of the Visual Rightness Theory of Pictorial Composition," *Acta Psychologica*, 103 (3), 261–80.
- Luangrath, Andrea Webb, Joann Peck, and Victor A. Barger (2017), "Textual Paralanguage and its Implications for Marketing Communications," *Journal of Consumer Psychology*, 27 (1), 98–107.
- Marinova, Detelina, Sunil K. Singh, and Jagdip Singh (2018), "Frontline Problem-Solving Effectiveness: A Dynamic Analysis of Verbal and Nonverbal Cues," *Journal of Marketing Research*, 55 (2), 178–92.
- McShane, Lindsay, Ethan Pancer, Maxwell Poole, and Qi Deng (2021), "Emoji, Playfulness, and Brand Engagement on Twitter," *Journal of Interactive Marketing*, 53, 96–110.
- Milosavljevic, Milica, Vidhya Navalpakkam, Christof Koch, and Antonio Rangel (2012), "Relative Visual Saliency Differences Induce Sizable Bias in Consumer Choice," *Journal of Consumer Psychology*, 22 (1), 67–74.
- Moe, Wendy W. and David A. Schweidel (2017), "Opportunities for Innovation in Social Media Analytics," *Journal of Product Innovation Management*, 34 (5), 697–702.
- Moffett, Jordan W, Judith Anne Garretson Folse, and Robert W. Palmatier (2021), "A Theory of Multiformat Communication: Mechanisms, Dynamics, and Strategies," *Journal of the Academy of Marketing Science*, 49, 441–61.
- Overgoor, Gijs, William Rand, and Masoud Mazloom (2022), "Simplicity Is Not Key: Understanding Firm-Generated Social Media Images and Consumer Liking," *International Journal of Research in Marketing*, 39 (3), 639–55.
- Packard, Grant and Jonah Berger (2023), "The Emergence and Evolution of Consumer Language Research," *Journal of Consumer Research*, 51 (1), 42–51.
- Pancer, Ethan, Vincent Chandler, Maxwell Poole, and Theodore J. Noseworthy (2019), "How Readability Shapes Social Media Engagement," *Journal of Consumer Psychology*, 29 (2), 262–70.

- Patrick, Vanessa M. and Henrik Hagtvædt (2011), "Aesthetic Incongruity Resolution," *Journal of Marketing Research*, 48 (2), 393–402.
- Peracchio, Laura A. and Joan Meyers-Levy (2005), "Using Stylistic Properties of Ad Pictures to Communicate with Consumers," *Journal of Consumer Research*, 32 (1), 29–40.
- Pieters, Rik and Michel Wedel (2004), "Attention Capture and Transfer in Advertising: Brand, Pictorial, and Text-Size Effects," *Journal of Marketing*, 68 (2), 36–50.
- Pieters, Rik and Michel Wedel (2007), "Goal Control of Attention to Advertising: The Yarns Implication," *Journal of Consumer Research*, 34 (2), 224–33.
- Pieters, Rik, Michel Wedel, and Rajeev Batra (2010), "The Stopping Power of Advertising: Measures and Effects of Visual Complexity," *Journal of Marketing*, 74 (5), 48–60.
- Pieters, Rik, Michel Wedel, and Jie Zhang (2007), "Optimal Feature Advertising Design Under Competitive Clutter," *Management Science*, 53 (11), 1815–28.
- Poor, Morgan, Adam Duhachek, and H. Shanker Krishnan (2013), "How Images of Other Consumers Influence Subsequent Taste Perceptions," *Journal of Marketing*, 77 (6), 124–39.
- Pracejus, John W., G. Douglas Olsen, and Thomas C. O'Guinn (2006), "How Nothing Became Something: White Space, Rhetoric, History, and Meaning," *Journal of Consumer Research*, 33 (1), 82–90.
- Preacher, Kristopher J., Michael J. Zyphur, and Zhen Zhang (2010), "A General Multilevel SEM Framework for Assessing Multilevel Mediation," *Psychological Methods*, 15 (3), 209–33.
- Rosenholtz, Ruth and Zhenlan Jin (2005), "A Computational Form of the Statistical Saliency Model for Visual Search," *Journal of Vision*, 5(8), 777.
- Sahni, Navdeep S., S. Christian Wheeler, and Pradeep Chintagunta (2018), "Personalization in Email Marketing: The Role of Noninformative Advertising Content," *Marketing Science*, 37 (2), 236–58.
- Salamon, Maureen (2024), "Doomscrolling Dangers," Harvard Health Publishing (September 1), <https://www.health.harvard.edu/mind-and-mood/doomscrolling-dangers>.
- Sample, Kevin L., Henrik Hagtvædt, and S. Adam Brasel (2020), "Components of Visual Perception in Marketing Contexts: A Conceptual Framework and Review," *Journal of the Academy of Marketing Science*, 48 (3), 405–21.
- Schweidel, David A. and Wendy W. Moe (2014), "Listening in on Social Media: A Joint Model of Sentiment and Venue Format Choice," *Journal of Marketing Research*, 51 (4), 387–402.
- Sevilla, Julio and Robert J. Meyer (2020), "Leaving Something for the Imagination: The Effect of Visual Concealment on Preferences," *Journal of Marketing*, 84 (4), 109–26.
- Sheikh, Mahnoor (2024), "How the Twitter Algorithm Works in 2024 [6 Strategies]," *Sprout Social* (May 28), <https://sproutsocial.com/insights/twitter-algorithm/>.
- Sisodia, Ankit, Alex Burnap, and Vineet Kumar (2024), "Generative Interpretable Visual Design: Using Disentanglement for Visual Conjoint Analysis," *Journal of Marketing Research*, 62 (3), 405–28.
- Smith, Andrew N. and Eileen Fischer (2021), "Pay Attention, Please! Person Brand Building in Organized Online Attention Economies," *Journal of the Academy of Marketing Science*, 49, 258–79.
- Sutarwala, Umme (2024), "Social Media Statistics You Need to Know in 2024," *Sprinklr* (February 1), <https://www.sprinklr.com/blog/social-media-statistics/>.
- Tatler, Benjamin W. (2007), "The Central Fixation Bias in Scene Viewing: Selecting an Optimal Viewing Position Independently of Motor Biases and Image Feature Distributions," *Journal of Vision*, 7 (14), 1–17.
- Treisman, Anne M. and Garry Gelade (1980), "A Feature-Integration Theory of Attention," *Cognitive Psychology*, 12 (1), 97–136.
- Twitter (2022), "Creative Ad Specifications," (accessed December 10, 2022), <https://business.twitter.com/en/help/campaign-setup/creative-ad-specifications.html>.
- Van der Lans, Ralf, Rik Pieters, and Michel Wedel (2021), "Online Advertising Suppresses Visual Competition During Planned Purchases," *Journal of Consumer Research*, 48 (3), 374–93.
- Van Doorn, Jenny, Katherine N. Lemon, Vikas Mittal, Stephan Nass, Doreén Pick, Peter Pimer, and Peter C. Verhoef (2010), "Customer Engagement Behavior: Theoretical Foundations and Research Directions," *Journal of Service Research*, 13 (3), 253–66.
- Van Rompay, Thomas J.L., Peter W. De Vries, and Xaviera G. Van Venrooij (2010), "More Than Words: On the Importance of Picture–Text Congruence in the Online Environment," *Journal of Interactive Marketing*, 24 (1), 22–30.
- Velarde, Orana (2017), "Quick Tips for Creating Social Media Graphics Like a Designer," *Visme* (December 30) <https://visme.co/blog/add-text-to-image/>.
- Villaruel Ordenes, Francisco, Dhruv Grewal, Stephan Ludwig, Ko De Ruyter, Dominik Mahr, and Martin Wetzels (2019), "Cutting Through Content Clutter: How Speech and Image Acts Drive Consumer Sharing of Social Media Brand Messages," *Journal of Consumer Research*, 45 (5), 988–1012.
- Walker, Lorna, Paul R. Baines, Radu Dimitriu, and Emma K. Macdonald (2017), "Antecedents of Retweeting in a (Political) Marketing Context," *Psychology & Marketing*, 34 (3), 275–93.
- Wies, Simone, Alexander Bleier, and Alexander Edeling (2023), "Finding Goldilocks Influencers: How Follower Count Drives Social Media Engagement," *Journal of Marketing*, 87 (3), 383–405.
- Wilson, Ashley and Anjan Chatterjee (2005), "The Assessment of Preference for Balance: Introducing a New Test," *Empirical Studies of the Arts*, 23 (2), 165–80.
- Wolfe, Jeremy M (1998), "What Can 1 Million Trials Tell Us About Visual Search?" *Psychological Science*, 9 (1), 33–9.
- Wooley, Kaitlin and Marissa A. Sharif (2022), "The Psychology of Your Scrolling Addiction," *Harvard Business Review* (January 31), <https://hbr.org/2022/01/the-psychology-of-your-scrolling-addiction>.
- Zhang, Shunyun, Dokyun Lee, Param Vir Singh, and Kannan Srinivasan (2022), "What Makes a Good Image? Airbnb Demand Analytics Leveraging Interpretable Image Features," *Management Science*, 68 (8), 5644–66.
- Zhang, Jie, Michel Wedel, and Rik Pieters (2009), "Sales Effects of Attention to Feature Advertisements: A Bayesian Mediation Analysis," *Journal of Marketing Research*, 46 (5), 669–81.