ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DOTTORATO DI RICERCA IN

MANAGEMENT

Ciclo 37

Settore Concorsuale: 13/B2 - ECONOMIA E GESTIONE DELLE IMPRESE

Settore Scientifico Disciplinare: SECS-P/08 - ECONOMIA E GESTIONE DELLE IMPRESE

LESS IS MORE: HOW VIDEO DESIGN SHAPES THE EFFECTIVENESS OF GREEN-ORIENTED BRAND MESSAGES ON TIKTOK

Presentata da: Cristoforo Losito

Coordinatore Dottorato
Riccardo Fini

Supervisore
Elisa Montaguti

Co-supervisori
Stefania Farace
Annamaria Tuan

Esame finale anno 2026

ABSTRACT

Sustainability has become a key driver in consumption choices. Brands increasingly rely on short-form videos on social media to promote eco-friendly products and services, share their green values, or educate consumers on how to behave sustainably. But while videos are ubiquitous on social media, how video design shapes the effectiveness of brand-generated content is unclear. This work investigates how different video designs, from videos featuring only baseline elements to those enriched with text overlays and voice-over, shape consumer engagement of brand messages on social media, as well as whether, how and why the effect might differ in the context of green- (vs. non-green) oriented communication. A combination of automated video, audio, and text analyses with machine learning techniques on a dataset of thousands of TikTok videos posted by top European companies sheds light on these issues. Results show that, overall, using text overlay, voice-over, or both, on top of the baseline elements (i.e., unfolding visuals and caption) increases social media engagement among non-green posts. However, when brands promote green-oriented content, a richer video design (i.e., implementing text overlay, voice-over, or both) backfires, likely due to skepticism and psychological reactance. Findings provide actionable insights for more effective content design on social media, suggesting that a "less is more" approach is a solution to foster engagement with videos promoting sustainability of products, production processes, behaviors or lifestyles.

Keywords: Multimodality, sustainability, social media, video, brand-generated content, engagement

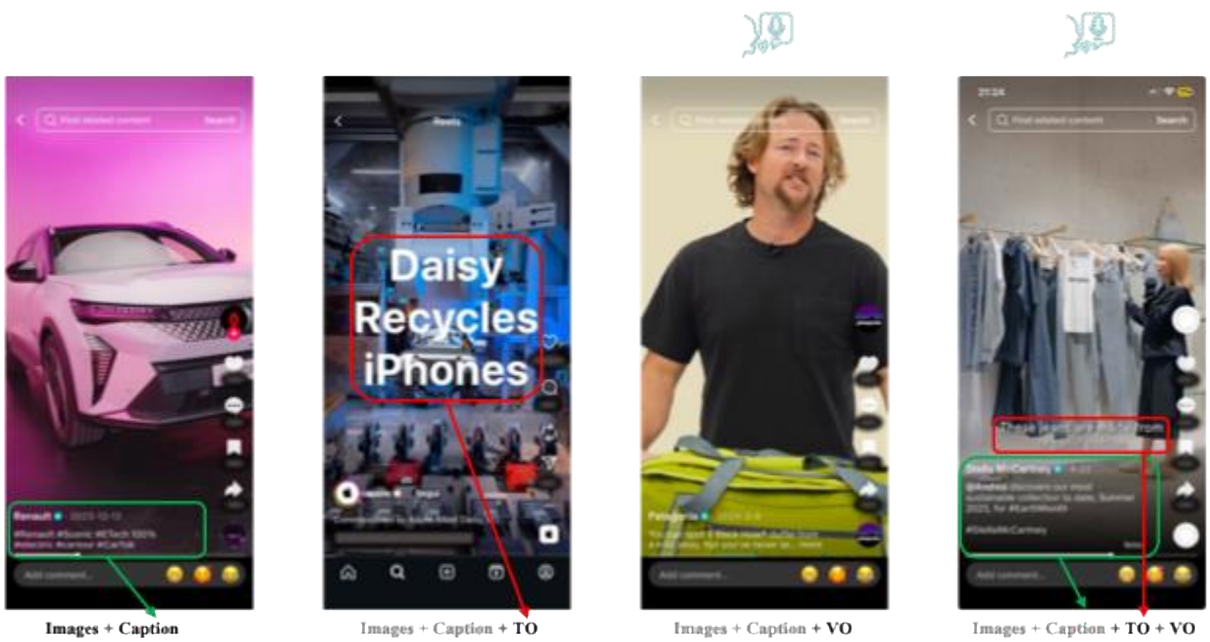
TABLE OF CONTENT

1 INTRODUCTION

Sustainability has become a key driver in consumption choices (White, Habib, and Hardisty 2019), with consumers increasingly looking for eco-friendly products and brands (Harvard Business Review 2023; First Insight 2022). Yet companies often struggle to communicate their environmental efforts effectively, as not just *what* to say, but also *how* to say it, is unclear (NYU CSB and Edelman 2025). The European Commission (2023) recently revealed that 53.3% of environmental claims are perceived as vague, misleading or unfounded, and highlights the importance of using the right "wording, imagery and overall product presentation, including the layout, choice of colors, images, pictures, sounds, symbols or labels" (European Commission 2023, p. 34) in brand messages communicating sustainability-related information.

In this context, mode enriched communication may be perceived as the best venue to convey complex topics or difficult to grasp messages such as those dealing with sustainability (White, Habib, and Hardisty 2019). Videos, as a rich medium, may be particularly effective in communicating green-oriented information, as they may allow businesses to convey more quickly and easily a large amount of information, build narratives, and foster emotional engagement through a blend of verbal and nonverbal cues (Casaburi 2024; The Ground We Walk On 2023). After all, individuals now spend over 17 hours per week watching videos (Forbes 2024a), and 87% of consumers have been persuaded to buy a product or service after watching a video (Wyzowl 2025). Indeed, an increasing number of companies are using particularly short-form videos on social media (e.g., TikTok videos, YouTube Shorts, and Instagram Reels) to promote eco-friendly products and services, share their green values, or educate consumers on how to behave sustainably. Though, these videos vary in their design from a multimodal perspective. Some videos feature only the baseline elements (i.e., visual content and caption), reflecting lower multimodal richness, while others can feature also text overlays, voice-over, or both, resulting in higher multimodal richness (Figure 1). For instance, Renault unveils its news sustainable concept car through merely a video coming on with the textual caption. Apple integrates also text overlays of keywords included in the caption to present Daisy – its recycling robot that disassembles old iPhones. Conversely, Patagonia uses only a voice-over, together with visuals and caption, to promote its recycled duffels. Stella McCartney showcases videos of eco-friendly alternatives to animal leather in videos combining voice-over and text-overlays.

Figure 1. Primary Components and Multimodal Design in Social Media Videos



Notes: Figure 1 presents typical screenshot examples of video design that differ in their richness from a multimodal perspective. The caption refers to the written text accompanying the video and typically appears outside its visual frame or requires users to click a “see more” button to view it fully. Text overlay (TO) refers to any text beyond trademarks or logo superimposed over the video. Voice-over (VO) refers to any voice (on screen or off-screen) that appears in a video, excluding lyrics from songs or background music.

Consistent with its relevance, a growing body of research has begun to examine how multimodal richness in social media posts shape the effectiveness of green-oriented messages

(Barberá-Tomás et al. 2019; Gonzalez-Arcos et al. 2021; Zhu, Cheng, and Wang 2025). For instance, Barberá-Tomás et al. (2019) show that pairing textual discourses with still images in anti-plastic pollution campaigns can help in guiding interpretation and fostering social change. However, recent research also evidenced as an excess of visual and verbal information in the context of user-generated pro-environmental videos can decrease engagement (Zhu, Cheng, and Wang 2025).

Despite the strategic relevance of owned media for brands (Batra and Keller 2016; Kumar et al. 2016; Liadeli, Sotgiu, and Verlegh 2023), and the high consumer demand for brand-generated video content on social media (Wyzowl 2025), videos have received limited attention in marketing research (Grewal et al. 2022; Grewal, Gupta, and Hamilton 2021). This is probably due to the inherently multimodal nature of a medium where visual, auditory, and textual elements co-occur and unfold over time, as well as technical difficulties in extracting consumer insights from highly unstructured data (Wang, Bendle, and Pan 2024). Extant work has tended to consider how textual (e.g., Cascio Rizzo et al. 2024; Zhou et al. 2021), audial (e.g., Chang, Mukherjee and Chattopadhyay 2023; Zhang, Zhou, and Lee 2024), and visual elements (e.g., Yang, Zhang, and Zhang 2024; Wang, Ding, and Hu 2025) *in isolation* affect consumer responses. However, less research has focused on whether and how multimodal cues should be *combined* in video design for effective marketing communication on social media (Holliday et al. 2023; Lapresta-Romero et al. 2024; Rajaram and Manchanda 2025), particularly when the context of sustainability is considered (Zhu, Cheng, and Wang 2025). While prior research in such context focused mostly on how to use visual cues to make advertising look "green" (e.g., Segev, Fernandes, and Hong 2016) *or* how to frame textual information to promote eco-friendly products, brands, and behaviors (e.g., Olsen, Slotegraaf, and Chandukala 2014; Kronrod, Amir, and Grinstein 2012), how the combination of multimodal cues shapes the effectiveness of green messages on social media is less clear (Barberá-Tomás et al. 2019; Gonzalez-Arcos et al. 2021; Zhu, Cheng, and Wang 2025).

Importantly, sustainability should not be viewed merely as a context for theory testing, but rather as a domain with distinct psychological and communicative specificities that may systematically alter how consumers interpret and respond to brand messages (Madan et al. 2023). Indeed, green-oriented communication unfolds within a prosocial and normative context in which skepticism toward corporate claims is often heightened (Leonidou and Skarmeas 2017) and where environmental issues and the consequences of (un)sustainable behaviors are often abstract and difficult to grasp (Meijers et al. 2022; White, Habib, and Hardisty 2019). In such a context, richer multimodal designs may facilitate message elaboration and enhance attitudes and behavioral responses by rendering green messages clearer, vivid and compelling (Meijers et al. 2025; Hartmann and Apaolaza-Ibáñez 2009). However, increasing multimodal richness may also yield unintended negative consequences, such as greater persuasion knowledge (Friestad and Wright 1994) and reactance (Brehm 1966). These domain-specific nuances heighten the need for empirical investigation grounded in real-world data, as the effectiveness of video design choices may hinge on the unique sensitivities and expectations elicited by green content.

Accordingly, this research investigates how different video designs – ranging from videos featuring only baseline elements to those enriched with text overlays and voice-over a) shape consumer engagement of brand messages on social media b) whether, how and why the effect might differ in the context of green (vs. non-green) communication.

We combine automated video, audio, and text analyses with machine learning techniques on a dataset of thousands of TikTok videos shared by top European companies to shed light on these issues. Our findings indicate that, overall, social media videos benefit from the use of text overlay and voice-over. However, when the brand promotes green-oriented messages, implementing a text overlay, voice-over, or a combination of both, on top of the baseline elements (i.e., visuals and caption) backfires.

This work makes several contributions. First, we deepen understanding around multimodal communication (Grewal et al. 2022; Grewal, Gupta, and Hamilton 2021; Packard and Berger 2024). While a growing body of research in video marketing has explored how textual (e.g., Zhou et al. 2021), audial (e.g., Chang, Mukherjee and Chattopadhyay 2023), and visual elements (e.g., Wang, Ding, and Hu 2025) *individually* affect consumer behavior, there has been less attention on their *joint* effect (Lapresta-Romero et al. 2024; Rajaram and Manchanda 2025). We fill this gap by examining how enriching video design with text overlay and voice-over shape consumer engagement of brand-generated content.

Second, we contribute to the literature on green-oriented communication by advancing understanding of multimodal strategies in dynamic content such as videos. Prior work has provided some insights into how brands can simultaneously leverage multiple modes (i.e., image, text) in static content to effectively promote eco-friendly brands, products or lifestyles (e.g., Schmuck, Matthes, and Naderer 2018; Barberá-Tomás et al. 2019), but less attention has been devoted to videos (Zhu, Cheng, and Wang 2025). While this stream of research highlights the positive effects of combining modes (e.g., Barberá-Tomás et al. 2019), our work shows that richer multimodal strategies can backfire in green-oriented videos and explores potential underlying mechanisms behind these effects.

Third, methodologically, we illustrate how combining automated video, audio, and text analysis tools with machine learning tools can deepen insights into consumer behavior (Grewal, Gupta, and Hamilton 2021; Wang, Bendle, and Pan 2024). While prior marketing research on videos has often examined key constructs *within a single mode* (e.g., arousal via text *or* speech; Cascio Rizzo et al. 2024), this work adopts a multimodal perspective and measures several constructs *across modes* (e.g., arousal across caption, text overlays and speech). By doing so, we offer a deeper conceptual and empirical understanding of how these measures can shape consumer behavior. Moreover, whereas most prior studies have applied computer vision to static content (e.g., Dang et al. 2025; Li and Xie 2020), we move beyond and apply it to videos to capture a key construct such as the semantic similarity between visuals and verbal cues, an aspect overlooked so far.

For practice, these results have clear implications. From companies and marketers to institutions and policy makers, everyone wants to improve the effectiveness of their green-oriented communication. Our findings indicate that a "less is more" approach that avoids the implementation of text overlays and voice-over in videos is a solution to avoid lower engagement for videos that promote a green lifestyle or consumption.

2 LITERATURE REVIEW

DIGITAL MULTIMODAL COMMUNICATION

Multimodal Communication: Definition, Properties of Modes, and Modal Interdependence

Multimodality is a way of characterizing communicative situations which rely on the combination of different forms of communication (i.e., modes) to be effective (Bateman, Wildfeuer, and Hiippala 2017). There is general consensus on conceptualizing a mode as a resource for making meaning and interpreting information (Bateman 2022). Modes such as images, written text, spoken language, and sounds can be integrated into a multimodal ensemble to convey richer, more nuanced messages than any single mode could achieve alone (Hull and Nelson 2005).

To understand how each mode can enrich multimodal communication, it is important to acknowledge first its distinct inherent properties – that is, how it structures information and engages the senses (Bateman, Wildfeuer, and Hiippala 2017). Different modes convey information in distinct ways, as they have different logics, potentials and constraints (Kress 2010). For instance, images are processed immediately, based on a principle of spatial and simultaneous organization, while language follows the logic of sequence and time, and it's based on grammar and syntax principles (Meyer et al. 2018; Kress and Van Leeuwen 2006). Noteworthy, despite some commonalities, spoken and written language have also different intrinsic properties that are related to how people encode information through different sensory systems (i.e., sight and hearing, respectively). In other words, beyond *what* is said or written, also *how* is said or written (i.e., paralanguage) plays a role in shaping meaning and might affect consumer responses (Luangrath, Xu, and Wang 2023). For instance, written language conveys meaning also through visual characteristics such as font type, color, and size, while spoken language inherently delivers meaning also through the audial properties of the sound waves (e.g., volume, pitch, tempo; Hildebrand et al. 2020; Van Zant and Berger 2020).

Given these differences among modes, the way individuals perceive, process, and respond to multimodal ensembles may vary depending on which modes are combined (e.g., visuals plus written text or visuals plus spoken language), and how they interplay and complement each other. Indeed, modes are inherently interdependent, as they typically need to be interpreted in relation to one another (Bateman, Wildfeuer, and Hiippala 2017; Bucher 2025). For instance, the visual, verbal and auditory information in a video derive their relevance, and meaning, in part by the context each mode creates for the other (Grewal, Gupta, and Hamilton 2021). Therefore, understanding how modes *jointly* – rather than separately – have an impact on consumer responses becomes crucial.

Despite the pervasiveness of multimodal messages, marketing research has only recently recognized the urgent need to investigate how the interplay among modes (e.g., image, text, audio) shapes consumer behavior, explicitly calling for further research on this topic (Grewal, Gupta, and Hamilton 2021; Grewal et al. 2022; Packard and Berger 2024).

The next paragraph reviews extant marketing research on multimodal communication in social media to examine which types of multimodal ensembles have been most frequently investigated, which specific combinations of visual, textual, and auditory cues have been considered within them, and how these combinations influence consumer responses. By doing so, we aim to identify research gaps and support the research objective of the present work.

Marketing Research on Multimodal Communication in Social Media

The first challenge in conducting a multimodal analysis lies in identifying what characterizes and differentiates various types of multimodal compositions on social media. Building on Bateman, Wildfeuer, and Hiippala's (2017) characterization of multimodal ensembles, social media messages can vary significantly depending on whether the content unfolds over time (i.e., dynamic vs. static content), and whether certain elements remain perceptually accessible or fade away (i.e., transience). From this perspective, short-form videos can be conceptualized as dynamic, audiovisual canvases, where certain elements (e.g., text overlay) may be transient (Grzenkowicz and Wildfeuer 2025). In contrast, image-based posts are static formats that do not typically feature any audio element, where non-transient visuals and text can co-exist and interplay.

Surprisingly, research on how the interplay among multiple elements across modes in dynamic, richer audiovisual content (i.e., videos) shapes consumer engagement on social media has been largely overlooked (See Table 1 for details). Yet, growing research in educational psychology has shown as the interplay among visual, auditory and textual elements in videos can shape learning outcomes (Schmidt-Borcherding et al. 2025; Bateman and Schmidt-Borcherding 2018; Mayer, Fiorella, and Stull 2020). For instance, people's learning improves when subtitles are added to an educational video in the learner's second language, as the written text helps compensate for reduced auditory comprehension (Mayer, Fiorella, and Stull 2020).

More specifically, the interaction between auditory cues (i.e., background music, speech, sounds) and other modes has received very limited attention. *Static* social media posts are not traditionally conceived as audiovisual content, although some platforms now allow users to integrate audio elements in specific formats (e.g., stories). An exception is the work by Lapresta-

Romero et al. (2024), who started exploring how specific multimodal combinations of visual, textual, and auditory cues in Instagram stories can shape social media engagement. Their findings suggest that audio elements can shape engagement differently depending on the content configuration. However, the study does not distinguish, for instance, between the impact of specific audio elements (e.g., speech, music) on engagement when combined with visuals, nor whether their influence varies depending on whether the story is video-based or not.

Concerning the interplay between visual and textual mode, prior research has provided valuable insights into how these elements interplay in *static* multimodal posts (e.g., still images and caption, still images with text overlays) to influence consumer responses. While some studies focus on *what* a message conveys (i.e., Adam 2025; Dang et al. 2025; Li and Xie 2020; Villarroel et al. 2019), other studies focus on *how* it is visually and textually constructed (i.e., Farace et al. 2025; Farace et al. 2020; Kanuri, Hughes, and Hodges 2024). Li and Xie (2020) find that the mere presence of an image in addition to the textual caption increases consumer engagement on Twitter, but not on Instagram. Focusing on image–text complementarity, Villarroel Ordenes et al. (2019) suggest that action-oriented images in combination with an informative textual caption facilitate consumer sharing. In a similar vein, Adam (2025) finds that a greater caption extension beyond the image increases consumer interest and engagement, thereby suggesting that providing additional information through text can enrich the visual message. While these studies examined how still images interplay with textual caption, other research focused on text overlays. Dang et al. (2025) show that posts emphasizing text overlay over visual content tend to get fewer likes and comments, unless the post is text oriented. In this case, the performance improves if text is more centered, informative, emotionally positive, and congruent with the pictorial content, and if the picture contains fewer prominent objects and visual clutter. Farace et al. (2025) investigate how to make a social media post more visually appealing and engaging by balancing the salience of text overlays with images. Specifically, they argue that placing the text overlay peripherally (vs. centrally) on a dynamic image and increasing its size can increase engagement. However, if it becomes too large, it might become visually unappealing, with negative consequences on engagement. However, it remains unclear whether the insights on image-text interplay in static multimodal posts may apply to dynamic audiovisual content such as videos. For instance, in video content, particularly text overlays can appear and disappear in the visual frame, vary in their position, size, and color, and interplay with audio elements such as voice-over or music, potentially shaping how audiences perceive, process, and engage with the message.

The next section reviews extant research on video effectiveness in marketing, focusing on the impact of individual modes and their interplay. In doing so, we aim at identifying which specific multimodal combination of elements in videos have been studied, which one deserves further attention, and why.

Table 1: Overview of Studies on Modes Interplay in Social Media Marketing Communication

Authors	Context	Method	Theoretical Background	Outcomes	Type of Multimodal Post	Modes Interplay	Modality of Interest				Main Findings
							Visual	Verbal		Audio	
								Text	Speech		
Villarroel Ordenes et al. 2019	Facebook, Twitter	Field data	Speech act theory (Searle 1969)	Virality	Static	Semantically (in)congruent pictures and words	✓	✓			Speech–image act complementarity affects consumer engagement. Both modes are independent, but a combination of different types of acts (e.g., informational text and directive image) makes the post more novel or surprising.
Li and Xie 2020	Twitter, Instagram	Field data	Visual advertising literature (e.g., Pieters and Wedel 2004)	Attention, retweets, likes	Static	Semantically relevant image for what is written in the caption	✓	✓			Image–text fit effects are context and platform specific. Specifically, image–text fit has a significant and positive effect on engagement for airline-related posts on Twitter but not on Instagram.
Farace et al. 2020	Twitter	Field data, Lab experiments, text and image analysis	Visual semiotics theory (Kress and van Leeuwen 2006)	Product evaluations	Static	Match between visual patterns in images and headlines (or caption)	✓	✓			Stronger product evaluations yield if regular visual patterns in ads images combines with verbal information conveying motion. A regular pattern creates

						conveying motion.					mental simulation, such that consumers imagine themselves experiencing the product.
Holiday et al. 2023	Instagram	Field data, facial expression analysis, computational textual analysis	Social exchange theory (e.g., Hayes, King and Ramirez Jr 2016), emotional contagion theory (Hatfield, Cacioppo and Rapson 1993)	Consumer brand engagement (likes, comments, views)	Dynamic	Emotions conveyed through visuals and text	✓	✓			Influencers’ facial (negative, not positive) emotion expressiveness drives consumer engagement, whereas textual emotion expressiveness has little impact. Branding fundamentally changes emotion’s impact.
Lapresta-Romero et al. 2024	Instagram	Fuzzy-set qualitative comparative analysis	Multimodality literature (e.g., Kress 2010), dual coding theory (Paivio 1978)	Social media engagement behaviors	Static and dynamic	Combinations of visual, auditory, linguistic, symbolic, social, and emotional elements	✓	✓	✓	✓	Specific content configurations (loud, affective, informative, relational) of elements across modes drive social media engagement.
Kanuri, Hughes, and Hodges 2024	Facebook	Field data, eye tracking	Elaboration likelihood model of persuasion (Petty and Cacioppo 1984)	Total engagement (clicks, likes, comments, reactions, and shares)	Static	Visual and textual complexity	✓	✓			Social media posts with more complex images draw more user attention, which then increases the likelihood of user engagement. Images with a high color complexity accompanied by complex text induce greater attention and engagement due to its ability to make a post seem more intriguing and distinctive.
Farace et al. 2025	Twitter, Instagram	Field data, managerial interviews, experiments	Multimodal aesthetics (Hagtvedt 2022), advertising literature (Pieters	Social media engagement (likes, comments,	Static	Image dynamism, text overlay size and centrality	✓	✓			A text overlay that is too large and placed centrally, when combined with dynamic images,

			and Wedel (2004)	retweets, favorites)							negatively affect social media engagement.
Adam 2025	Twitter	Field data, experiment	Multimodality literature (Heckler and Childers 1992; Houston, Childers, and Heckler 1987)	Social media engagement (likes, comments, shares)	Static	Caption extension beyond the image	✓	✓			Greater caption extension beyond the image increases consumer interest, which in turn enhances engagement.
Dang et al. 2025	Facebook, Instagram	Field data		Social media engagement (like, comment)	Static	Text overlay and pictorial content	✓	✓			Posts that emphasize text overlay over visual content tend to get fewer likes and comments. The performance of text-oriented posts improves if text is more centered, informative, emotionally positive, and congruent with the pictorial content, and if the picture contains fewer prominent objects and visual clutter.
Rajaram and Manchanda 2025	YouTube	Field data, interpretable deep learning	Influencer marketing literature (e.g., Leung et al. 2022), Literature on unstructured data via deep learning (e.g., Tian, Dew, and Iyengar 2024)	Verbal (i.e., comments) and non-verbal (like/dislike ratio) engagement, sentiment of engagement	Dynamic	Visual (i.e., face emotions, size of humans, packaged goods, and animals, brand logos, everyday objects), auditory (music duration, speech pace) and textual (brand mentions, presence of emotional words) features	✓	✓	✓	✓	*What* is said through words is more important than *how* it is said through imagery or music in predicting video engagement. Brand mentions and music in the beginning of a video are associated with sentiment expressed in comments, whereas visual stimuli (e.g., human faces, packaged goods) are linked with sentiment expressed through non-verbal engagement (like/dislike ratio).

Marketing Research on Video Effectiveness

Marketing research only recently started to investigate how textual (e.g., Cascio Rizzo et al. 2024; Zhou et al. 2021), audial (e.g., Chang et al. 2023; Zhang, Zhou, and Lee 2024), and visual features (e.g., Yang, Zhang, and Zhang 2024; Wang, Ding, and Hu 2025) in videos *individually* affect consumer responses (see Schraml 2025 for a comprehensive review). Studies have focused on a variety of topics, such as video advertising effectiveness (e.g., Teixeira, Wedel, and Pieters 2012; Stuppy et al. 2024), influencer marketing (e.g., Cascio Rizzo et al. 2023; Tian, Dew, and Iyengar 2024), online learning (Zhou et al. 2021), and livestreaming (Lin, Yao, and Chen 2021; Lu et al. 2021). For instance, in predicting consumption of educational videos (i.e., how long a video is watched), Zhou et al. (2021) find that positive sentiment in words used by an instructor increases video consumption. Conversely, Wang, Ding, and Hu (2025) focused on visual features, and found that face presence in user-generated videos positively influences consumer engagement, particularly when accounts for the 30-40% of frames. Meanwhile, other work looked at vocal cues: while Chang, Mukherjee, and Chattopadhyay (2023) found evidence that more narrating voices can facilitate persuasion, Zhang, Zhou, and Lee (2024) revealed that using an AI voice negatively affects engagement for those creators that are experienced and have higher identity disclosure level. Cascio Rizzo et al. (2024) reveals that the arousing effect of influencers' speech pitch in TikTok videos as a vocal cue potentially expressing arousal, showing that higher pitch (i.e., higher arousal) decreases engagement for macro-influencers but not for micro-influencers.

To the best of our knowledge, only few studies adopt a multimodal approach in examining how elements across modes within videos interact with each other in shaping consumer responses. Al-Qershi et al. (2022) focus on the complementarity among specific multimodal cues (visuals and speech). In line with research on the interplay between still images and written text (Li and Xie 2020; Villarroel et al. 2019), they found that complementary cues across modes can facilitate campaign success. More specifically, speech aids (i.e., words like “see”, “compare”) that complement visuals, along with a positive tone, can enhance the effectiveness of the campaign. Holiday et al. (2023) focus instead on the differential impact of emotions conveyed through multiple modes. In particular, they found that influencers’ negative emotions (i.e., anger, sadness) expressed through the face consistently drive consumer engagement, whereas emotions expressed through textual stimuli – whether positive or negative – have little impact. Such evidence is in line with the image superiority effect (Childers and Houston 1984) where images have advantages over words with regards to coding and retrieval of stored memory. In fact, pictures are processed and retrieved in memory more easily compared to text. Finally, Rajaram and Manchanda (2025) aim to document the relative importance of different modes in long-form influencer videos shared on social media, as well as their impact on different engagement measures. They found that *what* is said through words (i.e., caption, transcripts) is more influential than *how* it is said through imagery or acoustics (e.g., music, sounds) for predicting video engagement. In doing so, they identify and distinguish between theory-based features associated with verbal and non-verbal engagement measures. In particular, brand mentions and music duration in the beginning of a video are associated respectively with a decrease and increase of sentiment expressed in comments. In other words, viewers tend to express their sentiment toward auditory stimuli (e.g., brand mentions and music) through verbal engagement in comments. In contrast, visual stimuli (i.e., an increase in size of human faces or packaged goods) are linked with sentiment expressed through non-verbal engagement (like/dislike ratio). This is in line with prior research on sensory marketing suggesting that modes can drive different emotional and cognitive responses depending on the sensory system involved (Krishna 2012).

GREEN-ORIENTED MULTIMODAL COMMUNICATION

Brand communication can be qualified as green-oriented if meets one or more of the following criteria: (1) explicitly or implicitly addresses the relationship between a product/service and the biophysical environment, (2) promotes a green lifestyle with or without highlighting a product/service, and (3) presents a corporate image of environmental responsibility” (Banerjee, Gulas, and Iyer 1995, p. 22).

To understand how multimodality in videos can be leveraged to improve the effectiveness of green-oriented communication on social media, we reviewed extant work on how modes (i.e., text, images, speech, sounds) have been used – separately or jointly – to promote brands, eco-friendly products and services, or a green lifestyle (see Table 2 for details).

Extant work has studied how to effectively communicate sustainability in a variety of contexts, such as websites (e.g., Fennis et al. 2011; Parguel, Benoit-Moreau, and Larceneux 2011; Salnikova, Strizhakova, and Coulter 2022), social media (e.g., Kronrod et al. 2023; Mookerjee, Cornil, and Hoegg 2021; Barberá-Tomás et al. 2019; Gonzalez-Arcos et al. 2021; Zhu, Cheng, and Wang 2025), and advertising in particular (e.g., Chernev and Blair 2020; Gershoff and Frels 2015; Peloza, White, and Shang 2013; Royne et al. 2012; Winterich, Nenkov, and Gonzalez 2019).

Research focused mostly on verbal communication, and particularly on how textual messages can influence a variety of outcomes, such as the intention to behave sustainably (e.g., White, MacDonnell, and Dahl 2011; Zhang et al. 2023), the attitudes toward the brand (e.g., Olsen, Slotegraaf, and Chandukala 2014; Schmuck, Matthes, and Naderer 2018), and the volume of sales (Nickerson et al. 2022). Some of these studies focused on content (i.e., *what* is said; Goldstein, Cialdini, and Griskevicius 2008; Kronrod et al. 2023; Luchs et al. 2010; Newman, Gorlin, and Dhar 2014; Nickerson et al. 2022). For instance, Kronrod et al. (2023) find that combining discouraging (e.g., "don't close your eyes to our world") and encouraging messages (e.g., "include cultured meat in your meal") is more likely to foster sustainable behaviors that are novel to the target audience (e.g., lab-grown food). Atkinson and Rosenthal (2014) focus on the amount of environmental information, suggesting that specific (vs. general) green messages positively affect trust and attitudes toward the product. From this perspective, research suggested that even a subtle cue (e.g., a single word) can make a difference in shaping consumer responses. For instance, increasing the salience of product longevity by simply mentioning the word "durable" can encourage consumers to favor fewer higher quality products over multiple midrange-quality ones (Sun, Bellezza, and Paharia 2021). Similarly, drawing attention to the aesthetic flaws of products through "ugly" labeling can increase choice likelihood and purchases of unattractive products because it corrects for consumers' biased expectations on tastiness (Mookerjee, Cornil, and Hoegg 2021). Other work looked at *how* it is said, and to *whom* (e.g., Amatulli et al. 2019; Fennis et al. 2011; Kronrod, Grinstein and Wathieu 2012; White, MacDonnell, and Dahl 2011). On the other hand, Kronrod, Grinstein, and Wathieu (2012) show that environmental messages should be less assertive when targeting less environmentally concerned consumers to avoid reactance, as they perceive the topic as less important.

While the abovementioned stream of research focused on how to use verbal information to make green-oriented communication more effective, other work focused on how visual and auditory cues (e.g., birdsong, sea sound) can make the advertising look "green" (Parguel, Benoit-

Moreau, and Russell 2015; Segev, Fernandes and Hong 2016). Visual cues typically involve the use of nature-related colors (e.g., green, blue) and nature-evoking images, such as pictures involving mountains, forests, trees, and renewables sources of energy (e.g., sun, wind). Among colors, the color green in particular has been frequently adopted as a marketing tool for environmentally conscious consumption (Labrecque, Patrick, and Milne 2013). For instance, Pancer, McShane, and Noseworthy (2017) evidence that the green color in packaging, when combined with an eco-label, reinforce environmental associations and foster purchase intention of eco-friendly products. Similarly, research shows that consumers' exposure to nature-evoking images in green brand communications may lead to pleasant feelings similar to those experienced when in contact with the nature (i.e., "virtual nature experiences"), which in turn may affect positively brand attitude (Hartmann and Apaolaza-Ibánez 2009) and purchase intention (Peloza, White, and Shang 2013).

To sum up, research has demonstrated that textual, visual, and auditory cues can all play a role in influencing the effectiveness of green-oriented communication. However, only a few studies have examined how elements across modes *jointly* shape consumer responses. For instance, De Visser-Amundson, Peloza, and Kleijnen (2021) show that, when promoting rescue-based food the combination of multiple environmental cues across modes (i.e., verbal descriptions paired with green color) can reduce consumer demand by activating waste associations. Conversely, Schmuck, Matthes, and Naderer (2018) found that combining nature-evoking images with text can elicits a stronger emotional (cf. rational) response, with in turn positive consequences on the evaluation of the advertised products and brand. On social media, combining visual and verbal information in social media posts can effectively foster support for environmental causes by enhancing message clarity and emotional resonance (Barberá-Tomás et al. 2019; Gonzalez-Arcos et al. 2021; Zhu, Cheng, and Wang 2025). However, Zhu, Cheng, and Wang (2025) show that, in the context of pro-environmental user-generated videos, an excess of visual and verbal informativeness can decrease engagement.

The next section presents the conceptual development, which focuses on a) how different video designs – ranging from videos featuring only baseline elements to those enriched with text overlays and voiceovers – might shape consumer engagement of brand messages on social media b) whether, how and why the effect might differ in the context of green (vs. non-green) communication.

Table 2. Overview of Studies on the Effectiveness of Communication on Sustainability

Authors	Context	Method	Theoretical Background	Outcomes	Modes Interplay	Modality of Interest				Main Findings
						Visual	Verbal		Audio	
							Text	Speech		
Goldstein, Cialdini, and Griskevicius 2008	Print signs	Field experiments	Social comparison theory (Festinger 1954)	Reuse behavior			✓			Green messages integrating descriptive norms are more effective than traditional messages that focus merely on environmental protection.
Hartmann and Apaolaza-Ibánez 2009	Print ads	Experimental field study	Elaboration likelihood model (Petty and Cacioppo 1983)	Brand attitude	Natural imagery and textual information	✓	✓			Information on specific environmental product features increase brand attitude. Consumers' exposure to nature-evoking images in green brand communications may lead to pleasant feelings similar to those experienced when in contact with the nature (i.e., "virtual nature experiences"), which in turn may enhance brand attitude.
Luchs et al. 2010	Print ads	Implicit association test, field study, lab experiments	Literature on lay theories (e.g., Broniarczyk and Alba 1994; Raghunathan, Naylor, and Hoyer 2006), CSR (e.g., Luo and Bhattacharya 2006; Sen and Bhattacharya 2001)	Consumer preferences, Consumer perception of product ethicality			✓			Green messages providing explicit information about product strength counteract sustainability liability, as consumer typically perceive sustainable products as having lower functional performance.
Fennis et al. 2011	Website communication	Field and lab experiments	Literature on implementation intentions (e.g., Gollwitzer 1999), mental imagery (e.g., Keller and Block 1997;	Purchase behavior			✓			More (vs. less) vivid persuasive appeals (i.e., a narrative with more concrete wording) increase actual sustainable buying behavior.

			MacInnis and Price 1987)					
Parguel, Benoit-Moreau, and Larceneux 2011	Website design	Experiments	Attribution theory (e.g., Heider 1944; Kelley 1973)	Perceived corporate brand equity			✓	Poor independent sustainability ratings negatively affect corporate brand evaluations because consumers infer less intrinsic motives by the brand.
White, MacDonnell, and Dahl 2011	Print ads	Field and lab experiments	Construal level theory (Libermann and Trope 1998)	Recycling intention and behavior			✓	Verbal messages framed as losses (vs. gains) are more effective when combined with concrete information on how (vs why) to engage in a sustainable behavior.
Kronrod, Grinstein and Wathieu 2012	Google ads	Field and lab experiments	Psycholinguistics (e.g., Dillard and Shen 2005), communication literature on compliance (e.g., Edwards, Li, and Lee 2002)	Environmental compliance intention, click-through rate			✓	Individuals respond better to assertive (vs. non-assertive) language in green messages when perceive the issue as important.
Lin and Chang 2012	On-pack labeling, print ads	Field study, in-field experiments	Lay theories (e.g., (Broniarczyk and Alba 1994; Sujan and Dekleva 1987)	Product usage	–	✓	✓	Consumers perceive eco-friendly (vs. regular) product as less effective, particularly when more (vs. less) environmentally conscious. Therefore, green messages should explicitly label and endorse both product's effectiveness and sustainability to avoid overuse.
Royne et al. 2012	Print ads		Prospect Theory (Kahneman and Tversky 1979), accounting theory (Thaler 1985)	Perceived quality, perceived price			✓	Consumers perceive products advertised with environmental appeals as more costly, but not of lower quality if compared with products advertised with personal benefits.
Kidwell, Farmer, and Hardesty 2013	Print ads	Field and lab experiments	Literature on political ideology (e.g., Jost 2006), moral foundations (e.g., Graham et al. 2009; McAdams et al. 2008), and fluency (e.g., Reber,	Recycling behavior, purchase intent, water conservation intent			✓	Congruency between political ideology and appeal type positively affect sustainable behaviors. Communication appealing to "binding moral values" (e.g., duty, in-group affiliation) leads to more positive sustainable behaviors among Republicans, whereas appealing to "individualizing moral values" (e.g., fairness, individuality)

			Schwarz, and Winkielman 2004)					leads to more positive reactions among Democrats.
Peloza, White, and Shang 2013	Print ads	Field and lab experiments	Self-consistency (e.g., Stone and Cooper 2001; Thibodeau and Aronson 1992) and discrepancy theories (e.g., Higgins 1987), reactance theory (Brehm 1966; Brehm and Brehm 1981)	Purchase intention	Nature-evoking images, green color and product placement in the ads with textual information	✓	✓	Consumers in the private (vs. public) contexts prefer products promoted using a self-benefit (vs. ethical) appeal. In the private contexts, communications can increase preference for ethical products by subtly making people aware about the discrepancy between their actions and their ought standards, or by directly priming self-accountability.
White and Simpson 2013	Print ads	Field and lab experiments	Social identity theory (e.g., Brewer 1991; Tajfel and Turner 1986). Self-construal theory (e.g., Singelis 1994) also points to the n	Waste disposal behaviors			✓	Communication should ensure a match in terms of goal compatibility. When the collective (vs. individual) level of self is activated, injunctive and descriptive normative (vs. benefit) appeals are most effective in encouraging sustainable behaviors.
Atkinson and Rosenthal 2014	Eco-labels	Online experiments	Signaling Theory (Spence 1973; Erdem and Swait 1998)	Consumer trust, purchase intent			✓	Specific (vs. general) arguments positively affect trust and attitudes toward the product and the source, but not purchase intent.
White, Simpson, and Argo 2014	Press releases	Field study, experiments	Self-affirmation theory (e.g., Steele 1988)	water conservation, composting organics, and recycling			✓	Communicating a comparatively positive performance in sustainable behaviors of a dissociative reference group can lead consumers to increase their own positive behaviors, particularly in public (vs. private) settings.
Newman, Gorlin, and Dhar 2014	Print ads	Online experiments	Literature on intentionality (e.g., Heider 1958), lay theories of resource allocation (e.g., Chernev 2007;	Purchase intent			✓	Communicating green enhancements of a product as intended (vs. unintended) lead consumers to assume that the company diverted resources away from product quality, which in turns decrease purchase intent.

			Chernev and Carpenter 2001)							
Olsen, Slotegraaf, and Chandukala 2014	Overall communication	Field data	Social identity theory (e.g., Bhattacharya and Sen 2003; Du, Bhattacharya, and Sen 2007), framing theory (e.g., Entman 1993; Levin, Schneider, and Gaeth 1998)	Brand attitude			✓			Green new product innovations (GNPIs) can improve brand attitude. The quantity (but not the valence) of green claims influences brand attitude: fewer green claims positively affect brand attitude toward GNPIs. Higher-vice-related products tend to benefit more from GNPIs in terms of changing brand attitude.
Gershoff and Frels 2015	Print ads	Online experiments	Attribute centrality theory (e.g., Sloman, Love, and Ahn 1998)	Green product evaluation			✓			The degree of centrality of green attributes influences product evaluations. Messages emphasizing that a green attribute is central to the definition of the product may lead consumers to judge the entire product as green.
Parguel, Benoit-Moreau, and Russell 2015	Website design	Online experiments	Elaboration likelihood model (Petty and Cacioppo 1981)	Brand attitude, attitude toward the webpage	Visual elements (i.e., picture of a forest, green color in a tinted area) and sounds (i.e., birdsong)	✓	✓		✓	Executional elements evoking nature (i.e., picture of a forest, green color in a tinted area, birdsong) enhance brand environmental image and in turn attitude toward the brand. Adding environmental performance information can counterbalance the effect particularly among consumers with more topic knowledge.
Pancer, McShane, and Noseworthy 2017	Packaging design	Online experiments	Schema congruity theory (Mandler 1982; Meyers-Levy and Tybout 1989; Noseworthy et al. 2014)	Purchase intention	Color green and eco-labels	✓				Multiple environmental visual cues used in isolation (i.e., green color without an environmenta label or an environmental label without green color) reduce perceptions of product efficacy and, in turn, purchase intention.
Edinger-Schons et al. 2018	Online ads	Field and online experiments	Literature on consumer attributions about marketing CSR actions (e.g., Becker-Olsen, Cudmore, and Hill	Purchase intention			✓			Joint appeals (i.e., an intrinsic and extrinsic appeal together) decrease consumers' purchase intention of sustainable products. However, the effect does not hold for consumers with lower sustainable consumption involvement, who are more likely to

			2006; Sen and Bhattacharya 2001), persuasion theories (Fishbein 1979; Petty and Cacioppo 1986)					rely on the number of appeals as a cue for persuasion.
Schmuck, Matthes, and Naderer 2018	Print ads	Online experiments	Affect–reason–involvement (ARI) model (Buck et al. 2004; Buck and Chaudhuri 1994)	Brand attitude	Nature-evoking imagery and textual information	✓	✓	False verbal claims enhance perceived greenwashing, while vague verbal claims do not. Nature-evoking imagery in combination with vague or false claims elicits an affective mechanism that is stronger than rational greenwashing perceptions and in turn enhance consumers' evaluations of ads and brands.
Amatulli et al. 2019	digital marketing campaigns (e.g., videoclips)	In-field and online experiments	Cognitive theory of emotions (e.g., Lazarus 1991; Frijda 1986)	Donation for a pro-environmental cause, environmentally friendly choice			✓	Negatively (vs. positively) framed messages are more effective in leading people to engage in pro-environmental behaviors to reestablish a positive view of themselves.
Winterich, Nenkov, and Gonzalez 2019	Print ads, Google ads, posters	Field and online experiments	Literature on recycling (e.g., White, Macdonnell and Dahl 2011) and inspiration (Böttger et al. 2017; Thrash and Elliot 2003, 2004)	Disposal behavior, recycling intention, click-through rate, recycling rate, Recoverable fraction of landfill waste	Visual graphics combined with written text	✓	✓	Making consumers more aware of the transformation of recyclables into new products increases recycling behaviors by making them more inspired.
Orazi and Chan 2020	Print ads	Experiments	Literature on CSR Communication	Brand attitudes, purchase intention		✓		When specific (vs. vague) claims are countered by specific (vs. vague) external information, consumers report more negative brand attitudes and lower purchase intentions.
Chernev and Blair 2020	Print ads	Experiments	Moral cognition and moral reasoning theories (e.g., Baron 1993; Campbell and Winterich 2018; Greene 2013; Haidt, 2001), literature on sustainability-liability effect (e.g.,	Perceive product performance			✓	Messages framing sustainability as a company activity (vs. a product feature) can increase the perceived performance of the product, thus counteracting the sustainability-liability effect.

			Li and Chang 2010, Luchs et al. 2010)							
Sun, Bellezza and Paharia 2021	Online communication	Field data, online experiments, conjoint surveys	Accessibility-diagnosticity model (Lynch et al. 2015) and the scope insensitivity bias (Chang and Pham 2018)	Disposal behaviors, choice (i.e., one high-end product vs. multiple midrange products)			✓			Increasing the salience of product durability by mentioning the word “durable” is effective in encouraging the sustainable consumption of fewer high-quality products (vs. multiple midrange products).
Gonzalez-Arcos et al. 2021	News media, social media posts, sustainability reports	Archival data, interviews, ethnography	theory of consumer resistance in social practice change	Sustainable behaviors	Textual description combined with images	✓	✓			Multimodal content can emphasize the scope of the intervention more clearly and leverage emotions to reduce consumer resistance to embrace environmental practices.
Nickerson et al. 2022	CSR press releases	Field data, online experiments	Sociopsychological theory (Carlisle et al. 2012), image restoration theory (Benoit 1997)	Brand sales, purchase intention			✓			CSR initiatives that genuinely aim to reduce a brand’s negative externalities (i.e., “corrective” or “compensating”) increase sales, whereas CSR actions focused on philanthropy (i.e., “cultivating goodwill”) can undermine sales. Consumers may perceive philanthropic actions as less sincere.
Mookerjee, Cornil, and Hoegg 2021	Facebook ads, online ads	Field and online experiments	Literature on unattractive produce (e.g., Grewal et al. 2019) and two-sided persuasion (e.g., Ein-Gar, Shiv, and Tormala 2011)	Click-through rate, purchase likelihood, choice			✓			Messages drawing attention to the aesthetic flaw of unattractive produce via “ugly” labeling increase purchases, choice likelihood, and click-throughs.
de Visser-Amundson, Peloza, and Kleijnen 2021	Product promotions	Field experiment, lab and online experiments, online survey	Literature on: mental imagery (e.g., Jia et al. 2017; Wyer, Hung, and Jiang 2008; Zhao, Hoeffler, and Dahl 2009), color effect on preferences (e.g., Labrecque, Patrick, and Milne 2013),	Consumer thoughts, purchase intention, attitude, product choice	Green color and textual description	✓	✓			When an environmental benefit appeal promotes rescue-based food (RBF) with multiple (vs. single) environmental cues (e.g., verbal descriptions paired with green color), consumer demand decreases (vs. increases).

Salnikova, Strizhakova, and Coulter 2022	Website statements	Field data, experiments	Regulatory focus theory (Higgins 1998), Construal level theory (Trope and Liberman 2010), theory of framing effects (Tversky and Kahneman, 1981)	consumer engagement with the environmental sustainability initiatives		✓	Consumers with a strong global (vs. local) identity are more engaged with environmental sustainability initiatives when the communication adopts a framing congruent with their global identity (i.e., promotion-focused, distant spatial, and proximal temporal framing).
Kronrod et al. 2023	Facebook ads, print ads, posting signs	Field and online experiments, field survey	Sustainable education literature (e.g., Gifford 2011; White, Habib and Hardisty 2019)	New pro-environmental behaviors		✓	Combining encouraging and discouraging messages is more effective in driving pro-environmental behaviors that are novel to the target audience because it enhances perceptions of learning and informativeness.
Zhang et al. 2023	Print ads	Online experiments	Elaboration likelihood model (Petty and Cacioppo 1986), construal level theory (e.g., Trope and Liberman 2010), information processing theory (Reber, Schwarz and Winkielman 2004)	Green purchase intention	✓		For functional (vs. emotional) appeals, positive (vs. negative) facial expressions on utilitarian (vs. hedonic) products are more likely to increase green purchase intention.
Xu et al. 2023	Poster ads	Field and online experiments	Schema theory (Fiske and Taylor 1991), self-congruency theory (Sirgy 1982)	Click-through rate, drop-off behavior, product choice, package recycling, index of purchase preferences for eco-friendly products, purchase intention		✓	consumers with a cyclical (vs. linear) temporal perspective are more likely to purchase a green product when the promotional appeal is congruent with their temporal perspective.
Yan, Keh, and Murray 2023	Posters, online ads	Field and online experiments	Appraisal-tendency framework (Han, Lerner and Keltner 2007).	Recycling intention and behavior, reuse intention, sign-up behavior for a plastic-free campaign, purchase intention		✓	Pride (awe) appeals in green ads increase sustainable behavior and intentions when the self-enhancement (self-trascendence) value is emphasized.

3 CONCEPTUAL DEVELOPMENT

How Video Design Shapes Consumer Engagement

The power of video content lies in its inherent richness. According to media richness theory, the effectiveness of a communication format depends on its *information richness* – that is, its capacity to enhance understanding by providing specific information within a certain time interval while minimizing the potential for ambiguous interpretations (Daft and Lengel 1986).

Videos are inherently rich media, as they can simultaneously integrate visual, textual, and auditory cues to enhance understanding (Moffett, Folse, and Palmatier 2021). However, information richness can vary significantly even within the same format, depending on the number of cues and the specific mode through which such cues are delivered (Grewal, Gupta, and Hamilton 2021). In this regard, videos on social media differ considerably in their design, thereby resulting in different levels of richness from a multimodal perspective. Some videos on social media rely solely on baseline elements (i.e., visual content and text caption), thus reflecting lower richness, while others incorporate additional verbal cues – such as text overlays, voice-over, or both – yielding higher richness.

Verbal cues embedded in richer media play a central role in persuasion and marketing communication (Orazi et al. 2025). Recent empirical evidence in the context of social media videos further supports this view, showing that *what is said* in words is more influential than *how it is said* through imagery or acoustics (i.e., music, sounds) in capturing variation in engagement (Rajaram and Manchanda 2025). On social media, where attention is scarce (Forbes 2024b) and content is often quickly scrolled (Wooley and Sharif 2022; Tam and Inzlicht 2024), text overlays and voice-overs can direct viewers' focus to the message's key arguments and making them easier to follow as the visual content unfolds. Moreover, these verbal elements can help shape the meaning of visuals – whose interpretation can be open-ended (Kress and Van Leeuwen 2006; Meyer et al. 2018) – thereby enhancing understanding and, consequently, engagement. Building on these premises, we next elaborate on how text overlays and voice-over can shape – separately and jointly – social media engagement with videos.

Enriching Video Design with Text Overlay

Text overlays – unlike caption – are embedded within the visual frame of the video and can, for instance, appear and disappear in the video frame, vary in their position, font type, size, and color. Therefore, brands can use text overlays to visually capture the attention and emphasize key messages (Dang et al. 2025; Farace et al. 2025).

Prior research on image–text interplay in marketing shows that adding textual elements to visual content can shape meaning, improve message comprehension, and enhance product evaluations and engagement on social media (Adam 2025; Dang et al. 2025; Farace et al. 2020; Holiday et al. 2023; Li and Xie 2020; Kanuri, Hughes, and Hodges 2024; Villarroel et al. 2019). For instance, Adam (2025) shows that providing additional information through textual caption can enrich the visual message and foster consumer interest and engagement. Similarly, Villarroel Ordenes et al. (2019) suggest that the semantic complementarity between images and text can boosts engagement. More specifically, action-oriented images are more effective in stimulating consumer sharing of brand messages when accompanied by informational or emotional content instead of direct calls to action. These studies, although focused on off-frame text in static social media posts, converge on the idea that textual elements can enrich visuals by guiding interpretation and emphasizing the key messages, with in turn positive consequences on engagement. Building on it, we might expect that in-frame text overlays could be even more effective in enhancing understanding in more dynamic audiovisual content such as videos, given their spatial contiguity and the potential for temporal alignment with the visuals.

Insights from educational psychology offer a complementary perspective on why text overlays may enhance the effectiveness of brand video communication on social media. According to cognitive theory of multimedia learning (CTML; Mayer 2021, 2024) meaningful learning occurs when individuals actively *select* relevant information from graphics (e.g., animation, video, illustrations, photos) and words (written text or speech) to be processed in working memory, mentally *organize* the material into coherent verbal and visual structures, and *integrate* them with each other and with relevant knowledge retrieved from long-term memory. Within this framework, the spatial contiguity principle posits that people learn more effectively when corresponding words and graphics are presented close together in space while the temporal contiguity principle holds that learning improves when they are presented simultaneously rather than successively (Mayer, Fiorella, and Stull 2020; Mayer and Moreno 2002). Consistent with these principles, text overlays in videos – by appearing directly within the visual frame and in sync with the unfolding imagery – can enhance understanding by making the video easier to process, thereby increasing the likelihood of engagement. To sum up, in-frame overlays in social media videos can help viewers track the core arguments as imagery unfolds, improving comprehension. This, in turn, fosters more favorable engagement behaviors. Formally:

H_1: Individuals will express lower social media engagement for videos that present only visual content and a text caption, compared to posts embedding visuals, caption and text overlay.

We next examine the role of voice-over, which – while operating through a different sensory channel (i.e., auditory) – can likewise facilitate message comprehension and triggers engagement behaviors.

Enriching Video Design with Voice-Over

Unlike text overlays, which are processed through the visual sensory system, voice-over engages the auditory channel, thereby providing another pathway through which consumers can process brand messages (Tavassoli and Lee 2003). Prior research shows that humans are particularly sensitive to vocal stimuli among many competing sensory signals (Horowitz 2012; Belin, Fecteau, and Bédard 2004). Hearing human voices quickly draws attention and triggers early and deeper cognitive elaboration (Charest et al. 2009; Rutten et al. 2019).

Prior research on spoken language in video marketing has begun to examine how voice contributes to persuasion (Al-Qershi et al. 2022; Chang, Mukherjee, and Chattopadhyay 2023; Wang et al. 2021; Wang, Zhang, and Jiang 2024). Voice-over in short video ads has been shown to increase purchase intention by lowering cognitive load, particularly when they are not AI-generated (Wang, Zhang, and Jiang 2024). Speech can indeed complement visuals by directing attention to the core aspects of the message (e.g., through words like "see", "compare"; Al-Qershi et al. 2022) or by fostering mental simulation of the advertised product (Ceylan, Diehl, and Wood 2024). In line with these findings, research in the context of social media shows that "load" configurations – that is, multimodal posts characterized by the presence of visual elements and speech without text – in firm-generated Instagram stories can increase social media engagement (Lapresta-Romero et al. 2024).

Insights from educational psychology offer a complementary perspective on the role of voice-over in facilitating comprehension and learning, while also highlighting its differential impact if compared with the integration of on-screen text (Mayer 2021). According to CTML, pairing human narration with visuals engages complementary sensory channels, thereby facilitating the selection, organization, and integration of information in working memory (Mayer 2024). Moreover, people tend to achieve deeper understanding when learning from a combination of visual content and auditory narration, compared to visual content paired with on-screen text (i.e., modality principle; Mayer 2024). This principle builds upon the literature on dual coding (Paivio 1991) and cognitive overload (Kirschner 2002): given that human cognitive capacity is limited, presenting both in-frame written words and moving images can overload the visual channel, as written text and imagery are initially processed through the same sensory system.

In a nutshell, we expect that videos combining visuals with words – whether written or spoken – positively affect engagement, with voice-over proving more effective than overlays, as it engages the auditory channel and avoids overloading the visual system. Formally:

H_{2a}: Individuals will express lower social media engagement for videos that present only visual content and a text caption, compared to posts embedding visuals, caption and voice over.

H_{2b}: Individuals will express lower social media engagement for videos that present visual content, text caption, and text overlay, compared to posts embedding visuals, text caption, and voice-over.

When Text Overlay Meets Voice-Over: A Richer Multimodal Design

Companies can also integrate text overlay and voice-over *simultaneously* in videos to enhance their richness and increase consumers' opportunities to process brand information (MacInnis, Moorman, and Jaworski 1991). By embedding text directly within the visual frame while concurrently presenting spoken narration, videos leverage complementary sensory channel (i.e., visual and auditory) to deliver information. Prior research in marketing suggests that on-screen text matching spoken language have positive effects on video effectiveness (Brasel and Gips 2013; Wang, Zhang, and Jiang 2024). For instance, Brasel and Gips (2013) show that same-language subtitling in TV commercials has a strong effect on advertising processing, slightly lowering attention to visual elements and enhancing brand memory and recall of key verbal information (e.g., slogans, product features). Moreover, adding subtitles significantly increases the desire to learn more about the brand and usage intent. Similarly, Wang, Zhang, and Jiang (2024) show that when the voice-over matches the presence of subtitles in short video ads, consumers are more likely to exhibit higher levels of endorsement and purchase intention.

Though, research in educational psychology evidenced mixed findings regarding the joint presence of on-screen text and voices with visuals in enhancing processing, learning, and memory (Albers et al. 2023; Kalyuga, Chandler, and Sweller 2004; Mayer and Johnson 2008; Mayer and Moreno 2002). According to CTML, people learn better from graphics and narration than from graphics, narration, and on-screen written text (i.e., redundancy principle, Mayer 2021), because the simultaneous processing of verbal cues can overload working memory and increase extraneous processing (Kalyuga, Chandler, and Sweller 2004; Mayer and Moreno 2002). Nonetheless, research has identified conditions under which the contentual overlap of information across modes may even support learning (Albers et al. 2023). For instance, people learn better from an instructional video when subtitles are added to a narrated video in the learner's second language (i.e., subtitle principle; Mayer, Fiorella, and Stull 2020), as the written text helps compensate for reduced auditory comprehension. Content redundancy can also be beneficial for learning when only few keywords – rather than the whole text – are added to graphics and narration because a short written text may facilitate essential processing while not creating extraneous processing (Mayer and Johnson 2008).

Although research in educational psychology has shown that the presence of multiple verbal cues across modes can overload working memory and impair learning outcomes (Mayer 2021), we expect that such risks are likely attenuated in the context of social media videos, which are typically brief and optimized for fast-paced viewing. Therefore, in the context of social media, we argue that the simultaneous presence of text overlays and voices, compared to their absence, is more likely to foster engagement. While the overwhelming volume of content, combined with shrinking attention spans, increases the likelihood of peripheral processing and rapid scrolling (Tam and Inzlicht 2024), implementing multiple verbal cues (i.e., text overlay and voice-over) across complementary sensory channels (i.e., sight and hearing) can help capture attention and facilitate information processing. A deeper processing, in turn, enhances the probability of engaging with the post (Kanuri, Hughes, and Hodges 2024). Formally:

H_3: Individuals will express lower social media engagement for videos that present only visual content and a text caption, compared to posts embedding visuals, caption, text overlay and voice over.

How Video Design Shapes Consumer Engagement in Green-Oriented Communication

Companies increasingly rely on social media videos to showcase eco-friendly practices, communicate environmental values, and promote eco-friendly products. As such, understanding how consumers process green-oriented video communication is key to improve its effectiveness.

Prior research in this context suggests that multimodal richness in social media posts can increase persuasive effectiveness of green-oriented messages (Barberá-Tomás et al. 2019; Gonzalez-Arcos et al. 2021; Zhu, Cheng, and Wang 2025). For instance, Barberá-Tomás et al. (2019) show that pairing textual discourses with images can help in guiding interpretation and educating consumers, ultimately fostering social change. Similarly, Zhu, Cheng, and Wang (2025) find that, in pro-environmental videos, a moderate amount of informativeness, measured as a function of the number of both visual and verbal cues, can increase engagement on social media. However, they also show that an excess of information can decrease engagement, likely due to cognitive overload.

In this regard, while richer multimodal designs can enhance the effectiveness of green-oriented messages, they may also yield unintended negative consequences. While environmental claims are increasingly used by companies to appeal to consumers, they also attract greater scrutiny (Orazi and Chan 2020). Indeed, communication around corporate sustainable practices often elicits skepticism (Leonidou and Skarmeas 2017). Consumers face difficulties in distinguishing firms genuinely committed to social causes from those using sustainability as a strategic tool for profit (Nickerson et al. 2021; Parguel, Benoit-Moreau, and Larceneux 2011; Skarmeas and Leonidou 2013). These concerns are further exacerbated by the content of green claims, which can be vague, misleading, or unverifiable (Carlson, Grove, and Kangun 1993; Leonidou et al. 2011; Segev, Fernandes and Hong 2016). Beyond *what* is said, *how* it is said also may shape effectiveness, as both visual and verbal information could mislead consumers regarding their environmental practices or the environmental benefits of their products (Chen and Chang 2013; Parguel, Benoit-Moreau, and Russell 2015 Schmuck, Matthes, and Naderer 2018).

While richer video designs featuring text overlay and voice-over may enhance message richness and facilitate cognitive elaboration of green messages (Hartmann and Apaolaza-Ibánez 2009), this deeper processing may not be always beneficial. In prosocial and normative contexts where skepticism is already heightened – such as the one of sustainability – such elaboration is more likely to activate persuasion knowledge (Friestad and Wright 1994). In other words, using multiple elements across modes might backfire in those contexts where persuasive intent is more likely to be detected, scrutinized, and resisted. Consumers might become more skeptical when perceiving companies to make more effort to promote sustainability of their products and activities – that is, when they use a video design with additional verbal cues (e.g., text overlay, voice-over) to persuade consumers (Friestad and Wright 1994). After all, consumers could perceive marketing premises (i.e., making profit through endless sales; Csikszentmihalyi 2000) as truly in conflict with the achievement of sustainability goals (i.e., consuming less, consuming better; Kilbourne 2004; Sen et al. 2024; White, Habib, and Hardisty 2019). This could lead to negative reactions, such as the decision to disengage with the content (Campbell and Kirmani 2000; Eisend and Tarrahi 2022). This would resonate with prior research showing that excessive information around sustainability can lead to negative consumer attitudes and behaviors. For instance, prior research shows that more green claims (e.g., "natural, biodegradable, no pesticide") within the same campaign or advertisement can decrease brand attitude, likely because they trigger skepticism (Olsen, Slotegraaf, and Chandukala 2014). This mechanism may occur not only within a single mode (e.g., when information is conveyed through text), but also across modes (e.g., when both text, visuals, and speech convey information).

Beyond consumers' reactions to perceived persuasive attempts, a cognitive mechanism may also underlie consumers' lower engagement with richer video designs in the context of sustainability. In this regard, the presence of multiple modes might negatively affect consumers' ease of processing the information (Graf, Mayer, and Landwehr 2018), particularly when conveying non-redundant information (Ceylan, Diehl and Proserpio 2024), with, in turn, negative consequences on engagement. This issue may be particularly pronounced in green-oriented communication, as sustainability topics are often abstract and difficult to grasp for consumers (White, Habib, and Hardisty 2019). When processing becomes effortful, consumers may also question brand accuracy in providing information, a key dimension of authenticity (Nunes, Ordanini, and Giambastiani 2021). Reduced fluency can thus make the message appear less genuine and increase likelihood of confusion and perception of greenwashing (Chen and Chang 2013).

Moreover, the use of richer video designs integrating text overlays, voice-over, or both may also make the assertive tone of the message more salient. Prior research evidence that environmental messages are typically assertive in nature, aiming to encourage behavioral compliance and sustainable behavior change (e.g., Greenpeace's campaign "Stop talking. Start planting"; Kronrod et al., 2012; Milfeld and Pittman 2024). Assertive messages are effective in conveying urgency but can threaten consumers' perceived freedom of choice and autonomy (Brehm 1966; Rosenberg and Siegel 2018; Shen and Dillard 2005). As such, we might expect that when assertive messages are made salient by implementing text overlay and voice-over in video design, the risk of triggering reactance increases, with potential negative consequences on engagement.

Also, the prescriptive and normative nature of green messages may influence how video design shapes consumer responses. Green messages, as an educational intervention, “assume that informing the public about the environmental consequences of their actions should result in increased pro-environmental intentions and behavior” (Bolderdijk et al. 2013, p. 1). Increasing knowledge is often conceived as the first route to sustainable behavior change (Gifford and Nilson 2014). However, such messages often imply what individuals *should* do to act responsibly and sustainably, often requiring personal effort and costs for the benefit of others (White, Habib, and Hardisty 2019). Richer video design implementing text overlays or voice-over may make the normative stance of the green message more salient. This, in turn, may inadvertently trigger psychological reactance, ultimately reducing engagement (Brehm 1966).

Furthermore, the degree of congruency between brand’s typical content and its green-oriented messages may also influence how video design shape consumer responses. Prior research on incongruence shows that when communication deviates from established expectations, it can attract greater attention and cognitive elaboration (Heckler and Childers 1992). In the contest of green-oriented communication, consumers typically react negatively to perceived incongruencies between brand’s communication messages and its established identity (Ginder, Kwon, and Byun 2021; Salnikova, Strizhakova, and Coulter 2022; Torelli, Monga, and Kaikati 2012; White and Willness, 2009). On social media, such incongruence negatively influences how consumers react to social media posts, because the mismatch between the brand and the post content may increase scrutiny and raise suspicions of opportunism (Montaguti, Valentini, and Vecchioni 2023), or even greenwashing (Ginder, Kwon, and Byun 2021). Richer video designs that integrate text overlays and voice-over may further accentuate such inconsistencies in style, content, or narrative, ultimately reducing engagement.

To sum up, while multimodal richness can foster elaboration and improve attitudes and behaviors also in response to green messages (e.g., Hartmann and Apaolaza-Ibáñez 2009; Zhu, Cheng, and Wang 2025), it may also yield unintended negative effects. In particular, richer video design can trigger persuasion knowledge (Friestad and Wright 1994), affect processing fluency (Graf, Mayer, and Landwehr 2018), heighten the salience of assertive tones (Kronrod, Grinstein and Wathieu 2012) and the normative and educational-oriented stance of the message (Gifford and Nilson 2014), or exacerbate perceived incongruencies with the brand identity (Montaguti, Valentini, and Vecchioni 2023).

These contrasting perspectives raise an open question about whether richer video designs enhance or undermine engagement in green-oriented brand communication on social media. Our empirical investigation seeks to shed light on this issue.

4 EMPIRICAL ANALYSIS

Research Context and Data

To test how video design shapes consumer engagement of brand messages on social media, and whether, how, and why the effect might differ in the context of green (vs. non-green) communication, we turn to the field. Our research focuses on short-form videos shared by brands on TikTok, a leading video-based platform with more than 1.59 billion active users (DemandSage 2025). Unlike other social platforms such as Instagram or Facebook, TikTok is inherently centered on short-form video content, and brands increasingly invest on it to increase brand awareness, amplify content diffusion, and drive sales (TikTok for Business 2025). Moreover, TikTok's algorithm tends to favor videos that generate high engagement, which is often driven by an effective integration of design features such as text overlay and voice-over with the visual content (Sprout Social 2025; TikTok for Business 2022). This makes TikTok a particularly suitable and relevant context for our research purposes.

We selected the European companies from the 2023 Global Fortune 500 list with a verified TikTok account. We focused on large-scale companies, given the substantial resources they can allocate to communication activities, their broad audience reach and visibility, and the potential for significant societal impact. Nevertheless, enhancing the effectiveness of green-oriented communication is crucial for these firms to foster long-term brand value (Brand Finance 2025). Moreover, the European context is particularly relevant for our study because consumers ask for a more effective communication on environmental impact of their consumption choices and sustainability has become the focus of growing institutional and regulatory concerns (European Commission 2023).

We considered only one verified profile per each company according to the following criteria: if a company had a single global TikTok account, we included that profile. If no global account was available, we focused on the European one. In cases where companies presented multiple European accounts to promote their business in different countries, we selected the European account with the highest number of followers. In the case of corporate groups, we included all affiliated brands with a verified TikTok presence, as each brand pursue an independent communication strategy. We collected 6194 organic TikTok posts using the TikTok Research API. We included in the main sample only video-based posts and excluded other types of post (e.g., image-based posts, carousels). We further excluded those posts not available on the brand's profile at the moment of data collection, for instance due to content removal or inconsistencies with the verified account. The final sample resulted in 5,749 video-based posts by 52 companies across 21 industries from 1st January 2023 to 31st December 2023 (See Web Appendix A for details about data collection and cleaning processes).

Measures

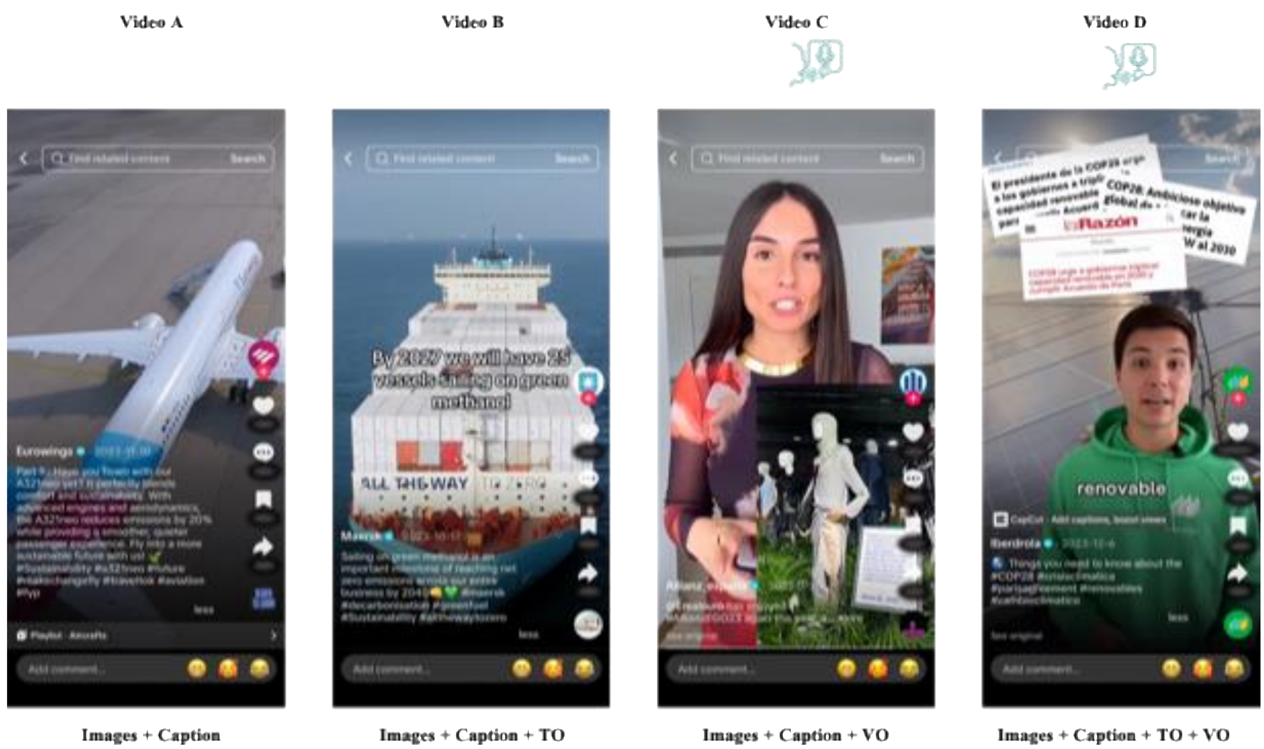
Engagement. We measured the engagement as the sum of like, comments and shares, as these behaviors on social media reflect consumers interest and support content diffusion and sales (Kumar et al. 2016; Liadeli, Sotgiu, and Verlegh 2023).

Video Design. The first author manually annotated the multimodal composition in all videos. Specifically, videos featuring the baseline elements (i.e., visuals and caption) represent 24.96% of the sample (n = 1,435), 28.20% include text overlays (n = 1,622), and 4.30% (n = 247) a voice-over. Finally, videos presenting both text overlays and voice-over sum up to 42.53% (n = 2,445) of the total sample.

Green Claim Videos. Following Banerjee's, Gulas, and Iyer (1995) definition of green-oriented communication, the first author and a research assistant independently coded the presence of visual, textual, or vocal sustainability-related cues in videos (Krippendorff's α

= .853). All disagreements were resolved through independent coding from the other authors. Figure 2 shows screenshot examples from our dataset of videos presenting green-oriented information for each video design. For instance, videos were coded as green when promoting products or services with reduced environmental impact through words (e.g., “sustainability”, “renewables”) or sentences in the caption or text overlay, as well as when discussing corporate environmental initiatives or topics related to environmental issues.

Figure 2. Examples of Green Claim Videos



Notes: Video A available at https://www.tiktok.com/@eurowings/video/7299833808767995169; Video B available at https://www.tiktok.com/@maersk_official/video/7290902871065890081; Video C available at https://www.tiktok.com/@allianz_es/video/7202299971901590790; Video D available at https://www.tiktok.com/@iberdrola/video/7309453253610147105.

Overall, 13.69% of the videos (n = 787) included at least one sustainability-related cue, qualifying the post as green-oriented. Table 3 and Table 4 show respectively the descriptive statistics of social media posts by industry and by brand, distinguishing green claim videos – that is, posts cueing sustainability of products, production processes, behaviors or lifestyles through visuals, text, or spoken words – from posts without such cues. Interestingly, companies communicating sustainability the most on TikTok in our sample belong to the two sectors that contribute the most to the emissions in Europe – namely, automotive and energy (N = 645) – according to European Environment Agency (EEA 2024). In contrast, industries whose core business is inherently less compatible with sustainability-oriented goals (e.g., airlines, fast fashion) appear to devote less effort to green-oriented communication, likely due to the reputational risks and accusations of greenwashing that might accompany their communication efforts.

Table 3. Sample Distribution by Industry

Industry	**Number of Posts (n)**	**Posts over Sample**	**non-Green Posts**	**Green Posts**	**Green Claim per Industry**	**Green Claim over Sample**
	N	**%**	**N**	**N**	**%**	**%**
Energy	59	1.03	3	56	94.92	7.12
Utilities	426	7.41	189	237	55.63	30.11
Petroleum Refining	25	.43	14	11	44.00	1.40
Motor Vehicles & Parts	1,082	18.82	741	341	31.52	43.33
Shipping	97	1.69	78	19	19.59	2.41
Engineering & Construction	15	.26	13	2	13.33	.25
Mail, Package, and Freight Delivery	77	1.34	68	9	11.69	1.14
Food & Drug Stores	300	5.22	276	24	8.00	3.05
Banks: Commercial and Savings	145	2.52	134	11	7.59	1.40
Wholesalers: Food and Grocery	333	5.79	308	25	7.51	3.18
Railroads	199	3.46	190	9	4.52	1.14
Aerospace & Defense	47	.82	45	2	4.26	.25
Telecommunications	429	7.46	414	15	3.50	1.91
Commercial Banks	64	1.11	62	2	3.13	.25
Pharmaceuticals	70	1.22	68	2	2.86	.25
Insurance: Life, Health (stock)	259	4.51	245	14	5.41	1.78
Airlines	458	7.97	452	6	1.31	.76
Specialty Retailers	1292	22.47	1290	2	.15	.25
Apparel	8	.14	8	0	.00	.00
Computer Software	7	.12	7	0	.00	.00
Household and Personal Products	357	6.21	357	0	.00	.00
Total	5,749		4,962	787		

Notes: companies were grouped into industries based on the classification used in the Global Fortune 500 list.

Table 4. Sample Distribution by Brand

Brand Profile	Industry	Number of Posts	Posts over Sample	non-Green Posts	Green Posts	Green Claim per Account	Green Claim over Sample
		N	%	N	N	%	%
e.on_italia	Energy	53	.92	2	51	96.23	6.48
engieitalia	Energy	6	.10	1	5	83.33	.64
peugeot	Motor Vehicles & Parts	13	.23	3	10	76.92	1.27
opel	Motor Vehicles & Parts	35	.61	10	25	71.43	3.18
naturgy	Utilities	212	3.69	73	139	65.57	17.66
iberdrola	Utilities	214	3.72	116	98	54.20	12.45
volkswagen	Motor Vehicles & Parts	167	2.90	83	84	50.30	10.67
repsol	Petroleum Refining	25	.43	14	11	44.00	1.40
volvotrucks	Motor Vehicles & Parts	262	4.56	152	110	41.98	13.98
vauxhall	Motor Vehicles & Parts	12	.21	7	5	41.67	.64
chrysler	Motor Vehicles & Parts	45	.78	29	16	33.56	2.03
daimlertruck	Motor Vehicles & Parts	31	.54	21	10	32.26	1.27
mercedesbenz	Motor Vehicles & Parts	83	1.44	57	26	31.33	3.30
citroen_fr	Motor Vehicles & Parts	88	1.53	67	21	23.86	2.67
maersk_official	Shipping	97	1.69	78	19	19.59	2.41
renault	Motor Vehicles & Parts	90	1.57	73	17	18.89	2.16
ramtrucks	Motor Vehicles & Parts	41	.71	34	7	17.07	.89
groupebouygues	Engineering & Construction	15	.26	13	2	13.33	.26
dhl	Mail, Package, and Freight Delivery	77	1.34	68	9	11.69	1.14
continental	Motor Vehicles & Parts	49	.85	44	5	10.20	.64
edeka	Wholesalers: Food and Grocery	227	3.95	204	23	10.13	2.92
sncfvoyageurs	Railroads	47	.82	43	4	8.51	.51
carrefourfrance	Food & Drug Stores	300	5.22	276	24	8.00	3.05
creditagricole	Banks: Commercial and Savings	145	2.52	134	11	7.59	1.40
jeep	Motor Vehicles & Parts	14	.24	13	1	7.14	.13
allianz_es	Insurance: Life, Health (stock)	259	4.51	245	14	5.41	1.78
airbus	Airlines	47	.82	45	2	4.26	.25
brusselsairlines	Airlines	25	.43	24	1	4.00	.13
deutschetelekom	Telecommunications	204	3.55	196	8	3.92	1.02
telefonica	Telecommunications	217	3.77	210	7	3.26	.89
deutschebahn	Railroads	152	2.64	147	5	3.29	.64
maserati	Motor Vehicles & Parts	92	1.60	89	3	3.26	.38
santander_es	Commercial Banks	64	1.11	62	2	3.13	.25
eurowings	Airlines	135	2.35	131	4	2.96	.51
sanofi	Pharmaceuticals	70	1.22	68	2	2.86	.25
alfaromeo_global	Motor Vehicles & Parts	42	.73	41	1	2.38	.13
metro_de	Wholesalers: Food and Grocery	106	1.84	104	2	1.89	.25
zarahome	Specialty Retailers	77	1.34	76	1	1.30	.13
lufthansa	Airlines	117	2.04	116	1	.85	.13
pullandbear	Specialty Retailers	443	7.71	442	1	.23	.13
ikeaitalia	Specialty Retailers	71	1.23	71	0	.00	.00
Sap	Computer Software	7	.12	7	0	.00	.00
lorealparis	Households and Personal Products	357	6.21	357	0	.00	.00
bershka	Specialty Retailers	354	6.16	354	0	.00	.00
stradivarius	Specialty Retailers	203	3.53	203	0	.00	.00
flyswiss	Airlines	164	2.85	164	0	.00	.00
massimodutti	Specialty Retailers	144	2.50	144	0	.00	.00
austrianairlines	Airlines	17	.30	17	0	.00	.00
dodge	Motor Vehicles & Parts	17	.30	17	0	.00	.00
dior	Apparel	8	.14	8	0	.00	.00
orange	Telecommunications	8	.14	8	0	.00	.00
bmw	Motor Vehicles & Parts	1	.02	1	0	.00	.00
Total		5,749		4,962	787		

Notes: companies were grouped into industries based on the classification used in the Global Fortune 500 list.

Control Variables. We consider multiple covariates that might influence engagement.

First, we control for several linguistic features[1]. Regarding key caption features, we focused on the number of hashtags, mentions, and emojis. Hashtags could increase post visibility and discovery, be an indicator of trending topics on the platform, and also lead consumers to make inferences on the video content, which may ultimately influence engagement (Valsesia and Diehl 2022). Similarly, mentions can increase impressions and engage users by directly tagging other brands or users (Walker, Dimitriu, and Macdonald 2017). Emojis can increase playfulness and convey emotions, thereby influencing engagement (Luangrath, Xu, and Wang 2023; McShane et al. 2021). We also control for other key linguistic properties which, unlike caption-specific key features, have been measured across all available verbal cues (i.e., text overlay, voice-over, caption). In this regard, we control for valence and arousal using Mohammad's (2025) VAD lexicon, since the language positivity and intensity can influence engagement (Cascio Rizzo et al. 2024; Berger and Milkman 2012). We also control for readability using Flesh Kinkaid scores (Pancer et al. 2019), concreteness using Paetzold and Specia's (2016) concreteness ratings, because easy-to-read and concrete posts might be easier to process, thus triggering engagement. Moreover, we control for verbal density, as a higher number of words across verbal cues per second might convey more information, but also overwhelm users and cause information overload, thereby affecting engagement.

Second, we control for key visual features over the video. Following prior literature, we used an open-source video mining tool from Schwenzow et al. (2021) to control for both structural features (i.e., technical aspects of the video) such as colorfulness and saturation, and content features, such as the share of frames with faces, and facial emotional states (i.e.,

[1] Text overlays for each post were manually annotated by the first author. Voice-over transcripts were first obtained by using OpenAI's Whisper model (Radford et al. 2023), and then manually refined by the first author for accuracy. Given that the posts were originally in different languages (e.g., English, French, Spanish, German), we translated textual captions, text overlays, and voice-overs into English before computing linguistic measures.

happiness, sadness, fear, anger, surprise, disgust), that have been shown to influence engagement (Wang, Ding, and Hu 2025).

Third, we account for key audial features that play a role in shaping consumer behavior and persuasion (Hildebrand et al. 2020; Van Zant and Berger 2020; Wang et al. 2021). Specifically, we used the Librosa Python's library (McFee et al. 2015) to measure volume (Xu et al. 2025; Chang, Mukherjee, and Chattopadhyay 2023), pitch (Apple, Streeter, and Krauss 1979; Lowe and Haws 2017), and speech rate (Mairesse et al. 2007; Chattopadhyay et al. 2003).

Fourth, we control for video duration, as longer videos may convey more information, but can also increase cognitive fatigue and reduce viewer attention in fast-scrolling contexts such as TikTok (Tam and Inzlicht 2024). As such, we also control for its squared term to account for potential boredom caused by excessively long videos (Li, Shi, and Wang 2019). The variable was log-transformed to correct for skewness.

Fifth, we control for the (log) number of views, given that a video viewed by a larger audience might receive more likes, comments, or shares (Li and Xie 2020).

Finally, we include time fixed effects for month, weekday, and time of the day (Kanuri, Chen, and Sridhar 2018), and brand fixed effects to capture both observable (e.g., number of followers, industry sector) and unobservable (e.g., account objectives, follower profiles) differences across accounts (Dang et al. 2025; Farace et al. 2025).

Table 5 shows the full list of measures, their operationalizations, and sources, and Table 6 reports summary statistics. See Web Appendix B for correlations (Table WB1).

Table 5: Measurement in the Field Study

Variable	Definition and Operationalization	Source	Related Research
Dependent Variable			
Engagement	Sum of likes, comments, and shares	TikTok API	Liadeli, Sotgiu, and Verlegh 2023
Independent Variables			
Video Design	multi-categorical variable for the following video designs: a) Images + Caption; b) Images + Caption + TO; c) Images + Caption + VO; d) Images + Caption + TO + VO)	Human Annotation	
Green Claim	This indicates whether a video includes a green-oriented visual, textual, or vocal cue or not; dummy coded	Human Annotation	Banerjee, Gulas, and Iyer 1995
Additional Controls			
Views	Total number of views for a video on TikTok	TikTok API	Li and Xie 2020
Video Controls			
Video Duration	Duration of the video in seconds, measured by using an open-source tool from Schwenzow et al. 2021	Video Mining	Zhou et al. 2021
Audial Controls			
Volume	Intensity of the sound in decibels, measured by using Python's LibROSA library	Audio Mining	Chang et al. 2023; Liu et al. 2018
Pitch	Fundamental frequency of the sound in hertz, measured by using Python's LibROSA library	Audio Mining	Cascio Rizzo et al. 2024; Zhang, Wang, and Chen 2020
Speech Rate	Number of words per second, measured by dividing the number of words in the speech to the duration of the video	Audio Mining	Chang et al. 2023; Mairesse 2007
Visual Controls			
Faces	Share of frames with at least one detected face over all analyzed frames, measured by using an open-source tool from Schwenzow et al. 2021	Video Mining	Wang, Ding, and Hu 2025
Faces Emotions (e.g., happy, sad)	mean of confidence levels at video level for each emotion across all detected faces in %, measured by using an open-source tool from Schwenzow et al. 2021	Video Mining	Wang, Ding, and Hu 2025
Colorfulness	As defined in Hasler and Süsstrunk (2003), measured by using an open-source tool from Schwenzow et al. 2021	Video Mining	Li and Xie 2020
Color Saturation	Degree of difference from a pure color to the respective gray having the same luminosity, measured by using an open-source tool from Schwenzow et al. 2021	Video Mining	Zhou et al. 2021
Caption Controls			
Valence	Valence score from Mohammad's (2025) VAD lexicon, averaged across caption, TO, and VO – if available	Text Mining	Berger and Milkman 2012
Arousal	Arousal score from Mohammad's (2025) VAD lexicon, averaged across caption, TO, and VO – if available	Text Mining	Berger and Milkman 2012
Concreteness	How much a word refers to an actual, tangible, or "real" entity, measured from Paetzold and Specia's (2016) dictionary; score averaged across caption, TO, and VO – if available	Text Mining	Packard and Berger 2021
Readability	Text readability score measured by using Flesch-Kincaid Grade Level measure, averaged across caption, TO and VO – if available	Text Mining	Pancer et al. 2019
Hashtags	Number of hashtags included in the caption	Text Mining	Valsesia et al. 2022
Mentions	Number of mentions included in the caption	Text Mining	Villarroel et al. 2019
Emoji	Number of emojis included in the caption	Text Mining	Luangrath, Xu, and Wang 2023
Verbal Density	Sum of word count across TO, VO and caption, divided by the video duration	Text Mining	
Alternative Explanations			
Verbal Similarity	Mean of cosine similarities between a) caption and TO b) caption and VO c) TO and VO – if available	Text Mining	
Visual-Verbal Similarity	Mean of cosine similarities between the embeddings of video frames and the available verbal elements (caption, TO and VO), measured by using the Open AI's Contrastive Language-Image Pre-Training (CLIP ViT-B/32) model	Video Mining	Deng et al. 2025
Informative (cf. Commercial) Goal	Measure of how informative (vs. commercial) a post goal is, calculated as the proportion of informative words to the total of informative and commercial words	Text Mining	Cascio Rizzo et al. 2024
Assertiveness	Assertiveness score measured by using Khan's and Kronrod (2025) tool for measuring language assertiveness	Text Mining	Kronrod, Grinstein, and Wathieu 2012
Brand Congruity	Cosine similarity measured at brand level between the aggregated bag of words (including caption, TO, and VO) of those including green (vs. non-green) claims	Text Mining	Montaguti, Valentini, and Vecchioni 2023
Fixed Effects			
Brand	Dummy variables capturing the brands	TikTok API	Dang et al. 2025; Farace et al. 2025
Time	Dummy variables capturing the posting time in relation to the month, whether the post was published on a weekend, and the time of day (i.e., AM or PM)	TikTok API	Kanuri, Chen and Sridhar 2018

Table 6. Summary Statistics

Variable	N	Mean	SD	Min	Max
Like	5749	5776.873	30011.88	0	1202097
Comment	5749	141.29	1329.41	0	51616
Share	5749	171.274	1640.906	0	64619
Engagement	5749	6089.437	32016.696	0	1285080
Green Wordcount Ratio	5748	.008	.023	0	.333
Greenness	5749	.573	1.631	0	7
(log)Views	5748	8.732	3.09	.693	17.367
(log)Video Duration	5749	3.094	.656	1.188	6.329
Volume	5743	-68.104	5.963	-79.814	0
Pitch	5707	125.259	132.812	65.406	2010.73
Speech Rate	5749	1.183	1.472	0	5.293
Colorfulness	5748	.506	.103	0	.72
Saturation	5748	.29	.119	0	.964
Faces	5749	.425	.343	0	1
Happy	5749	.189	.229	0	1
Sad	5749	.166	.175	0	1
Surprise	5749	.027	.065	0	.993
Fear	5749	.139	.161	0	1
Angry	5749	.074	.118	0	1
Disgust	5749	.005	.024	0	.995
Hashtag	5749	4.154	2.703	0	24
Mentions	5749	.222	.532	0	12
Emoji	5749	1.377	1.406	0	15
Valence	5729	.227	.125	-.65	.872
Arousal	5729	-.045	.103	-.592	1
Concreteness	5674	350.064	31.54	242.426	595.731
Readability	5748	20.224	15.661	-3.01	333.021
Verbal Density	5749	3.404	2.541	0	27.177
Verbal Similarity	5749	.206	.205	0	1
Visual-Verbal Similarity	5748	.239	.032	.088	.404
Informative Goal	5749	.596	.292	0	1
Brand Congruity	4398	.794	.163	.192	.965
Assertiveness	5749	1.7	1.733	0	5

Notes: Time and brand fixed effects are not included.

Model Estimation

Given that our dependent measure was a count variable and overdispersed with no excess of zeros (i.e., less than 1%), we run a negative binomial regression ($p < .001$, likelihood ratio test). The model takes the following form:

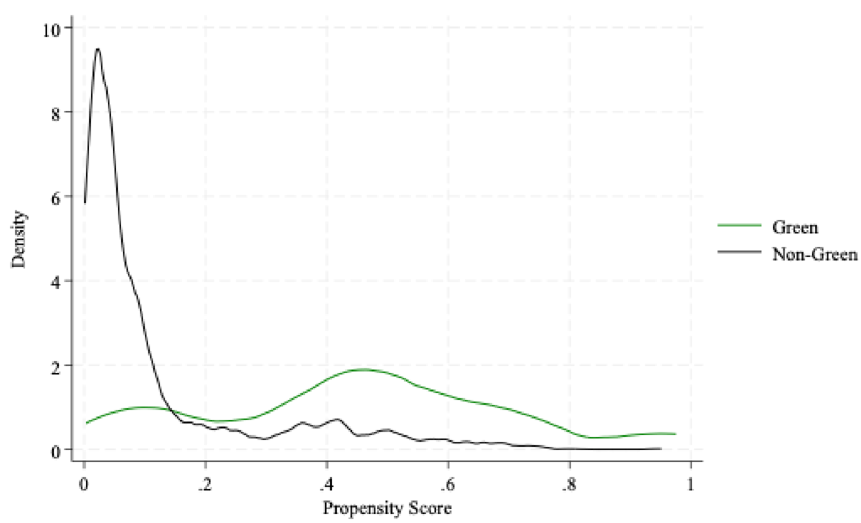
$$\log(\mu_{ibt}) = \alpha + \sum_{k} \beta_k \mathrm{VideoDesign}_{ik} + \gamma \mathrm{GreenClaim}_i + \sum_{k} \delta_k (\mathrm{VideoDesign}_{ik} \times \mathrm{GreenClaim}_i) + \boldsymbol{Z}_i'\theta + \boldsymbol{A}_i'\varphi + \boldsymbol{V}_i'\psi + \boldsymbol{T}_i'\omega + \iota_b + \tau_t \quad (1)$$

$$Y_{ibt} \sim NegBin(\mu_{ibt}, r)$$

where μ_{ibt} is the expected value of engagement received by the post i posted by the brand b at time t. $VideoDesign$ is a categorical variable capturing the multimodal configuration of the post i. Accordingly, $VideoDesign_{ik}$ is a set of dummy variables indicating whether post i adopts a video design k, depending on the presence of text overlay, voice-over, or both. Videos featuring only visual content and caption serve as the reference category. GreenClaim_i is a binary indicator which is equal to one if the post i includes a green message, and zero otherwise. β_k, γ, and δ_k are the key parameters of interest in our study, representing respectively the impact of multimodal compositions, green claim presence, and the interaction between the two on social media engagement. $\boldsymbol{Z}_i'$ is the vector of baseline controls per each post i posted by the brand b at time t, including (log) video duration, its squared term, and the (log) number of views of the post i. $\boldsymbol{A}_i'$ is the vector of audio controls per each post i posted by the brand b at time t, which includes volume, pitch, and speech rate. $\boldsymbol{V}_i'$ is the vector of visual controls per each post i posted by the brand b at time t, which accounts for colorfulness, saturation, face presence and related emotions. $\boldsymbol{T}_i'$ is the vector of textual controls per each post i posted by the brand b at time t, and includes valence, arousal, concreteness, readability, the number of hashtags, emojis, and mentions. ι_b and τ_t are brand and time fixed effects, respectively; brand fixed effects account for time-invariant unobserved heterogeneity across brands (e.g., brand popularity and positioning, follower base on social media, differences in posting strategies), while time fixed effect account for posting time patterns (e.g., seasonality, whether the post was published during weekend or not).

Noteworthy, whether to post a sustainability-oriented content on social media can be a strategic and deliberate decision of the company, which may correlate with other factors that also influence engagement. Therefore, following prior research (e.g., Li and Xie 2020), we use propensity score matching (PSM) to alleviate endogeneity concerns about self-selection bias and adjust for the differences in the treatment and control group (Rosenbaum and Rubin 1983). In our study, the propensity score is the predicted probability to post a green claim conditional on the value of the covariates. When the propensity scores for two observations are close enough to each other, the treatment is considered random. Both observable and unobservable factors may lead to endogeneity and bias our inferences. Therefore, we included brand and time fixed effects in our matching model. We adopted a 1:1 nearest-neighbor matching algorithm without replacement and a caliper of .2 to match a green post with a non-green post. This approach is particularly suitable for our setting, where the distributions of propensity scores for treated and control posts only partially overlap, while most control posts are concentrated at lower scores (Figure 3). Under such conditions, kernel matching would rely on distant and dissimilar control observations for matching, introducing smoothing bias (Heckman, Ichimura, and Todd 1997, 1998). In contrast, nearest-neighbor matching ensures that each treated post is matched with a directly comparable control unit.

Figure 3. Propensity Score for Green and non-Green Posts before Matching



The resulting matched sample included 1256 videos, half green and half non-green (Table 7). After matching, balance diagnostics indicate adequate covariate balance between treated and control units (B = 23.5, R = .72). Figure 4 shows the propensity score after matching, while Table 8 presents descriptive statistics for each variable before and after matching.

Figure 4. Propensity Score for Green and non-Green Posts after Matching

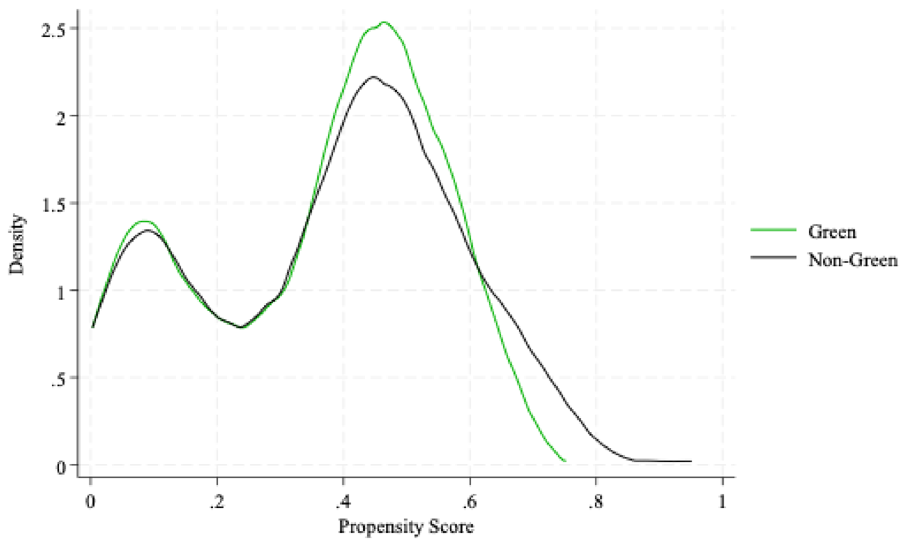


Table 7. Frequency Distribution of Green (vs. non-Green) Videos by Design

	Before Matching		After Matching	
Video Design	**Non-Green**	**Green**	**Non-Green**	**Green**
	# (%)	**# (%)**	**# (%)**	**# (%)**
Image+Caption	1,366 (27.53)	69 (8.77)	153 (24.36)	57 (9.08)
Image+Caption+TO	1,364 (27.49)	257 (32.66)	174 (27.71)	220 (35.03)
Image+Caption+VO	219 (4.41)	29 (3.68)	27 (4.30)	20 (3.18)
Image+Caption+TO+VO	2,013 (40.57)	432 (54.89)	274 (43.63)	331 (52.71)
Total	**4,962**	**787**	**628**	**628**

Table 8. Descriptive Statistics, Before and After Matching

A. Discrete Variables					
		Before Matching		**After Matching**	
	Level	**Count**	**Percentage**	**Count**	**Percentage**
Green Claim Presence	Yes	787	13.69	628	50.00
Weekday	Yes	917	15.95	212	16.88
Pm	Yes	4,113	71.54	813	64.73
Month	January	345	6.00	88	7.01
	February	380	6.61	59	4.70
	March	460	8.00	77	6.13
	April	417	7.25	96	7.64
	May	456	7.93	105	8.36
	June	515	8.96	103	8.20
	July	490	8.52	83	6.61
	August	516	8.98	115	9.16
	September	532	9.25	123	9.79
	October	541	9.41	136	10.83
	November	554	9.64	135	10.75
	December	543	9.45	136	10.83
Brand	airbus	47	0.82	5	0.40
	alfaromeo_global	42	0.73	2	0.16
	allianz_es	259	4.51	27	2.15
	austrianairlines	17	0.30	–	–
	bershka	354	6.16	–	–
	bmw	1	0.02	–	–
	brusselsairlines	25	0.43	2	0.16
	carrefourfrance	300	5.22	50	3.98
	chrysler	45	0.78	32	2.55
	citroen_fr	88	1.53	39	3.11
	continental	49	0.85	9	0.72
	creditagricole	145	2.52	22	1.75
	daimlertruck	31	0.54	19	1.51
	deutschebahn	152	2.64	10	0.80
	deutschetelekom	204	3.55	16	1.27
	dhl	77	1.34	20	1.59
	dior	8	0.14	–	–
	dodge	17	0.30	–	–
	e.on_italia	53	0.92	2	0.16
	edeka	227	3.95	47	3.74
	engieitalia	6	0.10	1	0.08
	eurowings	135	2.35	8	0.64
	flyswiss	164	2.85	–	–
	groupebouygues	15	0.26	3	0.24
	iberdrola	214	3.72	179	14.25
	ikeaitalia	71	1.23	–	–
	jeep	14	0.24	1	0.08
	lorealparis	357	6.21	–	–
	lufthansa	117	2.04	2	0.16
	maersk_official	97	1.69	38	3.03
	maserati	92	1.60	6	0.48
	massimodutti	144	2.50	–	–
	mercedesbenz	83	1.44	48	3.82
	metro_de	106	1.84	4	0.32
	naturgy	212	3.69	138	10.99
	opel	35	0.61	15	1.19
	orange	8	0.14	–	–
	peugeot	13	0.23	4	0.32
	pullandbear	443	7.71	2	0.16
	ramtrucks	41	0.71	14	1.11
	renault	90	1.57	31	2.47
	repsol	25	0.43	23	1.83
	sanofi	70	1.22	4	0.32
	santander_es	64	1.11	4	0.32
	sap	7	0.12	–	–
	sncfvoyageurs	47	0.82	7	0.56
	stradivarius	203	3.53	–	–
	telefonica	217	3.77	14	1.11
	vauxhall	12	0.21	10	0.80
	volkswagen	167	2.90	160	12.74
	volvotrucks	262	4.56	236	18.79
	zarahome	77	1.34	2	0.16

B. Continuous Variables										
	Before Matching					After Matching				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
Like	5749	5,776.873	30,011.88	0	1,202,097	1256	5,234.537	24,861.515	1	537,886
Comment	5749	141.29	1,329.41	0	51,616	1256	96.59	548.134	0	11,631
Share	5749	171.274	1,640.906	0	64,619	1256	189.176	1,323.065	0	20,562
Engagement	5749	6,089.437	32,016.696	0	1,285,080	1256	5,520.303	26,317.19	1	561,924
(log)Views	5748	.008	.023	0	.333	1256	.023	.037	0	.333
Green Wordcount Ratio	5749	.573	1.631	0	7	1256	2.045	2.51	0	7
Greenness	5748	8.732	3.09	.693	17.367	1256	8.545	3.057	.693	16.574
(log)Video Duration	5749	3.094	.656	1.188	6.329	1256	3.2	.69	1.188	5.415
Volume	5743	-68.104	5.963	-79.814	0	1255	-67.877	5.584	-79.629	0
Pitch	5707	125.259	132.812	65.406	2,010.73	1249	122.258	125.111	65.409	2,010.73
Speech Rate	5749	1.183	1.472	0	5.293	1256	1.308	1.491	0	5.293
Colorfulness	5748	.506	.103	0	.72	1256	.498	.098	.12	.716
Saturation	5748	.29	.119	0	.964	1256	.296	.114	.028	.933
Faces	5749	.425	.343	0	1	1256	.32	.33	0	1
Happy	5749	.189	.229	0	1	1256	.191	.24	0	1
Sad	5749	.166	.175	0	1	1256	.144	.171	0	.998
Surprise	5749	.027	.065	0	.993	1256	.027	.065	0	.708
Fear	5749	.139	.161	0	1	1256	.132	.172	0	1
Angry	5749	.074	.118	0	1	1256	.072	.128	0	.996
Disgust	5749	.005	.024	0	.995	1256	.005	.041	0	.995
Hashtag	5749	4.154	2.703	0	24	1256	4.036	2.573	0	24
Mentions	5749	.222	.532	0	12	1256	.113	.368	0	3
Emoji	5749	1.377	1.406	0	15	1256	1.136	1.336	0	11
Valence	5729	.227	.125	-.65	.872	1253	.225	.121	-.354	.862
Arousal	5729	-.045	.103	-.592	1	1253	-.038	.099	-.39	.786
Concreteness	5674	350.064	31.54	242.426	595.731	1233	351.072	28.944	265.125	595.731
Readability	5748	20.224	15.661	-3.01	333.021	1256	22.073	14.181	.388	110.852
Verbal Density	5749	3.404	2.541	0	27.177	1256	3.729	2.542	.102	15.103
Verbal Similarity	5749	.206	.205	0	1	1256	.23	.201	0	.82
Visual-Verbal Similarity	5748	.239	.032	.088	.404	1256	.243	.033	.145	.381
Assertiveness	5749	1.7	1.733	0	5	1256	1.76	1.67	0	5
Informative Goal	5749	.596	.292	0	1	1256	.615	.3	0	1
Brand Congruity	4398	.794	.163	.192	.965	1256	.87	.113	.192	.965

Results

Table 9 shows the results after matching[2]. We report also incident rate ratios (IRR) which reflects the proportional change in engagement due to each variable (IRR > 1 indicate a positive effect and more engagement; IRR < 1 indicates a negative effect and less engagement). Noteworthy, when a post is green-oriented, enriching video design with text overlays, voice-over, or both significantly decreases engagement (Table 9, Model 1). Results are consistent even after accounting for baseline controls (i.e., number of views, video duration, brand and time fixed effects; Table 9, Model 2), and also multimodal features (Table 9, Model 3). In other words, using text overlays (IRR = .634, b = -.456, SE =.180, z = -2.53, p = .011), voice-over (IRR = .287, b = -1.247, SE =.313, z = -2.23, p = .000), or a combination of both (IRR = .600, b =

[2] For completeness, results from the full sample (i.e., before matching) are reported in Web Appendix B, Table WB2.

-.511, SE =.173, z = -2.95, p = .003) on top of the baseline elements (i.e., visuals and caption) negatively affects the engagement of videos promoting sustainability through textual, vocal or visual cues (Table 9, Model 3). Specifically, green claim videos including text overlays are associated with a 42.3% decrease in engagement; green claim videos enriched with voice-over are associated with a 76.8% decrease in engagement; green claim videos including both text overlays and voice-over are associated with a 56.9% decrease in engagement. Thus, our findings suggest that companies on social media should carefully implement text overlays or voice-over when communicating sustainability through videos, as a richer video design is likely to decrease engagement.

Table 9: Results After PSM

DV: Engagement	Model 1: only IVs			Model 2: Baseline controls			Model 3: Multimodal controls		
	IRR	**Coeff.**	***p***	**IRR**	**Coeff.**	***p***	**IRR**	**Coeff.**	***p***
Independent Variables:									
Video Design									
Image+Caption+TO	.696	-.362 (.199)	.069	1.008	.008 (.105)	.937	.910	-.095 (.116)	.416
Image+Caption+VO	1.552	.440 (.375)	.242	1.189	.173 (.203)	.393	.807	-.214 (.228)	.348
Image+Caption+TO+VO	.528	-.639 (.182)	.000	1.074	.071 (.108)	.512	.719	-.330 (.148)	.026
Green Claim:									
Yes	2.001	.694 (.279)	.013	1.012	.012 (.139)	.933	1.202	.184 (.155)	.235
Interactions: Video Design * Green Claim									
Image+Caption+TO * Green Claim	.234	-1.451 (.334)	.000	.726	-.320 (.168)	.057	.634	-.456 (.180)	.011
Image+Caption+VO * Green Claim	.081	-2.512 (.600)	.000	.313	-1.161 (.303)	.000	.287	-1.247 (.313)	.000
Image+Caption+TO+VO * Green Claim	.271	-1.305 (.315)	.004	.770	-.261 (.158)	.099	.600	-.511 (.173)	.003
Additional Controls:									
Log (Views)				1.868	.625 (.013)	.000	1.836	.608 (.013)	.000
Log (Video Duration)				1.468	.384 (.317)	.227	.705	-.349 (.353)	.323
Log (Video Duration)^2				.941	-.061 (.049)	.217	1.070	.068 (.057)	.235
Audial Controls									
Volume							1.016	.016 (.007)	.022
Pitch							1.000	-.000 (.000)	.072
Speech Rate							1.237	.212 (.050)	.000
Visual Controls									
Colorfulness							1.177	.163 (.308)	.596
Saturation							235.316	.856 (.241)	.000
Faces							.869	-.141 (.108)	.193
Happy							1.226	.204 (.117)	.082
Sad							.859	-.152 (.152)	.318
Surprise							.866	-.144 (.423)	.733
Fear							1.234	.210 (.152)	.167
Angry							1.358	.306 (.218)	.161
Disgust							.530	-.634 (.625)	.310
Verbal Controls									
Caption Hashtags							1.014	.014 (.014)	.337
Caption Mentions							.844	-.170 (.082)	.037
Caption Emoji							.945	-.057 (.024)	.019
Valence							.588	-.532 (.228)	.020
Arousal							1.609	.475 (.289)	.100
Concreteness							.998	-.002 (.001)	.095
Readability							.990	-.010 (.004)	.013
Verbal Density							.992	-.008 (.025)	.752
Fixed Effects									
Time FE					Included			Included	
Brand FE					Included			Included	
Intercept	9288	9.137 (.145)	.000	4.337	1.467 (.645)	.023	91.385	4.515 (.972)	.000
N		1,256			1,256			1,225	
AIC		21,738.490			19,222.506			18,640.759	
Log likelihood		-10,860.245			-9,547.252			-9,236.380	

Notes: IRR = incident rate ratio. An IRR greater than (less than) 1 indicates a positive (negative) effect. Standard errors are in parentheses. TO = text overlay, VO = voice over.

Exploring the Process in the Field

In this section, we first seek insights into the underlying mechanisms explaining the contrasting effects of richer video designs for green- versus non-green communication. We then turn to the subsample of green-oriented posts to explore whether any of these factors helps explain within-context variability in video design's impact on engagement.

We begin by exploring the reasons why video design's impact on engagement may differ between green and non-green communication. One explanation might be that the presence of multiple modes might limit consumers' ease of processing the information (Graf, Mayer, and Landwehr 2018), especially when conveying non-redundant information (Ceylan, Diehl and Proserpio 2024). This may be particularly problematic in green-oriented communication, since sustainability is an inherently complex and multifaceted phenomenon (Madan et al. 2023), often difficult to grasp for consumers (White, Habib, and Hardisty 2019). To rule out this possibility, we control for content redundancy – that is, the semantic overlap of information across modes (Table 10, Model 1).

Another explanation may lie in the implicit educational, normative stance often embedded in green messages. Green messages, as an informational intervention, "assume that informing the public about the environmental consequences of their actions should result in increased pro-environmental intentions and behavior" (Bolderdijk et al. 2013, p. 1). While increasing knowledge is often conceived as the first route to sustainable behavior change (Gifford and Nilson 2014), a richer video design may cause the message to come across as overtly educational or prescriptive. The reinforcement of the message's normative stance may inadvertently trigger psychological reactance, ultimately reducing engagement (Brehm 1966).

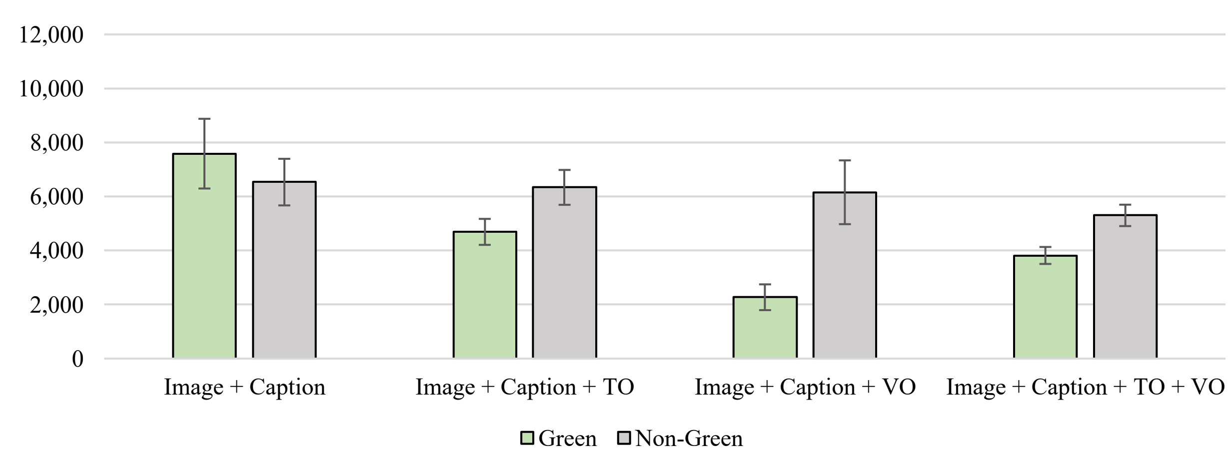
Therefore, we account for the goal of the post (that is, informative vs. commercial, see Table 10, Model 2).

Furthermore, we control for the level of brand congruency in communication style of posts including (vs. not including) a green claim (Table 10, Model 3; Montaguti, Valentini, and Vecchioni 2023). Coherently, if posts featuring a green claim seem incongruent with a brand's typical communication style, consumers might respond negatively.

We also explore the role of assertiveness of the message (Kronrod, Grinstein, and Wathieu 2012), as it might threaten consumers' perceived freedom of choice (Shen and Dillard 2005). The use of multiple modes could increase the salience of assertive messages, particularly when the message replicates across modes, thereby causing reactance and disengagement (Brehm 1966). As a consequence, we control for it (see Table 10, Model 4).

The results hold even after accounting for all these alternative explanations (Figure 5; Table 10, Model 5), thereby suggesting that none of them fully explain our results. In other words, the reason why video design exerts contrasting effects on engagement for green- versus non-green oriented posts cannot be fully attributed to message assertiveness, semantic redundancy across modes, post goal, or congruency in communication style.

Figure 5: Predictive Engagement by Type of Post (Video Design * Green Claim)



Beyond these factors, it might be reasonable to think that the negative effect of richer video designs in the context of green-oriented communication might be ascribed to the mere presence of multiple verbal elements across modes. Although our field data do not allow us to empirically test this interpretation, consumers might be more likely to be skeptical and activate persuasion knowledge (Friestad and Wright 1994) when companies make greater efforts to promote sustainability of their products and activities – in other words, when they employ richer video designs with text overlay and voice-over to persuade consumers. After all, consumers could perceive traditional marketing premises (i.e., making profit through endless sales; Csikszentmihalyi 2000) as truly in conflict with the achievement of sustainability goals (i.e., consuming less, consuming better; Kilbourne 2004; Sen et al. 2024; White, Habib, and Hardisty 2019). As a result, this would suggest that beyond *what* is said, also *how* it is said can shape the effectiveness of green-oriented communication, leading also consumers to perceive greenwashing attempts (Chen and Chang 2013).

We further examined, within green-oriented posts, whether these factors could account for differences in how video design shapes engagement. In particular, video design including text overlay, voice-over, or both exerts a significant negative effect on engagement when accounting separately for semantic similarity across modes (b_{TO} = -.522, SE = .174, t = -3.00, p = .003; b_{VO} = -1.433, SE = .272, t = -5.27, p = .000; b_{VO+TO} = -.876, SE = .224, t = -3.91, p = .000; Table 11, Model 1a), type of post (b_{TO} = -.587, SE = .167, t = -3.42, p = .000; b_{VO} = -1.500, SE = .272, t = -5.51, p = .000; b_{VO+TO} = -.952, SE = .219, t = -4.35, p = .000; Table 11, Model 2a), brand congruency in communication (b_{TO} = -.592, SE = .166, t = -3.57, p = .000; b_{VO} = -1.520, SE = .270, t = -5.63, p = .000; b_{VO+TO} = -.963, SE = .218, t = -4.42, p = .000; Table 11, Model 3a), and message assertiveness (b_{TO} = -.592, SE = .168, t = -3.52, p = .001; b_{VO} = -1.519, SE = .270, t

= -5.63, p = .000; b_{VO+TO} = -.963, SE = .219, t = -4.40, p = .000; Table 11, Model 4a). Notably, the negative effect of richer video designs on engagement becomes not significant when the interaction with the post's goal orientation is included (Table 11, Model 2b). Specifically, when the green-oriented post tends to be more informational than commercial-oriented, using text overlays (b = -2.633, SE =.857, z = -3.07, p = .002), voice-over (b = -2.924, SE =1.030, z = -2.84, p = .005), or a combination of both (b = -2.712, SE =.864, z = -3.14, p = .002) on top of the baseline elements (i.e., visuals and caption) significantly lowers engagement (Table 11, Model 2b). In other words, results show that when the post tends to be more informative than commercial-oriented, richer video designs negatively affect engagement (Table 11, Model 2a-2b). Similar interaction patterns do not emerge for semantic similarity across modes (Table 11, Model 1a; Model 1b), brand congruency in communication (Table 11, Model 3a; Model 3b), and message assertiveness (Table 11, Model 4a; Model 4b).

Table 10: Alternative Explanations: Between-Context Analysis

DV: Engagement	Model 1: Similarity across modes		Model 2: Type of Post		Model 3: Brand Congruity		Model 4: Assertiveness		Model 5: Full Model	
	Coeff.	***p***	**Coeff.**	***p***	**Coeff.**	***p***	**Coeff.**	***p***	**Coeff.**	***p***
Independent Variables:										
Video Design										
Image+Caption+TO	.088 (.108)	.416	.032 (.105)	.758	.008 (.105)	.937	-.024 (.106)	.818	-.030 (.121)	.802
Image+Caption+VO	.320 (.204)	.117	.247 (.204)	.227	.173 (.203)	.393	.160 (.203)	.431	-.059 (.229)	.796
Image+Caption+TO+VO	.229 (.117)	.051	.106 (.109)	.330	.071 (.108)	.512	.062 (.109)	.566	-.209 (.152)	.168
Green Claim:										
Yes	.004 (.138)	.974	.040 (.139)	.772	.012 (.139)	.933	-.012 (.139)	.933	.149 (.154)	.332
Interactions: Video Design * Green Claim										
Image+Caption+TO * Green Claim	-.314 (.167)	.059	-.333 (.167)	.047	-.320 (.168)	.057	-.301 (.168)	.073	-.450 (.178)	.011
Image+Caption+VO * Green Claim	-1.115 (.301)	.000	-1.172 (.303)	.000	-1.161 (.303)	.000	-1.137 (.302)	.000	-1.147 (.310)	.000
Image+Caption+TO+VO * Green Claim	-.233 (.158)	.139	-.280 (.158)	.077	-.261 (.158)	.099	-.232 (.159)	.145	-.478 (.172)	.005
Additional Controls:										
Log (Views)	.621 (.013)	.000	.628 (.013)	.000	.625 (.013)	.000	.625 (.013)	.000	.604 (.013)	.000
Video Controls:										
Log (Video Duration)	.445 (.314)	.156	.418 (.317)	.187	.384 (.317)	.227	.369 (.317)	.244	-.117 (.351)	.739
Log (Video Duration)^2	-.062 (.049)	.202	-.064 (.049)	.195	-.061 (.049)	.217	-.058 (.049)	.240	.048 (.057)	.396
Audial Controls										
Volume									.016 (.007)	.021
Pitch									-.000 (.000)	.080
Speech Rate									.203 (.050)	.000
Visual Controls										
Colorfulness									.051 (.307)	.869
Saturation									.904 (.242)	.000
Faces									-.114 (.109)	.294
Happy									.224 (.115)	.052
Sad									-.146 (.153)	.340
Surprise									-.164 (.413)	.691
Fear									.251 (.151)	.096
Angry									.286 (.220)	.192
Disgust									-.776 (.621)	.212
Verbal Controls										
Caption Hashtags									.011 (.014)	.426
Caption Mentions									-.167 (.080)	.037
Caption Emoji									-.042 (.024)	.079
Valence									-.528 (.225)	.019
Arousal									.331 (.288)	.250
Concreteness									-.002 (.001)	.035
Readability									-.010 (.004)	.019
Verbal Density									.020 (.026)	.440
Verbal Similarity	-.591 (.191)	.002							-.629 (.203)	.002
Visual-Text Similarity	40.072 (.977)	.000							4.885 (.991)	.000
Assertiveness							-.031 (.017)	.058	-.026 (.017)	.133
Informative (cf. commercial) goal			-.272 (.089)	.002					-.271 (.088)	.002
Brand Congruity					-22.920 (13.854)	.098			-26.715 (13.692)	.051
Fixed Effects									-.030 (.121)	.802
Time FE	Included		Included		Included		Included		Included	
Brand FE	Included		Included		Included		Included		Included	
Intercept	.302 (.682)	.658	1.453 (.643)	.024	15.234 (8.510)	.073	1.542 (.646)	.017	19.139 (8.440)	.023
N	1,256		1,256		1,256		1,256		1,225	
AIC	19,202.748		19,215.115		19,222.506		19,220.945		18,606.368	
Log likelihood	-9,535.374		-9,542.558		-9,547.253		-9,545.472		-9,215.184	

Notes: Standard errors are in parentheses. TO = text overlay, VO = voice over

Table 11: Alternative Explanations: Within-Context Analysis

DV: Engagement	**Model 1a: Content Similarity**		**Model 1b: Content Similarity**		**Model 2a: Type of Post**		**Model 2b: Type of Post**		**Model 3a: Brand Congruity**		**Model 3b: Brand Congruity**		**Model 4a: Assertiveness**		**Model 4b: Assertiveness**	
	Coeff.	***p***	**Coeff.**	***p***	**Coeff.**	***p***	**Coeff.**	***p***	**Coeff.**	***p***	**Coeff.**	***p***	**Coeff.**	***p***	**Coeff.**	***p***
Independent Variables:																
Video Design																
Image+Caption+TO	-.522 (.174)	.003	-1.949 (1.192)	.102	-.587 (.167)	.000	.941 (.507)	.063	-.592 (.166)	.000	-.259 (1.089)	.812	-.592 (.168)	.000	-.623 (.189)	.001
Image+Caption+VO	-1.433 (.272)	.000	4.185 (4.755)	.379	-1.500 (.272)	.000	.160 (.674)	.813	-1.520 (.270)	.000	.642 (1.368)	.639	-1.519 (.270)	.000	-1.409 (.377)	.000
Image+Caption+TO+VO	-.876 (224)	.000	-.876 (1.340)	.513	-.952 (.219)	.000	.588 (.521)	.259	-.963 (.218)	.000	-.051 (1.084)	.963	-.963 (.219)	.000	-1.141 (.236)	.000
Alternative Explanations																
Informative (cf. commercial) goal					-.069 (.138)	.535	2.567 (.850)	.003								
Brand Congruity									.096	.996	-4.064 (19.188)	.832				
Assertiveness													.002	.951	-.055 (.054)	.314
Verbal Similarity	-.489 (277)	.078	-1.211 (2.477)	.625												
Visual-Text Similarity	1.828 (1.356)	.179	.099 (3.809)	.979												
Interactions Effects:																
Image+Caption+TO * Visual-Verbal Similarity * Verbal Similarity			-16.853 (14.688)	.256												
Image+Caption+VO * Visual-Verbal Similarity * Verbal Similarity			70.414 (60.686)	.246												
Image+Caption+TO+VO*Visual-Verbal Similarity*Verbal Similarity			–	–												
Image+Caption+TO * Informative							-2.633 (.857)	.002								
Image+Caption+VO * Informative							-2.924 (1.030)	.005								
Image+Caption+TO+VO * Informative							-2.712 (.864)	.002								
Image+Caption+TO * Brand Congruity											-.435 (1.279)	.734				
Image+Caption+VO * Brand Congruity											-2.608 (1.609)	.105				
Image+Caption+TO+VO * Brand Congruity											-1.130 (1.292)	.382				
Image+Caption+TO * Assertiveness															-.005 (.067)	.935
Image+Caption+VO * Assertiveness															-.028 (.144)	.848
Image+Caption+TO+VO * Assertiveness															.144 (.067)	.031
Controls																
Log(Views)	Included		Included		Included		Included		Included		Included		Included		Included	
Video Controls	Included		Included		Included		Included		Included		Included		Included		Included	
Audial Controls	Included		Included		Included		Included		Included		Included		Included		Included	
Visual Controls	Included		Included		Included		Included		Included		Included		Included		Included	
Verbal Controls	Included		Included		Included		Included		Included		Included		Included		Included	
Time and Brand Fixed Effects	Included		Included		Included		Included		Included		Included		Included		Included	
Intercept	3.753 (1.449)	.010	4.107 (1.602)	.010	4.478 (1.402)	.001	3.239 (1.450)	.025	4.463 (11.662)	.702	6.917 (5.956)	.562	4.519 (1.400)	.001	4.837 (1.405)	.001
N	614		614		614		614		614		614		614		614	
AIC	9,145.433		9,156.004		9,147.457		9,141.213		9,145.841		9,148.825		9,147.838		9,145.001	
Log likelihood	-4,492.716		-4,490.002		-4,494.728		-4,488.606		-4,494.921		-4,493.413		-4,494.919		-4,490.501	

Notes: Standard errors are in parentheses. TO = text overlay, VO = voice over

Robustness

We test in several ways the robustness of our model specification and measurements. See Table 12 for an overview, while details are in Web Appendix B.

First, we adopted a caliper of .1 (Table WB3: Model 1), and a caliper of .01 (Table WB3: Model 2) to allow for a cleaner comparison between green and non-green ads. We show that results remain consistent even when matching green-oriented posts with more similar non-green counterparts (i.e., when using a smaller caliper to increase matching precision).

Second, one could wonder whether the results are somehow driven by the modeling approach used. Therefore, we adopted an ordinary least squares (OLS) regression with a log-transformed dependent variable as an alternative modelling approach and find consistent results (Table WB3: Model 3).

Third, one could wonder whether the results hold for any measure of engagement (like, comment, share), given that measures such as like, comments and shares could require different effort and might be driven by different reasons (Swani and Labrecque 2020). Therefore, we investigated the effects on likes, comments, and shares, separately. Results after matching remain consistent for any measure of engagement (i.e., like, comment, share; Table WB3: Model 4, Model 5, and Model 6).

Fourth, one could wonder whether the effect is driven by the way our focal variable is measured. Therefore, we considered two specifications alternative to our dummy variable, namely a green word count ratio (Table WB3, Model 7) and a continuous measure of the post's green orientation (Table WB3, Model 8; see "Alternative Measures of Message Green Orientation" in Web Appendix B for details on how these measures have been computed). Results hold for any of these alternative measures.

Table 12: Robustness Checks

	What We Test?	How We Test It
Modelling		
Caliper	Are the results robust to different caliper thresholds when matching green and non-green posts (i.e., with stricter criteria for matching green and non-green posts)?	Results hold for: Caliper = .1 (b_{TO} = -.453, SE = .178, t = 2.54, p = .011; b_{VO} = -1.139, SE = .310, t = 3.67, p = .000; b_{VO+TO} = -.476, SE = .172, t = 2.77, p = .006); Caliper = .01 (b_{TO} = -.558, SE = .189, t = 2.95, p = .003; b_{VO} = -1.336, SE = .326, t = 4.10, p = .000; b_{VO+TO} = -.529, SE = .182, t = 2.91, p = .004).
Alternative Approach	Are the results driven by the modeling approach used?	Results hold if using OLS regression with log-transformed DV (b_{TO} = -.474, SE = .180, t = 2.63, p = .008; b_{VO} = -.873, SE = .297, t = 2.94, p = .003; b_{VO+TO} = -.342, SE = .170, t = 2.01, p = .045).
Alternative DV	Could the effect be driven by the dependent variable used?	Results hold for: Like (b_{TO} = -.397, SE = .178, t = 2.23, p = .026; b_{VO} = -1.134, SE = .311, t = 3.64, p = .000; b_{VO+TO} = -.427, SE = .172, t = 2.48, p = .013); Comment (b_{TO} = -.684, SE = .227, t = 3.01, p = .003; b_{VO} = -1.146, SE = .397, t = 2.89, p = .004; b_{VO+TO} = -.834, SE = .218, t = 3.83, p = .000); Share (b_{TO} = -.211, SE = .236, t = .89, p = .370; b_{VO} = -.899, SE = .446, t = 1.99, p = .044; b_{VO+TO} = -.430, SE = .233, t = 1.85, p = .065).
Measurement		
Alternative measure of the message's green orientation	Could the effect be driven by how the green orientation of the message is measured?	Green message orientation measured using a green wordcount ratio (b_{TO} = -5.984, SE = 1.994, t = 3.00, p = .003; b_{VO} = -18.899, SE = 5.999, t = 3.17, p = .002; b_{VO+TO} = -4.141, SE = 2.162, t = 1.92, p = .056), or a continuous variable (b_{TO} = -.127, SE = .066, t = 1.92, p = .056; b_{VO} = -.272, SE = .082, t = 3.32, p = .001; b_{VO+TO} = -.097, SE = .063, t = 1.54, p = .123).

General Discussion

Marketers increasingly use videos on social media to foster brand awareness, promote products and lifestyles, and inform consumers about their production processes. But while videos are ubiquitous on social media, how the interplay among modes (i.e., visual content, text, audio) within videos shape consumer engagement is unclear, particularly when the context of sustainability is considered. Given the urge and societal relevance of improving the effectiveness of communication around sustainability (European Commission 2023; Federal Trade Commission 2022), this work blends automated video, audio, and text analyses with machine learning tools on a dataset of 5749 TikTok videos posted by top European brands to shed light on how different video designs – ranging from videos featuring only baseline elements (i.e., visual content and caption) to those enriched with text overlays and voice-over – a) shape consumer engagement of brand messages on social media b) whether, how and why the effect might differ in the context of green (vs. non-green) communication.

The results demonstrate that, overall, for branded content a richer video design tends to have a positive impact on social media engagement on TikTok. In other words, a video-based post combining visual content and caption with text overlays, voice-over, or both, is associated with more likes, comments, or shares. However, our findings also show that when a video features a green claim, implementing these verbal cues backfires. Specifically, consumers are less likely to engage with green-oriented videos that implement a voice-over, text overlay, or a combination of both on top of the baseline elements (i.e., visuals and caption). The effect persisted in the field even accounting for a variety of alternative explanations such as the semantic similarity across modes, goal of the post, message assertiveness, brand congruency in communication style of posts including (vs. not including) a green claim, suggesting that none of these explanations fully explain our results. It might be reasonable to think that the negative effect of richer video designs in the context of green-oriented communication may reflect heightened consumers' skepticism triggered by the company's perceived stronger persuasive efforts when promoting sustainability initiatives through multiple modes. We also examined whether these factors could account for differences in how video design shapes engagement within green-oriented communication. Our findings suggest that richer video designs tend to negatively affect engagement when the post's goal orientation is more informative than commercial.

Theoretical Contributions

This work makes several contributions. First, as an answer to the calls for research on multimodality (Grewal et al. 2022; Grewal, Gupta, and Hamilton 2021; Packard and Berger 2024) it sheds light on how the interplay among specific features across modes in videos shapes consumer behavior. While prior work has tended to consider how textual (e.g., Cascio Rizzo et al. 2024; Zhou et al. 2021), audial (e.g., Chang, Mukherjee and Chattopadhyay 2023; Zhang, Zhou, and Lee 2024), and visual elements (e.g., Yang, Zhang, and Zhang 2024; Wang, Ding, and Hu 2025) *in isolation* affect consumer responses, there has been less attention on whether and how multiple elements across modes should be *combined* in video design for effective marketing communication on social media (Holliday et al. 2023; Lapresta-Romero et al. 2024; Rajaram and Manchanda 2025). Building on research in the fields of social media marketing communication (e.g., Li and Xie 2020), educational psychology (e.g., Mayer 2021), and neuroscience (e.g., Rutten et al. 2019), we fill this gap and contribute to the literature on video marketing (e.g., Chang, Mukherjee, and Chattopadhyay 2023; Rajaram and Manchanda 2025) by examining how enriching video design with two key elements such as text overlay and voice-over shape consumer engagement of brand-generated content. Our findings indicate that, overall, short brand-generated videos benefit from the use of text overlay and voice-over, in line with prior evidence concerning static content that show a positive effect of the joint use of multiple modes (e.g., Adam 2025; Lapresta-Romero et al. 2024; Villarroel et al. 2019). However, when the brand promotes green-oriented messages, implementing a text overlay, voice-over, or a combination of both, on top of the baseline elements (i.e., visuals and caption) is detrimental to engagement. Our focus on brand-generated content further differentiates our research from other work on the interplay among multiple modes in videos, which examine influencer-generated content (Rajaram and Manchanda 2025; Holliday et al. 2023). The two contexts differ substantially: while brand-generated content is designed to increase brand awareness (Colicev et al. 2018), support content diffusion (Hollebeek, Glynn, and Brodie 2014; Tellis et al. 2019), foster consumer-brand relationships (Van Doorn et al. 2010), and increase sales (Kumar et al. 2016), influencer-generated content is perceived as more trustworthy in engaging audiences (Leung et al. 2022). This may fundamentally alter how consumers respond to different video design strategies, particularly when they perceive a lack of authenticity – an issue that becomes even more critical in the context of sustainability (Pittman and Abell 2021).

Second, we contribute to the literature on communication around environmental sustainability. Despite the importance for brands of communicating sustainability on social media (Bharadwaj, Naik, and Nath 2022), literature falls short in investigating how brands can leverage simultaneously multiple modes to effectively promote green products and lifestyles on social media (Fernandez, Hartmann, and Apaolaza 2022). While a burgeoning stream of research focused on how to leverage visual cues to make the message look “green” (e.g., Parguel, Benoit-

Moreau, and Russell 2015; Segev, Fernandes and Hong 2016) or textual information to foster sustainable behaviors (e.g., Kronrod, Amir, and Grinstein 2012; Kronrod et al. 2023; Olsen, Slotegraaf, and Chandukala 2014), less attention has been devoted to how the joint presence of visual and verbal elements shape green communication's impact on social media (Barberá-Tomás et al. 2019; Gonzalez-Arcos et al. 2021), particularly in social media videos (Zhu, Cheng, and Wang 2025). This work contributes to this stream of research by unveiling contingencies within green-oriented video communication that suggest a careful use of modes, or social media engagement for green-oriented messages might likely suffer. Moreover, in demonstrating how video design shapes green-oriented communication's impact, we also explore potential underlying mechanisms behind these effects.

Third, from a methodological standpoint, we demonstrate how combining automated video, audio, and text analysis tools with machine learning techniques can provide novel insights into consumer behavior (Grewal, Gupta, and Hamilton 2021; Wang, Bendle, and Pan 2024). While prior marketing research on videos has typically examined key constructs *within a single mode* (e.g., arousal via text *or* speech; Cascio Rizzo et al. 2024), we adopt a multimodal approach and measure several constructs *across modes* (e.g., arousal across caption, text overlays and speech). By doing so, we provide a deeper conceptual and empirical perspective on how these measures can affect consumer behavior on social media. In addition, computer vision has been primarily applied to static (e.g., Dang et al. 2025; Li and Xie 2020) rather than to dynamic content (e.g., Zhou et al. 2021). We move beyond and apply computer vision to video content to measure a key construct such as the semantic similarity between visuals and verbal cues.

Practical Implications

These findings have clear practical implications. From companies and marketers to institutions and policy makers, everyone wants to improve the effectiveness of their green-oriented communication. Our findings indicate that a "less is more" approach that avoids the implementation of text overlays and voice-over in videos is a solution to avoid lower engagement for videos that promote a green lifestyle or consumption. Specifically, our findings indicate that, in case of green-oriented messages, including text overlays is associated with about 43 fewer likes, comments, or shares, whereas implementing voice-over is associated with an even stronger drop resulting in about 77 less likes, comments, or shares. Finally, combining text overlays and voice-over results in a about 57 fewer likes, comment, or shares, compared to green videos that feature only the baseline elements (visual content and caption). Therefore, we provide evidence that beyond *what* is said, also *how* it is said can shape effectiveness of green-oriented communication. Compared to more resource-intensive strategies, relying primarily on visuals and captions seems potentially easier and more cost-effective. Despite these benefits, only 8.77% of green-oriented videos in our sample presented only visual content and caption, suggesting a missed opportunity in current practice.

Importantly, while these findings suggest that a "less is more" approach can enhance the effectiveness of communication around sustainability, such strategy may have a double edge. The same design choices that make content more engaging may also be leveraged to increase the persuasiveness of messages that are actually misleading or deceptive, thereby fostering actually greenwashing. As a consequence, these actionable insights call for companies to act responsibly, using design choices not as a mere persuasive shortcut, but as a means to convey in a more effective way authentic environmental commitment.

Limitations and Future Research

Videos play a key role in contemporary marketing communication on social media, but there is surprisingly little empirical understanding of how the joint use of multiple elements across modes in videos shape consumer responses. While this work helps fill this gap, it inevitably leaves room for further inquiry.

First, our analysis focuses on how the mere presence in videos of text overlays and voice-over in combination with visual content and caption shape engagement. However, we do not account for specific properties of these verbal cues – for instance, text overlay transience and changes in color, size, and centrality (Grzenkowicz and Wildfeuer 2025), or voice-over features, such as gender, or timing. Future research could focus on how these video design features interplay with each other and shape user responses.

Second, although 93% of TikTok users consume content with sound on (TikTok for Business 2022), our field data do not capture whether the audio component was actually on during viewing. As a result, the estimated effect of voice-over should be interpreted with caution, as some users may have watched the video with the sound turned off. This limitation implies that the observed effect may reflect both auditory and non-auditory elements of video design. Future research could experimentally isolate the causal impact of voice-over in videos on engagement, as well as the impact of its specific properties (Hildebrand et al. 2020). More broadly, future research might examine whether the effectiveness of multimodal cues depends on viewing behaviors. For instance, highly involved users might be more likely to process multiple modes, whereas less involved users may rely primarily on visual cues and decide to switch off the audio, thereby affecting the effects of richer video designs on engagement.

Third, future research could examine how video design affects other key measures of social media engagement. In this research, we relied on the count of likes, comments, and shares as dependent variables, partly because they are well-established and relevant measures of engagement (Liadeli, Sotgiu, and Verlegh 2023), but also because of data collection limitations. Future work could consider the effect of video design on sustained attention (Berger, Moe, and Schweidel 2023). This metric might be particularly relevant in the context of green-oriented communication, where the inherent complexity of sustainability topics requires that viewers remain engaged long enough for the message to effectively inform, educate, or promote sustainable behavior (White, Habib, and Hardisty 2019).

Fourth, we focus on brand-generated video content from large companies, which typically have more resources to allocate to communication practices, broader audience reach, and higher visibility. As such, these firms are particularly influential in shaping consumer behavior, marketing communication trends and public discourse related to sustainability. However, whether the effects documented here extend to smaller companies, influencers, and user-generated content remains an open question, particularly in the context of green-oriented communication. For instance, while richer video designs by large firms promoting eco-friendly products or sustainable behaviors may inadvertently trigger persuasion knowledge and foster skepticism and perceptions of greenwashing, the same design from green content creators or consumers, which are typically perceived as more authentic and less profit-oriented, may be less likely to evoke such negative responses (Hassler et al. 2025).

Fifth, while our dataset covers a wide range of industries and products, green-oriented posts are mainly drawn from companies belonging to the automotive and energy sectors. This is partly because these industries are heavily involved in emission-reduction activities, but also due to our sampling criteria, which rely on companies listed in Global Fortune 500. Future research could examine whether our findings extend also to other high-impact industries such as food, fashion, and beauty, where sustainability strongly influences consumer choices (Brand Finance 2025).

Finally, the empirical analysis focuses on TikTok, an entertainment-driven, video-based platform with its own specific algorithm, platform affordances, and audience, whose mission is to "inspire creativity and bring joy[3]". As green-oriented content is often perceived as instructive and normative (Bolderdijk et al. 2013), it may contrast with users' expectations for playful, entertaining content. This could influence how users react to elements such as text overlays or voice-overs, which may inadvertently reinforce the prescriptive tone of the message, with, in turn, consequences on engagement. Future research could test the generalizability of these findings across platforms.

5 CONCLUSION

In conclusion, this work demonstrates how video design can shape the effectiveness of brand generated content on social media, with a particular emphasis on how to improve green-oriented communication's impact, a topic that is at the core of growing societal and governmental concern. Specifically, our findings indicate that in the field of green-oriented communication a "less is more" approach that avoids the implementation of text overlays and voice-over in videos is a solution to achieve higher levels of engagement on social media. In an era where sustainability is constantly in the spotlight, we provide evidence that *how* brands communicate is just as crucial as *what* they communicate, along with actionable insights on how to make such communication more effective.

[3] https://www.tiktok.com/about

6 REFERENCES

Albers, Fabian, Melanie Trypke, Ferdinand Stebner, Joachim Wirth, and Jan L. Plass (2023), "Different Types of Redundancy and Their Effect on Learning and Cognitive Load," British Journal of Educational Psychology, 93(2), 339–352.

Adam, Zitian (2025), "How Do Image Captions Drive Consumer Engagement on Social Media?," *Journal of Interactive Marketing*, 10949968251352408.

Al-Qershi, Osamah, Junbum Kwon, Shuning Zhao, and Zhaokun Li (2022), "Predicting Crowdfunding Success with Visuals And Speech in Video Ads and Text Ads," *European Journal of Marketing*, 56(6), 1610–1649.

Amatulli, Cesare, Matteo De Angelis, Alessandro Maria Peluso, Isabella Soscia, and Gianluigi Guido (2019), "The Effect of Negative Message Framing on Green Consumption: An Investigation of the Role of Shame," *Journal of Business Ethics*, 157(4), 1111–1132.

Atkinson, Lucy, and Sonny Rosenthal (2014), "Signaling the Green Sell: The Influence of Eco-Label Source, Argument Specificity, and Product Involvement on Consumer Trust," *Journal of Advertising*, 43 (1), 33–45.

Apple, William, Lynn A. Streeter, and Robert M. Krauss (1979), "Effects of Pitch and Speech Rate on Personal Attributions," *Journal of Personality and Social Psychology*, 37 (5), 715–27.

Banerjee, Subhabrata, Charles S. Gulas, and Easwar Iyer (1995), "Shades of Green: a Multidimensional Analysis of Environmental Advertising," *Journal of Advertising*, 24(2), 21–31.

Barberá-Tomás, David, Itziar Castelló, Frank G. de Bakker, and Charlene Zietsma (2019),"Energizing Through Visuals: How Social Entrepreneurs Use Emotion-Symbolic Work for Social Change," *Academy of Management Journal*, 62 (6), 1789–817.

Bateman John (2022), “Multimodality, where next? – Some meta-methodological considerations,” *Multimodality and Society*, 2(1) 41–63.

Bateman John , Janina Wildfeuer, and Tuomo Hiippala (2017), *Multimodality. Foundations, Research and Analysis—A Problem-Oriented Introduction*, Berlin; Boston: De Gruyter Mouton.

Batra, Rajeev, and Kevin Lane Keller (2016),“Integrating Marketing Communications: New Findings, New Lessons, and New Ideas,” *Journal of Marketing*, 80 (6), 122–45.

Belin, Pascal, Shirley Fecteau, and Catherine Bédard (2004), “Thinking the Voice: Neural Correlates of Voice Perception,” *Trends in Cognitive Sciences*, 8 (3), 129–35.

Berger, Jonah and Katherine L Milkman (2012),“What Makes Online Content Viral?” *Journal of Marketing Research*, 49 (2), 192–205.

Berger, Jonah, Wendy W. Moe, and David Schweidel (2023),“What Holds Attention? Linguistic Drivers of Engagement,” *Journal of Marketing*, 87 (5), 793–809.

Bharadwaj Neeraj, Prasad A. Naik, and Pravin Nath (2022), “Sustainability Communications and Corporate Brand Associations,” *Journal of Sustainable Marketing*, 3(1), 41–52.

Bolderdijk, Jan Willem, Madelijne Gorsira, Kees Keizer, Linda Steg (2013), “Values Determine the (In)Effectiveness of Informational Interventions in Promoting Pro-Environmental Behavior,” *Plos One*, 8(12), 1–7.

Brand Finance (2025), “Sustainability Perceptions Index 2025”, (accessed 20 July, 2025), from https://brandirectory.com/reports/sustainability-perceptions-index.

Brasel, S. Adam, and James Gips (2014), “Enhancing Television Advertising: Same-Language Subtitles Can Improve Brand Recall, Verbal Memory, And Behavioral Intent,” *Journal of the Academy of Marketing Science*, 42(3), 322–336.

Brehm, Jack W. (1966), *A Theory of Psychological Reactance*. New York: Academic Press.

Bucher, Hans-Juergen (2025), "Insights Into the Black Box of Multimodal Meaning-Making: Investigating the Reception of Multimodality Empirically," *Visual Communication*, 24(3), 543–569.

Campbell, Margaret C. and Amna Kirmani (2000), "Consumers' Use of Persuasion Knowledge: The Effects of Accessibility and Cognitive Capacity on Perceptions of an Influence Agent," *Journal of Consumer Research*, 27 (1), 69–83.

Carlson, Les, Stephen J. Grove, and Norman Kangun (1993), "A Content Analysis of Environmental Advertising Claims: A Matrix Method Approach," *Journal of Advertising*, 22 (3), 27–39.

Casaburi Ivana (September 18, 2024), "Brand Sustainability Starts with Strategic Communications," https://dobetter.esade.edu/en/sustainability-communications.

Cascio Rizzo, Giovanni Luca, Jonah Berger, Matteo De Angelis, and Rumen Pozharliev (2023), "How Sensory Language Shapes Influencer's Impact," *Journal of Consumer Research*, 50 (4), 810–25.

Cascio Rizzo, Giovanni Luca, 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) 107–128.

Ceylan, Gizem, Kristin Diehl, and Davide Proserpio (2024), "Words Meet Photos: When and Why Photos Increase Review Helpfulness," *Journal of Marketing Research*, 61(1), 5–26.

Ceylan, Gizem, Kristin Diehl, and Wendy Wood (2024), "From Mentally Doing to Actually Doing: A Meta-Analysis of Induced Positive Consumption Simulations," *Journal of Marketing*", 88(2), 21–39.

Chang, Hannah H., Anirban Mukherjee, and Amitava Chattopadhyay (2023), “More Voices Persuade: The Attentional Benefits of Voice Numerosity,” *Journal of Marketing Research*, 60(4), 687–706.

Charest, Ian, Cyril R. Pernet, Guillaume A. Rousselet, Ileana Quiñones, Marianne Latinus, Sarah Fillion-Bilodeau, Jean-Pierre Chartrand, and Pascal Belin (2009),“Electrophysiological Evidence for an Early Processing of Human Voices,” *BMC Neuroscience*, 10 (1), 127.

Chattopadhyay, Amitava, Darren W. Dahl, Robin J.B. Ritchie, and Kimary N. Shahin (2003),“Hearing Voices: The Impact of Announcer Speech Characteristics on Consumer Response to Broadcast Advertising,” *Journal of Consumer Psychology*, 13 (3), 198–204.

Chen, Yu-Shan and Ching-Hsun Chang (2013), “Greenwash and Green Trust: The Mediation Effects of Green Consumer Confusion and Green Perceived Risk,” *Journal of Business Ethics*, 114 (3), 489–500.

Chernev, Alexander and Sean Blair (2020),“When Sustainability is not a Liability: The Halo Effect of Marketplace Morality,” *Journal of Consumer Psychology*, 31 (3), 551–69.

Childers, Terry L., and Michael J. Houston (1984), “Conditions for a Picture-Superiority Effect on Consumer Memory,” *Journal of Consumer Research*, 11 (2), 643–654.

Colicev, Anatoli, Ashwin Malshe, Koen Pauwels, and Peter O’Connor (2018),“Improving Consumer Mindset Metrics and Shareholder Value Through Social media: The Different Roles of Owned and Earned Media,” *Journal of Marketing*, 82 (1), 37–56.

Csikszentmihalyi, Mihaly (2000), “The Costs and Benefits of Consuming,” *Journal of Consumer Research*, 27 (2), 267–72.

Daft, Richard L., and Robert H. Lengel (1986), “Organizational Information Requirements, Media Richness and Structural Design,” *Management Science*, 32(5), 554–571.

Dang, Chu (Ivy), Canice M. C. Kwan, Jayson S. Jia, and Yang Shi (2025), “When Words Meet Visuals: How Content Composition Drives Social Media Engagement for Marketer-Generated Content,” *Journal of Marketing Research*, 00222437251373042.

de Visser-Amundson, Anna, John Peloza, and Mirella Kleijnen (2021). How Association with Physical Waste Attenuates Consumer Preferences for Rescue-Based Food,” *Journal of Marketing Research*,” 58(5), 870–887.

DemandSage. (2025), “How Many People Use TikTok (Active Users Stats 2025),” (accessed September 29, 2025), from https://www.demandsage.com/tiktok-user-statistics/.

Edinger-Schons, Laura, Jenni Sipila ̈, Sankar Sen, Gina Mende, and Jan Wieseke (2018), “Are Two Reasons Better Than One? The Role of Appeal Type in Consumer Responses to Sustainable Products,” *Journal of Consumer Psychology*, 28 (4), 644–64.

Eisend, Martin, and Farid Tarrahi (2022), “Persuasion Knowledge in the Marketplace: A Meta-Analysis,” *Journal of Consumer Psychology*, 32(1), 3–22.

European Commission (2023). “Proposal for a Directive of the European Parliament and of the Council on Substantiation and Communication of Explicit Environmental Claims (Green Claims Directive). COM(2023) 166 final”, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52022SC0085.

European Environment Agency (2024), “Trends and Projections in Europe 2024,” Research Report No. 11/2024, EEA.

Farace, Stefania, Francisco Villarroel Ordenes, Dennis Herhausen, Dhruv Grewal, and Ko De Ruyter (2025), "Standing Out While Fitting in: Visual Design of Text Overlays in Social Media Communication," *Journal of Marketing*, 00222429251322773.

Federal Trade Commission (2022), "FTC Seeks Public Comment on Potential Updates to its 'Green Guides' for the Use of Environmental Marketing Claims, press release (December 14), https://www.ftc.gov/news-events/news/press-releases/2022/12/ftc-seeks-public-comment-potential-updates-its-green-guides-use-environmental-marketing-claims.

Fennis, Bob M., Marieke A. Adriaanse, Wolfgang Stroebe, and Bert Pol (2011), "Bridging the Intention–Behavior Gap: Inducing Implementation Intentions Through Persuasive Appeals," *Journal of Consumer Psychology*, 21 (3), 302–11.

Fernandez, Paula, Patrick Hartmann, and Vanessa Apaolaza-Ibánez (2022), "What Drives CSR Communication Effectiveness on Social Media? A Process-Based Theoretical Framework and Research Agenda," *International Journal of Advertising*, 41(3), 385–413.

Forbes (2024a), "Understanding the Power of Video for Marketing", (accessed June 26, 2025), from https://www.forbes.com/councils/forbesagencycouncil/2024/03/07/understanding-the-power-of-video-for-marketing/.

Forbes (2024b). "Short-Form Video Content: Capturing Attention in the Digital Age," (accessed June 27, 2025), from https://www.forbes.com/councils/forbesagencycouncil/2024/03/19/short-form-video-content-capturing-attention-in-the-digital-age/.

Friestad, Marian and Peter Wright (1994), "The Persuasion Knowledge Model: How People Cope with Persuasion Attempts," *Journal of Consumer Research*, 21 (1), 1–31.

Gershoff, Andrew D. and Judy K. Frels (2014), "What Makes It Green? The Role of Centrality of Green Attributes in Evaluations of the Greenness of Products," *Journal of Marketing*, 79 (1), 97–110.

Ginder, Whitney, Wi-Suk Kwon, and Sang-Eun Byun (2021),“Effects of Internal–External Congruence-Based CSR Positioning: An Attribution Theory Approach,” *Journal of Business Ethics*, 169 (2), 355–69.

Goldstein, Noah J., Robert B. Cialdini, and Vladas Griskevicius (2008), “A Room with a Viewpoint: Using Social Norms to Motivate Environmental Conservation in Hotels,” *Journal of Consumer Research*, 35 (3), 472–82.

Gonzalez-Arcos, Claudia, Alison M. Joubert, Daiane Scaraboto, Rodrigo Guesalaga, and Jörgen Sandberg (2021), “‘How Do I Carry All This Now?’ Understanding Consumer Resistance to Sustainability Interventions,” *Journal of Marketing*, 85 (3), 44–61.

Graf, Laura K.M., Stefan Mayer, and Jan R. Landwehr (2018), “Measuring Processing Fluency: One Versus Five Items,” *Journal of Consumer Psychology*, 28 (3), 393–411.

Grewal, Rajdeep, Sachin Gupta, and Rebecca Hamilton (2021), “Marketing Insights from Multimedia Data: Text, Image, Audio, and Video,” *Journal of Marketing Research*, 58 (6), 1025–1033.

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), 224–240.

Grzenkowicz, Maciej, and Janina Wildfeuer. "Addressing TikTok’s Multimodal Complexity: A Multi-Level Annotation Scheme for the Audio-Visual Design of Short Video Content,” *Digital Scholarship in the Humanities*, 1–24.

Hagtvedt, Henrik (2022),“A Brand (New) Experience: Art, Aesthetics, and Sensory Effects,” *Journal of the Academy of Marketing Science*, 50 (3), 425–8.

Hartmann, Patrick, and Vanessa Apaolaza-Ibánez (2009), “Green Advertising Revisited”, *International Journal of Advertising*, 28(4), 715–739.

Hassler, Corinne M. K., Martin Mende, Maura L. Scott, and Lisa E. Bolton (2025), “The Prosocial Ambassador Effect: Adopting an Ambassador Role Increases Sustainable Behavior,” *Journal of Marketing*, 89 (1), 19–38.

Heckler, S.E. and T.L. Childers (1992),“The Role of Expectancy and Relevancy in Memory for Verbal and Visual Information: What Is Incongruity?” *Journal of Consumer Research*, 18 (4), 475–92.

Heckman, James J., Hidehiko Ichimura, and Petra E. Todd. "Matching as an Econometric evaluation estimator: Evidence from Evaluating a Job Training Programme." *The Review of Economic Studies*,” 64(4), 605–654.

Heckman, James J., Hidehiko Ichimura, and Petra E. Todd. (1998), “Matching as an Econometric Evaluation Estimator,” *The Review of Economic Studies*, 65(2), 261–294.

Hildebrand, Christian, Fotis Efthymiou, Francesc Busquet, William Hampton, Donna Hoffman, and Novak Thomas (2020), “Voice Analytics in Business Research: Conceptual Foundations, Acoustic Feature Extraction, and Applications,” *Journal of Business Research*, 121, 364–74.

Holiday, Steven, Jameson L. Hayes, Haseon Park, Yuanwei Lyu, and Yang Zhou (2023), “A Multimodal Emotion Perspective on Social Media Influencer Marketing: The Effectiveness of Influencer Emotions, Network Size, and Branding on Consumer Brand Engagement Using Facial Expression and Linguistic Analysis,” *Journal of Interactive Marketing*, 55 (4), 414–39.

Hollebeek, Linda D., Mark S. Glynn, and Roderick J. Brodie (2014), “Consumer Brand Engagement in Social Media: Conceptualization, Scale Development and Validation,” *Journal of Interactive Marketing*, 28(2), 149–165.

Horowitz, Seth S. (2012), *The Universal Sense: How Hearing Shapes the Mind*. Bloomsbury.

Hovy, Dirk, Shiri Melumad, and J. Jeffrey Inman (2021),"Wordify: A Tool for Discovering and Differentiating Consumer Vocabularies," *Journal of Consumer Research*, 48 (3), 394–414.

Hull, Glynda A., and Mark Evan Nelson (2005), "Locating the Semiotic Power of Multimodality," *Written Communication*, 22(2), 224–261.

Kanuri, Vamsi K., Yixing Chen, and Shrihari Sridhar (2018), "Scheduling Content on Social Media: Theory, Evidence, and Application," *Journal of Marketing*, 82 (6), 89–108.

Kanuri, Vamsi K., Christian Hughes, and Brady T. Hodges (2024), "Standing out from the Crowd: When and Why Color Complexity in Social Media Images Increases User Engagement," *International Journal of Research in Marketing*, 41(2), 174-193.

Kilbourne, William E. (2004), "Sustainable Communication and the Dominant Social Paradigm: Can They be Integrated?," *Marketing Theory*, 4(3), 187–208.

Kress, Gunther R. (2010), *Multimodality: A Social Semiotic Approach to Contemporary Communication.* Routledge.

Kronrod, Ann, Amir Grinstein, and Luc Wathieu (2012), "Go Green! Should Environmental Messages Be So Assertive?" *Journal of Marketing*, 76 (1), 95–102.

Kronrod, Ann, Anat Tchetchik, Amir Grinstein, Lior Turgeman, and Vered Blass (2023), "Promoting New Pro-Environmental Behaviors: The Effect of Combining Encouraging and Disouraging Messages," *Journal of Environmental Psychology*, 86, 101945.

Kumar, Ashish, Ram Bezawada, Rishika Rishika, Ramkumar Janakiraman, and P.K. Kannan (2016), "From Social to Sale: The Effects of Firm-Generated Content in Social Media on Customer Behavior," *Journal of Marketing*, 80 (1), 7–25.

Labrecque, Lauren I., Vanessa M. Patrick, and George R. Milne (2013), "The Marketers' Prismatic Palette: A Review of Color Research and Future Directions," *Psychology & Marketing*, 30 (2), 187–202.

Lapresta-Romero Sara, Larissa Becker, Blanca Hernández-Ortega, Harri Terho, and José L. Franco (2024), "Getting the Recipe Right: How Content Combinations Drive Social Media Engagement Behaviors," *Journal of Interactive Marketing*, 1–22.

Leonidou, Leonidas C., Constantinos N. Leonidou, Dayananda Palihawadana, and Magnus Hultman (2011), "Evaluating the Green Advertising Practices of International Firms: A Trend Analysis," *International Marketing Review*, 28 (1), 6–33

Leonidou, Constantinos N., and Dionysis Skarmeas (2017), "Green Shades of Green: Causes and Consequences of Green Skepticism," *Journal of Business Ethics*, 144(2), 401–415.

Leung, Fine F., Flora F. Gu, Yiwei Li, Jonathan Z. Zhang, and Robert W. Palmatier (2022), "Influencer Marketing Effectiveness," *Journal of Marketing*, 86 (6), 93–115.

Li, Xi, Mengze Shi, and Xin Wang (2019),"Video Mining: Measuring Visual Information Using Automatic Methods," *International Journal of Research in Marketing*, 36 (2), 216–31.

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.

Liadeli, Georgia, Francesca Sotgiu, and Peetr W. Verlegh (2023),"A Meta-Analysis of the Effects of Brands' Owned Social Media on Social media Engagement and Sales," *Journal of Marketing*, 87 (3), 406–27.

Lin, Yan, Dai Yao, and Xingyu Chen (2021),"Happiness Begets Money: Emotion and Engagement in Live Streaming," *Journal of Marketing Research*, 58 (3), 417–38.

Lin, Ying-Ching, and Chiu-chi Angela Chang (2012), "Double Standard: The Role of Environmental Consciousness in Green Product Usage," *Journal of Marketing*, 76 (5), 125–34.

Liu, Xuan, Savannah Wei Shi, Thales Teixeira, & Michel Wedel (2018), "Video Content Marketing: The Making of Clips," *Journal of Marketing*, 82(4), 86–101.

Lowe, Michael L. and Kelly L. Haws (2017), "Sounds Big: The Effects of Acoustic Pitch on Product Perceptions," *Journal of Marketing Research*, 54 (2), 331–346.

Lu, Shijie, Dai Yao, Xingyu Chen, and Rajdeep Grewal (2021),"Do Larger Audiences Generate Greater Revenues Under Pay What You Want? Evidence from a Live Streaming Platform," *Marketing Science*, 40 (5), 964–84.

Luangrath, Andrea Webb, Yixiang Xu, and Tong Wang (2023), "Paralanguage classifier (PARA): An algorithm for automatic coding of paralinguistic nonverbal parts of speech in text," *Journal of Marketing Research*, 60 (2), 388–408.

Luchs, Michael G., Rebecca Walker Naylor, Julie R. Irwin, and Rajagopal Raghunathan (2010), "The Sustainability Liability: Potential Negative Effects of Ethicality on Product Preference," *Journal of Marketing*, 74 (5), 18–31.

Khan, Sameed, and Ann Kronrod 2025 (2025), "You Must Calculate That: Developing an Automated Tool to Measure Language Assertiveness," *Advances in Consumer Research*, 53.

Kidwell, Blair, Adam Farmer, and David M. Hardesty (2013), "Getting Liberals and Conservatives to Go Green: Political Ideology and Congruent Appeals," *Journal of Consumer Research*, 40 (2), 350–67.

Kirschner, Paul A. (2002), "Cognitive load theory: Implications of cognitive load theory on the design of learning," *Learning and Instruction*, 12(1), 1–10.

Kress, Gunther and Theo van Leeuwen (2006), *Reading Images. The Grammar of Visual Design*. Routledge.

Krishna, Aradhna (2012) “An Integrative Review of Sensory Marketing: Engaging the Senses to Affect Perception, Judgment and Behavior,” *Journal of Consumer Psychology*, 22(3), 332–351.

Kronrod, Ann, Amir Grinstein, and Luc Wathieu (2012), “Go Green! Should Environmental Messages Be So Assertive?” *Journal of Marketing*, 76 (1), 95–102.

Kronrod, Ann, Anat Tchetchik, Amir Grinstein, Lior Turgeman, and Vered Blass (2023), “Promoting New Pro-Environmental Behaviors: The Effect of Combining Encouraging and Disouraging Messages,” *Journal of Environmental Psychology*, 86, 101945.

Madan, Shilpa, Gita Venkataramani Johar, Jonah Berger, Pierre Chandon, Rajesh Chandy, Rebecca Hamilton, Leslie K. John, Aparna A. Labroo, Peggy J. Liu, John G. Lynch Jr., Nina Mazar, Nicole L. Mead, Vikas Mittal, Christine Moorman, Michael I. Norton, John Roberts, Dilip Soman, Madhu Viswanathan, and Katherine White (2023),“Reaching for Rigor and Relevance: Better Marketing Research for a Better World,” *Marketing Letters*, 34(1), 1–12.

Mairesse, François, Marilyn A. Walker, Matthias R. Mehl, and Roger K. Moore (2007), “Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text,” *Journal of Artificial Intelligence Research*, 30, 457–500.

Mayer, Richard E. (2021), *Multimedia learning* (3rd ed.). Cambridge University Press.

Mayer, Richard E. (2024), “The Past, Present, and Future of the Cognitive Theory of Multimedia Learning,” *Educational Psychology Review*, 36(1)–8.

Mayer, Richard E., Logan Fiorella, and Andrew Stull (2020), “Five Ways to Increase the Effectiveness of Instructional Video,” *Education Tech Research and Development*, 68(3), 837–852.

Mayer, Richard E., and Cheryl I. Johnson (2008), “Revising the Redundancy Principle in Multimedia Learning,” *Journal of Educational Psychology*, 100(2), 380–386.

Mayer, Richard E., and Moreno (2002), “Animation as an Aid to Multimedia Learning,” *Educational Psychology Review*, 14(1), 87–99.

MacInnis, Deborah J., Christine Moorman, and Bernard J. Jaworski (1991), “Enhancing and Measuring Consumers’ Motivation, Opportunity, and Ability to Process Brand Information from Ads,” *Journal of Marketing*, 55(4), 32–53.

McFee, Brian, Colin Raffel, Dawen Liang, Daniel P. Ellis, Matt McVicar, Eric Battenberg, and Oriol Nieto (2015, July), “Librosa: Audio and Music Signal Analysis in Python,” In *Proceedings of the 14th Python in Science Conference*, 8, 18–25.

McShane, Lindsay, Ethan Pancer, Maxwell Poole, and Qi Deng (2021), “Emoji, Playfulness, and Brand Engagement on Twitter,” *Journal of Interactive Marketing*, 53, 96–110.

Meijers, Marijn. H. C., Eline S. Smit, Kelly de Wildt, Sonja-Greetta Karvonen, Demi van der Plas, and L. Nynke van der Laan (2022), “Stimulating Sustainable Food Choices Using Virtual Reality: Taking an Environmental vs. Health Communication Perspective on Enhancing Response Efficacy Beliefs,” *Environmental Communication*, 16(1), 1–22.

Meijers, Marijn. H. C., Jeroen S. Lemmens, Yuyao Lu, Erin R. Newman-Grigg, Xinmiao Lan, Tessa Blanken, and Hanneke Hendriks, (2025), “Let's Talk About Climate Change: How Immersive Media Experiences Stimulate Climate Conversations,” *Journal of Environmental Psychology*, 104, 102610.

Melumad, Shiri, and Robert J. Meyer (2025), “How Listening Versus Reading Alters Consumers’ Interpretations of News,” *Journal of Marketing Research*, 62(2), 342–361.

Meyer, Renate E., Dennis Jancsary, Markus A. Höllerer, and Eva Boxenbaum (2018), “The Role of Verbal and Visual Text in the Process of Institutionalization,” *Academy of Management Review*, *43*(3), 392–418.

Milfeld, Tyler, and Matthew Pittman (2024), “Altercast Framing with Assertive Sustainability Messages: How Dominant Brands Can Motivate Non-Green Consumers,” International Journal of Advertising, 43(1), 173–201.

Mohammad, Saif M. (2025), “NRC VAD Lexicon v2: Norms for Valence, Arousal, and Dominance for over 55k English Terms,” arXiv preprint arXiv:2503.23547.

Montaguti, Elisa, Sara Valentini, and Federica Vecchioni (2023), “Content That Engages Your Customers: The Role of Brand Congruity and Promotions in Social Media,” *Journal of Interactive Marketing*, 58 (1), 16–33.

Mookerjee, Siddhanth, Yann Cornil, and JoAndrea Hoegg (2021), “From Waste to Taste: How ‘Ugly’ Labels Can Increase Purchase of Unattractive Produce,” *Journal of Marketing*, 85 (3), 62–77.

Newman, George E., Margarita Gorlin, and Ravi Dhar (2014), “When Going Green Backfires: How Firm Intentions Shape the Evaluation of Socially Beneficial Product Enhancements,” *Journal of Consumer Research*, 41 (3), 823–39

Nickerson, Dionne, Michael Lowe, Adithya Pattabhiramiah, and Alina Sorescu (2022),“The Impact of Corporate Social Responsibility on Brand Sales: An Accountability Perspective,” *Journal of Marketing*, 86 (2), 5–28.

Nunes, Joseph C., Andrea Ordanini, and Gaia Giambastiani (2021), “The Concept of Authenticity: What It Means to Consumers,” *Journal of Marketing*, 85 (4), 1–20.

NYU CSB and Edelman (2025), “Effective Sustainability Communications – A best Practice Guide for Brands & Marketers”, (last accessed July 26), from https://www.stern.nyu.edu/experience-stern/about/departments-centers-initiatives/centers-of-research/center-sustainable-business/research/research-initiatives/best-practice-guide-effective-sustainability-communications.

Olsen, Mitchell C., Rebecca J. Slotegraaf, and Sandeep R. Chandukala (2014), “Green Claims and Message Frames: How Green New Products Change Brand Attitude,” *Journal of Marketing*, 78 (5), 119–37.

Orazi, Davide C., Bhoomija Ranjan, and Yimin Cheng (2023), “Non-Face Emojis in Digital Marketing: Effects, Contingencies, and Strategic Recommendations,” *Journal of the Academy of Marketing Science*, 51(3), 570–597.

Orazi, Davide C., Anne Hamby, Dennis Herhausen, Tom van Laer, Stephan Ludwig Chahna Gonsalves, and Dhruv Grewal (2025), “Verbal Persuasion in Marketing: A Multimethod Meta-Analysis of Analytical and Narrative Processing,” *Journal of the Academy of Marketing Science*, 1–24.

Orazi, Davide C., and Eugene Y. Chan (2020), “They Did Not Walk the Green Talk!: How Information Specificity Influences Consumer Evaluations of Disconfirmed Environmental Claims,” *Journal of Business Ethics*, 163(1), 107–123.

Packard, Grant and Jonah Berger (2021), “How Concrete Language Shapes Customer Satisfaction,” *Journal of Consumer Research*, 47 (5), 787–806.

Packard, Grant and Jonah Berger (2024), “The Emergence and Evolution of Consumer Language Research,” *Journal of Consumer Research*, 51(1), 42–51.

Paivio, A. (1986). *Mental representations: A dual-coding approach*. Oxford University Press.

Pancer, Ethan, Lindsay McShane, and Theodore J. Noseworthy (2017),"Isolated Environmental Cues and Product Efficacy Penalties: The Color Green and Eco-Labels," *Journal of Business Ethics*, 143 (1), 159–77.

Parguel, Béatrice, Florence Benoît-Moreau, and Cristel Antonia Russell (2015) "Can Evoking Nature in Advertising Mislead Consumers? The Power of 'Executional Greenwashing'," *International Journal of Advertising*, 34(1), 107–134.

Parguel, Beatrice, Florence Benoît-Moreau, and Fabrice Larceneux (2011), "How Sustainability Ratings Might Deter 'Greenwashing:' A Closer Look at Ethical Corporate Communication," *Journal of Business Ethics*, 102 (1), 15–28.

Peloza, John, Katherine White, and Jingzhi Shang (2013), "Good and Guilt-Free: The Role of Self-Accountability in Influencing Preferences for Products with Ethical Attributes," *Journal of Marketing*, 77 (1), 104–19.

Pittman, Matthew, and Annika Abell (2021), "More Trust in Fewer Followers: Diverging Effects of Popularity Metrics and Green Orientation Social Media Influencers," *Journal of Interactive Marketing*, 56(1), 70–82.

Radford, Alec, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, Ilya Sutskever (2023), "Robust Speech Recognition via Large-Scale Weak Supervision," *Proceedings of the 40th International Conference on Machine Learning*, 202, https://proceedings.mlr.press/v202/radford23a.html.

Rajaram Prashant, and Puneet Manchanda (2025), "Unboxing Engagement in YouTube Influencer Videos: An Attention-Based Approach", SSRN (April 30), https://dx.doi.org/10.2139/ssrn.3752107

Royne, Marla B., Jennifer Martinez, Jared Oakley, and Alexa K. Fox (2012), “The Effectiveness of Benefit Type and Price Endings in Green Advertising,” *Journal of Advertising*, 41 (4), 85–102.

Rosenbaum, Paul R. and Donald B. Rubin (1983), “The Central Role of the Propensity Score in Observational Studies for Causal Effects,” *Biometrika*, 70 (1), 41–55.

Rutten, Sanne, Roberta Santoro, Alexis Hervais-Adelman, Elia Formisano, and Narly Golestani (2019),“Cortical Encoding of Speech Enhances Task-Relevant Acoustic Information,” *Nature Human Behaviour*, 3 (9), 974–87.

Rosenberg, Benjamin D., and Jason T. Siegel (2018), “A 50-Year Review of Psychological Reactance Theory: Do Not Read This Article,” *Motivation Science*, 4(4), 281.

Salnikova, Ekaterina, Yuliya Strizhakova, and Robin A. Coulter (2022), “Engaging Consumers with Environmental Sustainability Initiatives: Consumer Global–Local Identity and Global Brand Messaging,” *Journal of Marketing Research*, *59*(5), 983–1001.

Schmuck, Desirée, Jörg Matthes, and Brigitte Naderer (2018), "Misleading Consumers with Green Advertising? An Affect–Reason–Involvement Account of Greenwashing Effects in Environmental Advertising,” *Journal of Advertising*, 47(2), 127–145.

Schraml Christopher (2025), “Automated Video Analytics In Marketing Research: A Systematic Literature Review and a Novel Multimodal Large Language Model Method,” OSF preprint.

Schwenzow, Jasper, Jochen Hartmann, Amos Schikowsky, and Mark Heitmann (2021), “Understanding Videos at Scale: How to Extract Insights for Business Research,” *Journal of Business Research*, 123, 367–79.

Segev Sigal, Juliana Fernandes, and Cheng Hong (2016) “Is Your Product Really Green? A Content Analysis to Reassess Green Advertising,” *Journal of Advertising*, 45(1), 85–93.

Sen Sankar, Chitra Bhanu Bhattacharya , Kristin Lindrud, Silvia Bellezza, Yann Cornil, Shuili Du, Shreyans Goenka, Katharina Husemann, Eric J. Johnson, Cait Lamberton, Gergana Nenkov, Remi Trudel, Katherine White, and Karen Page Winterich (2024), "Enhancing Consumer and Planetary Well-Being by Consuming Less, Consuming Better," *Journal of Sustainable Marketing*," 5(1), 30–42.

Shen, Lijiang, and James Price Dillard (2005), "Psychometric Properties of the Hong Psychological Reactance Scale," *Journal of Personality Assessment*, 85(1), 74–81.

Sprout Social (2025), "How the TikTok Algorithm Works in 2025," (Last Accessed October 13, 2025), from https://sproutsocial.com/insights/tiktok-algorithm/#how.

Sun, Jennifer J., Silvia Bellezza, and Neeru Paharia (2021),"Buy Less, Buy Luxury: Understanding and Overcoming Product Durability Neglect for Sustainable Consumption," *Journal of Marketing*, 85(3), 28–43.

Stuppy, Anika, Jan R. Landwehr, and A. Peter McGraw (2024), "The Art of Slowness: Slow Motion Enhances Consumer Evaluations by Increasing Processing Fluency," *Journal of Marketing Research*, 61(2), 185–203.

Swani, Kunal, and Lauren I. Labrecque (2020), "Like, Comment, or Share? Self-presentation vs. brand relationships as drivers of social media engagement choices," *Marketing Letters*, 31(2), 279-298.

Tam, Katy Yuen Yan, and Michael Inzlicht (2024), "Fast-Forward to Boredom: How Switching Behavior on Digital Media Makes People More Bored," *Journal of Experimental Psychology: General*.

Tavassoli, Nader T. and Yih H. Lee (2003),“The Differential Interaction of Auditory and Visual Advertising Elements with Chinese and English,” *Journal of Marketing Research*, 40 (4), 468–80.

Teixeira, Thales, Michel Wedel, and Rik Pieters (2012), “Emotion-Induced Engagement in Internet Video Advertisements,” *Journal of Marketing Research*, 49(2), 144–159.

Tellis, Gerard J., Deborah J. MacInnis, Seshadri Tirunillai, and Yanwei Zhang (2019), “What Drives Virality (Sharing) of Online Digital Content? The Critical Role of Information, Emotion, and Brand Prominence,” *Journal of Marketing*, 83 (4), 1–20.

The Ground We Walk On (2024) “Four reasons why videos are a powerful tool for companies with an environmental message, ” (accessed 26 October, 2025), from https://thegroundwewalkon.com.au/blogs/news/the-visual-advantage-why-videos-are-a-powerful-tool-for-those-working-to-improve-the-environment.

TikTok for Business (2022), “The Power of Creative Elements in TikTok Ads,” (accessed July 20, 2025), from https://ads.tiktok.com/business/creativecenter/quicktok/online/Power_Creative_Elements/pc/en.

TikTok for Business (2025), “TikTok Continues to Lead Future Media Spend Increase and Innovation According to Kantar,” (Accessed October 13, 2025), from https://ads.tiktok.com/business/en/blog/kantar-media-reactions.

Tian, Zijun, Ryan Dew, and Raghuram Iyengar (2023),“Mega or Micro? Influencer Selection Using Follower Elasticity,” *Journal of Marketing Research*, 61 (3), 472–95.

Valsesia Francesca, Diehl Kristin (2022), “Let Me Show You What I Did Versus What I Have: Sharing Experiential Versus Material Purchases Alters Authenticity and Liking of Social Media Users,” *Journal of Consumer Research*, 49 (3), 430–49.

Van Der Heide, Brandon, Jonathan D. D'Angelo, and Erin M. Schumaker (2012),"The Effects of Verbal Versus Photographic Self-Presentation on Impression Formation in Facebook," *Journal of Communication*, 62 (1), 98–116.

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.

Van Zant, Alexander B. and Jonah Berger (2020), "How the Voice Persuades," *Journal of Personality and Social Psychology*, 118 (4), 661–682.

Villarroel 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.

Yang Jeremy, Juanjuan Zhang, Yuhan Zhang (2024), "Engagement that Sells: Influencer Video Advertising on TikTok," *Marketing Science*, 44 (2), 247–267.

Walker, Lorna, Paul R. Baines, Radu Dimitriu, and Emma K. Macdonald (2017), "Antecedents of Retweeting in a (Political) Marketing Context," *Psychology & Marketing*, 34 (3), 275–293.

Wang, Xinmeng, Zhe Zhang, and Qingyun Jiang (2024), "The Effectiveness of Human vs. AI Voice-Over in Short Video Advertisements: A Cognitive Load Theory Perspective," *Journal of Retailing and Consumer Services*, 81, 104005.

Wang, Xin (Shane), Neil Bendle, and Yinjie Pan (2024), "Beyond Text: Marketing Strategy in a World Turned Upside Down." *Journal of the Academy of Marketing Science* 52(4), 939–954.

Wang, Xin (Shane), Shijie Lu, X. I. Li, Mansur Khamitov, and Neil Bendle (2021), "Audio mining: The Role of Vocal Tone in Persuasion," *Journal of Consumer Research*, 48 (2), 189–211.

Wang, Xia, Ying Ding, and Ying Hu (2025), “The Impact of Face Presence in User-Generated Videos on Consumer Engagement: Insights into How, When, and Who,” *International Journal of Research in Marketing*.

White, Katherine, Rishad Habib, and David J. Hardisty (2019), “How to SHIFT Consumer Behaviors to Be More Sustainable: A Literature Review and Guiding Framework,” *Journal of Marketing*, 83 (3), 22–49.

White, Katherine, Rhiannon MacDonnell, and Darren W. Dahl (2011), “It’s the Mind-Set that Matters: The Role of Construal Level and Message Framing in Influencing Consumer Efficacy and Conservation Behaviors,” *Journal of Marketing Research*, 48 (3), 472–85.

White, Katherine and Bonnie Simpson (2013), “When Do (and Don’t) Normative Appeals Influence Sustainable Consumer Behaviors?” *Journal of Marketing*, 77 (2), 78–95.

White, Katherine, Bonnie Simpson, and Jennifer J. Argo (2014), “The Motivating Role of Dissociative Out-Groups in Encouraging Positive Consumer Behaviors,” *Journal of Marketing Research*, 51 (4), 433–47.

White, Katherine and Chelsea Willness (2009), “Consumer Reactions to Decreased Usage Messages: The Role of Elaborative Processing,” *Journal of Consumer Psychology*,” 19 (1), 73–87.

Winterich, Karen P., Gergana Y. Nenkov, and Gabriel E. Gonzales (2019), “Knowing What It Makes: How Product Transformation Salience Increases Recycling,” *Journal of Marketing*, 83 (4), 21–37.

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.

Wyzowl (2025), “Video Marketing Statistics 2025”, (last accessed July 20), from https://wyzowl.com/video-marketing-statistics/.

Xu, Haifeng, Yu Ding, Yu Ding, Qi Zhang, and Cheng Zhang (2025), “Whispers in Your Mind: the Role of Voice Features in Customer Acquisition and Retention,” *Journal of Marketing*, 1–22.

Zhang, Ke, Siqi Wang, Huan Yang, and Long Chen (2023), “Do Consumers Prefer Sad Faces on Eco-Friendly Products? How Facial Expressions on Green Products in Advertisements Influence Purchase Intentions,” *Journal of Advertising Research*, *63*(3), 274–289.

Zhang, Xiaoke, Mi Zhou, and Gene Moo Lee (2024), “AI Voice in Online Video Platforms: A Multimodal Perspective on Content Creation and Consumption,” (November 20), https://dx.doi.org/10.2139/ssrn.4676705.

Zhu, Rui (Juliet) and Joan Meyers-Levy (2005),“Distinguishing Between the Meanings of Music: When Background Music Affects Product Perceptions,” *Journal of Marketing Research*, 42(3), 333–45.

Zhu, Jingjie, Mingming Cheng, and Ying (Wendy) Wang (2025), “Viewer In-Consumption Engagement in Pro-Environmental Tourism Videos: A Video Analytics Approach,” *Journal of Travel Research*, 64(4), 716–735.

LESS IS MORE: HOW VIDEO DESIGN SHAPES THE EFFECTIVENESS OF GREEN-ORIENTED BRAND MESSAGES ON TIKTOK

WEB APPENDIX

Appendix A: Data Collection and Cleaning

Table WA1. Data Scraping Criteria

		#Post
Raw Data	Include All TikTok posts from European companies in Global Fortune 500 list with a verified TikTok profile, scraped through TikTok Research API. One verified profile per each company, according to the following criteria: if a company had a single global TikTok account, we included that profile. If no global account was available, we selected only the European one. In case of multiple European accounts, we selected the European account with the highest number of followers. For corporate groups [a], we included all affiliated brands with a verified TikTok profile.	6,194
Cleaned Data	Exclude scraped posts not available on the brand's profile at the time of data collection [b] (e.g., due to content removal or profile updates).	5894
	Exclude image-based posts and carousels	5749
Final Sample		5749

Notes: [a] The sample includes the following corporate groups: Lufthansa Group (Lufthansa, Swiss, Eurowings, Austrian Airlines, Brussels Airlines); Inditex Group (Bershka, Massimo Dutti, Stradivarius, Zara, Pull&Bear); Stellantis (Alfa Romeo Official, Chrysler, Citroën, Dodge, Jeep, Maserati, Opel, Peugeot, Ram Trucks, Vauxhall). [b] data were collected in March 2024.

Table WA2. Data Scraping Accuracy by Brand

Brand Profile	Industry	TikTok posts on brand's profile	TikTok image-based posts on brand's profile	Scraped TikTok video-based posts	TikTok video-based posts correctly retrieved
		#	#	#	%
deutschetelekom	Telecommunications	209	1	204	98.08
dhl	Mail, Package, and Freight Delivery	79	2	77	100.00
santander_es	Commercial Banks	64	0	64	100.00
dior	Apparel	207	4	8	3.94
repsol	Petroleum Refining	26	1	25	96.15
airbus	Aerospace & Defense	48	1	47	100.00
deutschebahn	Railroads	171	8	152	93.25
daimlertruck	Motor Vehicles & Parts	31	0	31	100.00
edeka	Wholesalers: Food and Grocery	230	2	227	99.56
sanofi	Pharmaceuticals	72	0	70	97.22
volvotrucks	Motor Vehicles & Parts	265	0	262	98.87
groupebouygues	Engineering & Construction	15	0	15	100,00
snfcvoyageurs	Railroads	50	1	47	95.92
ikeaItalia	Specialty Retailers	72	0	71	98.61
continental	Motor Vehicles & Parts	52	0	49	94.23
lorealparis	Household and Personal Products	371	14	357	100.00
naturgy	Utilities	212	0	212	100.00
metro_de	Wholesalers: Food and Grocery	115	0	106	92.17
volkswagen	Motor Vehicles & Parts	184	3	167	92.27
mercedesbenz	Motor Vehicles & Parts	186	14	83	48.26
allianz_es	Insurance: Life, Health (stock)	262	3	259	100.00
bmw	Motor Vehicles & Parts	149	0	1	0,67
creditagricole	Banks: Commercial and Savings	152	1	145	96.03
e.on_italia	Energy	59	0	53	89.83
carrefourfrance	Food & Drug Stores	325	11	300	95.54
engieitalia	Energy	6	0	6	100,00
maersk_official	Shipping	98	0	97	98.98
renault	Motor Vehicles & Parts	99	4	90	94.74
orange	Telecommunications	150	10	8	5.71
telefonica	Telecommunications	223	5	217	99.54
iberdrola	Utilities	221	0	214	96.83
sap	Computer Software	21	0	7	33.33
zara	Specialty Retailers	113	0	0	0.00
pullandbear	Specialty Retailers	457	12	443	99.55
massimodutti	Specialty Retailers	144	0	144	100.00
bershka	Specialty Retailers	371	8	354	97.52
stradivarius	Specialty Retailers	211	5	203	98.54
zarahome	Specialty Retailers	77	0	77	100.00
alfaromeo_global	Motor Vehicles & Parts	42	0	42	100.00
chrysler	Motor Vehicles & Parts	45	0	45	100.00
citroen_fr	Motor Vehicles & Parts	88	0	88	100.00
dodge	Motor Vehicles & Parts	49	0	17	34.69
jeep	Motor Vehicles & Parts	31	0	14	45.16
maserati	Motor Vehicles & Parts	93	0	92	98.92
opel	Motor Vehicles & Parts	76	0	35	46.05
peugeot	Motor Vehicles & Parts	35	0	13	37.14
ramtrucks	Motor Vehicles & Parts	45	2	41	95.35
vauxhall	Motor Vehicles & Parts	12	0	12	100.00
lufthansa	Airlines	122	2	117	97.50
flyswiss	Airlines	176	10	164	98.80
eurowings	Airlines	139	2	135	98.54
austrianairlines	Airlines	22	0	17	77.27
brusselsairlines	Airlines	26	1	25	100,00

Notes: companies were grouped into industries based on the classification used in the Global Fortune 500 list.

Web Appendix B: Field Study

Table WB1. Correlation Matrix

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)	(25)	(26)	(27)	(28)	(29)	(30)	(31)	(32)	(33)
(1) Like																																	
(2) Comment	.594*																																
(3) Share	.714*	.503*																															
(4) Engagement	.999*	.624*	.742*																														
(5) Green Wordcount Ratio	-.005	-.006	-.002	-.005																													
(6) Greenness	-.032*	-.021	-.011	-.032*	.539*																												
(7) (log)Views	.294*	.132*	.179*	.290*	-.032*	-.039*																											
(8) (log)Video Duration	-.026*	-.032*	-.023	-.027*	.026*	.158*	-.129*																										
(9) Volume	.009	.006	.011	.010	.013	-.020	.049*	-.184*																									
(10) Pitch	-.008	-.006	-.005	-.008	-.019	-.040*	.005	-.088*	.050*																								
(11) Speech Rate	-.043*	.039*	-.024	-.040*	.001	.185*	-.148*	.492*	-.247*	-.113*																							
(12) Colorfulness	-.017	.012	-.015	-.016	-.049*	-.033*	-.004	-.051*	-.007	-.044*	.087*																						
(13) Saturation	.008	.006	.000	.008	.012	.038*	-.064*	.069*	-.022	-.022	.105*	.040*																					
(14) Faces	-.029*	.007	-.041*	-.029*	-.085*	-.052*	-.037*	.085*	-.102*	-.024	.300*	.201*	.019																				
(15) Happy	-.029*	-.001	-.024	-.028*	-.036*	-.012	-.030*	.113*	.013	-.011	.083*	.077*	.070*	.208*																			
(16) Sad	.003	.007	.007	.003	-.053*	-.046*	-.015	.094*	-.033*	-.021	.058*	.078*	-.011	.179*	-.175*																		
(17) Surprise	-.021	-.015	-.016	-.021	.046*	.104*	-.049*	.126*	-.050*	-.037*	.216*	.024	.062*	.169*	-.026	-.055*																	
(18) Fear	.002	-.008	.023	.003	-.016	.017	-.040*	.115*	-.054*	-.032*	.107*	.103*	.028*	.063*	-.112*	.117*	.140*																
(19) Angry	-.004	.003	-.017	-.004	.024	.042*	-.063*	.086*	-.018	-.013	.050*	.022	.013	.055*	-.085*	.048*	.056*	.025															
(20) Disgust	-.018	-.007	-.012	-.018	-.001	.025	-.022	.045*	-.032*	-.026*	.069*	.046*	.021	.055*	.002	.006	.012	.028*	.084*														
(21) Hashtag	-.045*	-.015	-.012	-.043*	.024	.022	-.158*	-.005	.000	.013	.008	-.005	.018	.072*	.052*	-.012	.040*	.001	.008	.042*													
(22) Mentions	.020	-.011	-.003	.018	-.080*	-.080*	.010	.054*	-.028*	.002	.002	.025	-.030*	.122*	.034*	.073*	.025	.004	.002	-.008	-.024												
(23) Emoji	-.059*	.032*	-.052*	-.057*	-.012	.052*	-.087*	.015	-.025	.000	.111*	.075*	.023	.103*	.055*	.016	.052*	.022	.036*	.026	.178*	.058*											
(24) Valence	-.011	-.001	-.025	-.011	.002	-.057*	.067*	-.060*	.062*	.027*	-.151*	-.034*	.037*	-.044*	.014	-.042*	-.028*	-.017	-.025	.002	.024	-.007	.005										
(25) Arousal	.033*	.024	.008	.033*	.032*	-.032*	.021	-.062*	.073*	.004	-.124*	-.039*	.021	-.016	.033*	-.045*	-.021	-.037*	-.005	.009	.035*	.008	-.036*	.102*									
(26) Concreteness	.011	.011	.021	.012	.001	-.030*	.056*	.010	.035*	.008	-.106*	.007	-.060*	-.127*	-.063*	.019	-.024	.020	-.016	-.021	-.005	-.005	.011	.075*	-.164*								
(27) Readability	-.027*	-.009	-.016	-.026*	.018	.168*	-.134*	.684*	-.170*	-.074*	.601*	-.032*	.064*	.173*	.064*	.052*	.140*	.067*	.061*	.046*	.077*	.054*	.025	-.063*	-.067*	-.055*							
(28) Verbal Density	-.049*	.077*	-.021	-.044*	.030*	.206*	-.160*	.177*	-.161*	-.079*	.777*	.072*	.128*	.226*	.069*	.002	.182*	.057*	.033*	.060*	.072*	-.044*	.149*	-.152*	-.101*	-.161*	.510*						
(29) Verbal Similarity	-.048*	.015	-.027*	-.046*	.045*	.173*	-.142*	.381*	-.130*	-.083*	.524*	-.020	.103*	.097*	.090*	.009	.143*	.062*	.064*	.057*	.062*	-.064*	.108*	-.089*	-.048*	-.102*	.426*	.587*					
(30) Visual-Text Similarity	.015	.038*	.035*	.017	.044*	-.019	.037*	-.159*	.028*	.039*	-.163*	.001	-.048*	-.164*	-.027*	-.102*	-.076*	-.067*	-.035*	.005	.091*	-.033*	-.007	-.007	.009	.063*	-.119*	-.032*	.059*				
(31) Informative Goal	-.017	-.049*	-.012	-.019	-.026*	.054*	-.052*	.168*	-.053*	-.024	.150*	-.024	.036*	.036*	.009	.009	.028*	.032*	-.004	.020	.017	.019	-.019	-.009	.033*	-.051*	.150*	.132*	.125*	-.013			
(32) Brand Congruity	-.017	.019	-.001	-.015	.101*	.159*	-.127*	.155*	-.015	-.049*	.349*	.029	.054*	.145*	.047*	.047*	.122*	.040*	.039*	.019	-.085*	-.030*	.148*	-.108*	-.091*	-.038*	.186*	.282*	.213*	-.019	.002		
(33) Assertiveness	-.031*	.007	-.020	-.030*	-.035*	.052*	-.014	.201*	-.061*	.005	.193*	-.019	.016	.000	.007	.032*	.045*	.036*	.017	.000	-.017	.000	.038*	-.076*	-.082*	-.028*	.202*	.188*	.157*	-.040*	.075*	.048*	

Notes: * $p < 0.05$.

Table WB2. Results before PSM

DV: Engagement		**Model 1**			**Model 2**	
	IRR	**Coeff.**	***p***	**IRR**	**Coeff.**	***p***
Independent Variables:						
Video Design:						
Image+Caption+TO	1.092	.088 (.047)	.058	1.113	.107 (.048)	.026
Image+Caption+VO	1.525	.422 (.093)	.000	1.735	.551 (.096)	.000
Image+Caption+TO+VO	1.127	.119 (.071)	.093	1.140	.131 (.071)	.065
Green Claim:						
Yes	.740	-.301 (.053)	.000	1.060	.058 (.145)	.689
Interactions: Video Design * Green Claim						
Image+Caption+TO * Green Claim				.666	-.406 (.162)	.012
Image+Caption+VO * Green Claim				.156	-1.858 (.247)	.000
Image+Caption+TO+VO * Green Claim				.720	-.328 (.157)	.036
Additional Controls:						
Log (Views)	1.925	.655 (.007)	.000	1.926	.655 (.007)	.000
Video Controls:						
Log (Video Duration)	.317	-1.149 (.202)	.000	.313	-1.163 (.202)	.000
Log (Video Duration)^2	1.189	.173 (.034)	.000	1.194	.178 (.034)	.000
Audial Controls						
Volume	1.009	.009 (.003)	.012	1.009	.009 (.003)	.011
Pitch	1.000	-.000 (.000)	.024	1.000	-.000 (.000)	.018
Speech Rate	1.028	.028 (.028)	.318	1.028	.028 (.028)	.328
Visual Controls						
Colorfulness	1.603	.472 (.144)	.001	1.609	.476 (.143)	.001
Saturation	.825	-.192 (.126)	.128	.827	-.190 (.126)	.132
Faces	.994	-.006 (.051)	.909	.994	-.006 (.050)	.907
Happy	1.043	.042 (.066)	.525	1.036	.035 (.066)	.596
Sad	.840	-.174 (.083)	.035	.853	-.158 (.082)	.054
Surprise	.529	-.637 (.223)	.004	.521	-.652 (.223)	.003
Fear	.880	-.127 (.087)	.143	.867	-.143 (.087)	.099
Angry	1.033	.033 (.124)	.792	1.036	.036 (.123)	.773
Disgust	.657	-.420 (.514)	.414	.633	-.457 (.512)	.372
Verbal Controls						
Caption Hashtags	.990	-.010 (.008)	.219	.989	-.011 (.008)	.185
Caption Mentions	1.150	.140 (.030)	.000	1.146	.136 (.030)	.000
Caption Emoji	1.013	.013 (.012)	.267	1.013	.013 (.012)	.255
Valence	1.024	.024 (.115)	.835	1.021	.021 (.116)	.858
Arousal	1.083	.080 (.142)	.573	1.081	.078 (.142)	.583
Concreteness	1.000	.000 (.000)	.714	1.000	.000 (.000)	.651
Readability	.996	-.004 (.002)	.042	.995	-.005 (.002)	.023
Verbal Density	1.013	.012 (.014)	.379	1.013	.013 (.014)	.363
Fixed Effects						
Time FE		Included			Included	
Brand FE		Included			Included	
Intercept	39.394	3.674 (.477)	.000	39.046	3.665 (.476)	.000
N		5,620			5,620	
AIC		88,714.224			88,673.818	
Log likelihood		-44,264.112			-44,240.909	

Notes: IRR = incident rate ratio. An IRR greater than (less than) 1 indicates a positive (negative) effect. Standard errors are in parentheses. TO = text overlay, VO = voice over.

Alternative Measures of Message Green Orientation

We considered two measurements alternative to our dummy variable qualifying a post as green-oriented, namely a green word count ratio and a continuous measure of the post's green orientation.

First, we used Wordify (Hovy, Melumad, and Inman 2021) to identify an initial list of words that might be related to green claims, with the aim of constructing a dictionary of representative or indicative words that may occur in a green-oriented post. Wordify identifies the most representative words within a given a priori-coded category – in our case, a post classified as green-oriented. Therefore, we input into the tool the full text of caption, text overlay, and voice-over for each post (when available). Given our need of using the tool for vocabulary discovery, we set a 15% inclusion threshold to retain a broader lexical base for subsequent annotation. Wordify returned an initial list of 283 words. We then refined the list by selecting only those words that were judged as contextually relevant based on human and GPT-based annotation. Specifically, the first author manually annotated whether to include each word in the dictionary. In parallel, following Blanchard's et al. (2025) best practices for an effective GenAI use in annotation tasks, we instructed OpenAI's GPT 4.1 to act as a second coder based on the following prompt:

"*Your task is to act as an independent coder for a dictionary-building task aimed at identifying the most representative or indicative words that might occur in a green claim. A green claim is defined as "any ad that meets one or more of the following criteria: (1) explicitly or implicitly addresses the relationship between a product/service and the biophysical environment, (2) promotes a green lifestyle with or without highlighting a product/service, (3) presents a corporate image of environmental responsibility" (Banerjee, Gulas, and Iyer 1995, p. 22). Please analyze each element in the column "Word". This column may contain single words (e.g., "climate"), compound words (e.g., "eco-friendly"), merged* *expressions (e.g., "electriccar", "greenvibe"), or multi-word expressions (e.g., "climate change", "anti waste"). Code identically any word deriving from the same root (e.g., "recycle," "recycled", "recycling", "recyclable"). Please give me a response in column "GPT-4.1" as a string in the format: [1] if the word is typically representative or indicative of a green claim, and in the format [0] if the word is not typically representative or indicative of a green claim. Do not provide any additional explanation or text.*"

Inter-coder reliability was high (Krippendorff's alpha = .76), and any disagreement was resolved by a third human coder (i.e., a coauthor). We ended up with the following 52 n-grams that are likely to be related with green claims: "*electric*", "*energy*", "*sustainability*", "*renewable*", "*save*", "*waste*", "*sustainable*", "*green*", "*makeitalygreen*", "*hybrid*", "*planet*", "*environment*", "*allelectric*", "*vegan*", "*recycle*", "*etech*", "*emission*", "*organic*", "*tree*", "*plant*", "*reduce*", "*vegetarian*", "*veggie*", "*solar panel*", "*earth*", "*eco*", "*wind*", "*solar*", "*responsible*", "*transition*", "*recycled material*", "*wind farm*", "*biodiversity*", "*energyefficiency*", "*recycled*", "*greenvibe*", "*electriccar*", "*recycling*", "*electricity*", "*footprint*", "*environmental*", "*carbon*", "*fuelefficiency*", "*anti waste*", "*sustainable future*", "*climate change*", "*hydrogen*", "*climate*", "*electromobility*", "*electrictrucksinreality*", "*electrictruck*", "*sea*".

Using this dictionary, we computed the green word count ratio, measured as the sum of sustainability-related word count across text overlay, voice-over, and caption, divided by the total word count.

Second, we considered a continuous measure of greenness, based on a 1–7-point Likert scale annotation by the first author, which captures the extent of environmental focus in the post (Banerjee, Gulas, and Iyer 1995). This measure assesses both the centrality of green-oriented information within the overall message as well as the level of detail conveyed across modes. As such, posts featuring only vague environmental references (e.g., product-oriented posts including just a word like “sustainability”, “renewables”, “electric” in the caption or text overlay) were coded as having a shallow greenness, typically ranging from 1 to 3 depending on how consistently and prominently such green-oriented cues were expressed across modes. For instance, a post including generic terms like “electric” or “sustainable” only in the caption would fall at the lower end of the scale, whereas the same vague message reiterated through multiple modes (e.g., nature-evoking visuals, text overlay and voice-over) would be assigned a slightly higher score within the shallow greenness range. The same logic was applied to assess moderately green and deep-green posts. Posts not addressing environmental issues in detail, but mentioning specific issues (e.g., anti-waste recipes and behaviors, or information on electric vehicle energy consumption), were coded as having a moderate level of greenness. Posts that focused solely on environmental issues and discussed them in depth (e.g., a detailed description of company’s commitment for the environment or about how to behave sustainably) were coded as having a deep level of greenness.

While our dummy-coded measure qualifying a post as green was used for matching purposes, we employed these two alternative measures (i.e., green word count ratio, continuous greenness score) as focal predictors in the outcome model to assess the robustness of our findings to different operationalizations of green message orientation. Results are reported in Table WB3 (see respectively Model 7 and Model 8 for details).

Table WB3. Robustness Tests for Field Study

DV: Engagement	**Model 1: PSM Caliper = .1**			**Model 2: PSM Caliper = .01**		
	IRR	**Coeff.**	***p***	**IRR**	**Coeff.**	***p***
Independent Variables:						
Video Design						
Image+Caption+TO	.966	-.034 (.121)	.777	1.006	.006 (.125)	.964
Image+Caption+VO	.933	-.070 (.229)	.761	1.071	.068 (.237)	.773
Image+Caption+TO+VO	.806	-.216 (.152)	.155	.854	-.157 (.159)	.321
Green Claim:						
Yes	1.155	.145 (.154)	.349	1.266	.236 (.164)	.151
Interactions: Video Design * Green Claim						
Image+Caption+TO * Green Claim	.636	-.453 (.178)	.011	.572	-.558 (.189)	.003
Image+Caption+VO * Green Claim	.320	-1.139 (.310)	.000	.263	-133.577 (.326)	.000
Image+Caption+TO+VO * Green Claim	.621	-.476 (.172)	.006	.589	-.529 (.182)	.004
Additional Controls:						
Log (Views)	1.827	.603 (.013)	.000	1.854	.617 (.014)	.000
Video Controls:						
Log (Video Duration)	.872	-.137 (.352)	.697	1.127	.119 (.361)	.741
Log (Video Duration)^2	1.052	.051 (.057)	.369	1.017	.017 (.058)	.777
Audial Controls						
Volume	1.016	.016 (.007)	.018	1.017	.017 (.007)	.016
Pitch	1.000	-.000 (.000)	.076	1.000	-.000 (.000)	.111
Speech Rate	1.228	.205 (.050)	.000	1.175	.161 (.053)	.002
Visual Controls						
Colorfulness	1.015	.015 (.308)	.961	.819	-.199 (.316)	.528
Saturation	2.347	.853 (.242)	.000	2.440	.892 (.248)	.000
Faces	.891	-.116 (.109)	.288	.950	-.051 (.113)	.652
Happy	1.247	.221 (.115)	.055	1.400	.336 (.123)	.006
Sad	.860	-.150 (.153)	.326	.816	-.203 (.156)	.194
Surprise	.941	-.061 (.407)	.882	.821	-.197 (.425)	.643
Fear	1.258	.229 (.151)	.128	1.293	.257 (.157)	.101
Angry	1.317	.276 (.221)	.211	1.168	.155 (.223)	.487
Disgust	.459	-.779 (.620)	.209	.470	-.755 (.638)	.237
Verbal Controls						
Caption Hashtags	1.013	.013 (.014)	.365	1.007	.007 (.015)	.656
Caption Mentions	.854	-.158 (.079)	.047	.809	-.212 (.082)	.010
Caption Emoji	.958	-.043 (.024)	.075	.967	-.034 (.025)	.176
Valence	.583	-.540 (.226)	.017	.551	-.596 (.239)	.013
Arousal	1.372	.316 (.288)	.272	1.365	.311 (.313)	.320
Concreteness	.998	-.002 (.001)	.036	.999	-.001 (.001)	.237
Readability	.990	-.010 (.004)	.019	.990	-.010 (.004)	.018
Verbal Density	1.020	.020 (.026)	.443	1.045	.044 (.028)	.113
Verbal Similarity Index	.542	-.612 (.204)	.003	.481	-.733 (.213)	.001
Visual-Text Similarity	135.577	4.910 (.991)	.000	126.671	4.842 (1.032)	.000
Assertiveness	.973	-.027 (.017)	.109	.973	-.027 (.018)	.130
Informative (cf. Commercial) Goal	.759	-.275 (.088)	.002	.726	-.321 (.093)	.001
Brand Congruity	.000	-26.039 (13.704)	.057	.000	-26.997 (13.975)	.053
Fixed Effects						
Time FE		Included			Included	
Brand FE		Included			Included	
Intercept	150'000'000	18.825 (8.446)	.026	1128.349	7.029 (9.029)	.436
N		1,224			1,163	
AIC		18,597.168			19	
Log likelihood		-9,210.584			-8,771.512	

Notes: IRR = incidence rate ratio. An IRR greater than (less than) 1 indicates a positive (negative) effect. Standard errors are in parentheses. TO = text overlay, VO = voice over

	Model 3: DV: Log-transformed Engagement		Model 4: DV: Like		
	Coeff.	*p*	IRR	Coeff.	*p*
Independent Variables:					
Video Design					
Image+Caption+TO	.096 (.121)	.428	.970	-.031 (.121)	.799
Image+Caption+VO	-.089 (.210)	.671	.967	-.034 (.230)	.884
Image+Caption+TO+VO	-.120 (.146)	.412	.831	-.185 (.152)	.223
Green Claim:					
Yes	.129 (.154)	.403	1.115	.109 (.155)	.480
Interactions: Video Design * Green Claim					
Image+Caption+TO * Green Claim	-.474 (.180)	.008	.673	-.397 (.178)	.026
Image+Caption+VO * Green Claim	-.873 (.297)	.003	.322	-1.134 (.311)	.000
Image+Caption+TO+VO * Green Claim	-.342 (.170)	.045	.652	-.427 (.172)	.013
Additional Controls:					
Log (Views)	.568 (.014)	.000	1.837	.608 (.013)	.000
Video Controls:					
Log (Video Duration)	.330 (.327)	.313	.919	-.084 (.350)	.810
Log (Video Duration)^2	-.047 (.053)	.371	1.038	.038 (.056)	.504
Audial Controls					
Volume	.002 (.007)	.744	1.017	.017 (.007)	.015
Pitch	-.000 (.000)	.242	1.000	-.000 (.000)	.096
Speech Rate	.164 (.048)	.001	1.234	.210 (.050)	.000
Visual Controls					
Colorfulness	.217 (.289)	.454	1.070	.068 (.308)	.827
Saturation	.528 (.240)	.028	2.567	.943 (.242)	.000
Faces	-.104 (.106)	.326	.905	-.099 (.109)	.360
Happy	.040 (.111)	.721	1.239	.215 (.115)	.063
Sad	-.154 (.156)	.324	.869	-.140 (.153)	.361
Surprise	.069 (.407)	.865	.728	-.318 (.414)	.443
Fear	.138 (.150)	.357	1.296	.259 (.151)	.086
Angry	.107 (.201)	.595	1.279	.246 (.218)	.259
Disgust	-.562 (.606)	.353	.464	-.768 (.625)	.219
Verbal Controls					
Caption Hashtags	.013 (.013)	.343	1.007	.007 (.014)	.618
Caption Mentions	-.103 (.076)	.175	.852	-.160 (.080)	.046
Caption Emoji	-.033 (.024)	.161	.958	-.043 (.024)	.073
Valence	-.217 (.224)	.333	.560	-.579 (.224)	.010
Arousal	-.047 (.278)	.865	1.379	.321 (.289)	.267
Concreteness	-.001 (.001)	.202	.998	-.002 (.001)	.051
Readability	-.004 (.004)	.348	.992	-.008 (.004)	.039
Verbal Density	-.022 (.024)	.358	1.009	.009 (.026)	.731
Verbal Similarity Index	-.451 (.197)	.022	.541	-.614 (.202)	.002
Visual-Text Similarity	2.173 (.947)	.022	143.967	4.970 (.992)	.000
Assertiveness	-.004 (.017)	.815	.979	-.021 (.017)	.223
Informative (cf. Commercial) Goal	-.118 (.086)	.168	.771	-.260 (.088)	.003
Brand Congruity	-14.109 (13.907)	.311	.000	-27.509 (13.697)	.045
Fixed Effects					
Time FE	Included			Included	
Brand FE	Included			Included	
Intercept	10,478 (8.595)	.223	29’000’000	19.484 (8.443)	.021
N	1,225			1,225	
AIC	3,120.647			18,427.537	
Log likelihood				-9,125.769	

Notes: IRR = incidence rate ratio. An IRR greater than (less than) 1 indicates a positive (negative) effect. Standard errors are in parentheses. TO = text overlay, VO = voice over

	Model 5: DV: Comment			Model 6: DV: Share		
	IRR	Coeff.	*p*	IRR	Coeff.	*p*
Independent Variables:						
Video Design						
Image+Caption+TO	.852	-.160 (.157)	.307	.920	-.083 (.169)	.623
Image+Caption+VO	.715	-.336 (.297)	.258	.619	-.480 (.336)	.153
Image+Caption+TO+VO	.538	-.621 (.196)	.002	.609	-.496 (.216)	.022
Green Claim:						
Yes	1.863	.622 (.193)	.001	1.067	.065 (.202)	.748
Interactions: Video Design * Green Claim						
Image+Caption+TO * Green Claim	.504	-.684 (.227)	.003	.809	-.211 (.236)	.370
Image+Caption+VO * Green Claim	.318	-1.146 (.397)	.004	.407	-.899 (.446)	.044
Image+Caption+TO+VO * Green Claim	.434	-.834 (.218)	.000	.650	-.430 (.233)	.065
Additional Controls:						
Log (Views)	1.622	.483 (.016)	.000	2.562	.941 (.023)	.000
Video Controls:						
Log (Video Duration)	.541	-.614 (.445)	.168	1.000	-.000 (.460)	1.000
Log (Video Duration)^2	1.182	.168 (.072)	.019	1.039	.038 (.074)	.609
Audial Controls						
Volume	1.016	.016 (.009)	.067	1.020	.020 (.010)	.035
Pitch	1.000	-.000 (.000)	.538	1.000	.000 (.000)	.353
Speech Rate	1.274	.242 (.065)	.000	1.252	.225 (.074)	.002
Visual Controls						
Colorfulness	.330	-1.109 (.404)	.006	.479	-.737 (.439)	.093
Saturation	2.315	.839 (.317)	.008	2.663	.979 (.329)	.003
Faces	.940	-.062 (.148)	.677	.595	-.520 (.168)	.002
Happy	.942	-.060 (.143)	.675	1.086	.082 (.160)	.607
Sad	1.017	.017 (.198)	.934	.763	-.271 (.213)	.204
Surprise	1.892	.638 (.538)	.236	1.050	.048 (.602)	.936
Fear	1.096	.091 (.195)	.639	.921	-.082 (.213)	.701
Angry	1.396	.334 (.281)	.235	1.126	.119 (.295)	.686
Disgust	.302	-1.197 (.780)	.125	.285	-1.254 (.762)	.100
Verbal Controls						
Caption Hashtags	1.028	.028 (.018)	.135	1.039	.038 (.021)	.071
Caption Mentions	.668	-.403 (.106)	.000	.749	-.290 (.120)	.016
Caption Emoji	.973	-.028 (.031)	.373	.917	-.086 (.033)	.009
Valence	.808	-.213 (.300)	.478	1.619	.482 (.299)	.107
Arousal	2.044	.715 (.372)	.055	.446	-.808 (.366)	.027
Concreteness	.996	-.004 (.001)	.003	.998	-.002 (.001)	.061
Readability	.977	-.023 (.005)	.000	.988	-.012 (.006)	.041
Verbal Density	1.076	.073 (.033)	.027	.994	-.006 (.037)	.870
Verbal Similarity Index	.599	-.513 (.267)	.055	1.476	.389 (.294)	.185
Visual-Text Similarity	322.418	5.776 (1.282)	.000	204.088	5.319 (1.416)	.000
Assertiveness	.942	-.060 (.022)	.007	.974	-.027 (.024)	.256
Informative (cf. Commercial) Goal	.698	-.359 (.116)	.002	1.182	.167 (.130)	.199
Brand Congruity				.164	-1.806 (17.019)	.916
Fixed Effects						
Time FE		Included			Included	
Brand FE		Included			Included	
Intercept	9.363	2.237 (.1.279)	.080	.016	-4.144 (10.507)	.693
N		1,225			1,256	
AIC		10,204.585			7,976.969	
Log likelihood		-5,014.293			-3,899.486	

Notes: IRR = incidence rate ratio. An IRR greater than (less than) 1 indicates a positive (negative) effect. Standard errors are in parentheses. TO = text overlay, VO = voice over

	Model 7: Green Wordcount Ratio as Alternative IV			Model 8: DV: Greenness as alternative IV		
	IRR	Coeff.	*p*	IRR	Coeff.	*p*
Independent Variables:						
Video Design						
Image+Caption+TO	.916	-.087 (.111)	.430	.910	-.095 (.116)	.413
Image+Caption+VO	.835	-.181 (.210)	.390	.856	-.156 (.218)	.476
Image+Caption+TO+VO	.744	-.295 (.150)	.048	.729	-.317 (.151)	.036
Green Claim:						
Yes	1.678	.518 (1.674)	.757	1.053	.051 (.062)	.404
Interactions: Video Design * Green Claim						
Image+Caption+TO * Green Claim	.003	-5.984 (1.994)	.003	.881	-.127 (.066)	.056
Image+Caption+VO * Green Claim	.000	-18.611 (5.999)	.002	.762	-.272 (.082)	.001
Image+Caption+TO+VO * Green Claim	.016	-4.141 (2.162)	.056	.907	-.097 (.063)	.123
Additional Controls:						
Log (Views)	1.833	.606 (.013)	.000	1.832	.606 (.013)	.000
Video Controls:						
Log (Video Duration)	.836	-.180 (.348)	.606	.899	-.107 (.351)	.761
Log (Video Duration)^2	1.053	.052 (.056)	.355	1.048	.047 (.057)	.411
Audial Controls						
Volume	1.014	.014 (.007)	.043	1.015	.015 (.007)	.030
Pitch	1.000	-.000 (.000)	.097	1.000	-.000 (.000)	.085
Speech Rate	1.198	.180 (.049)	.000	1.223	.201 (.051)	.000
Visual Controls						
Colorfulness	.949	-.052 (.311)	.868	1.107	.102 (.309)	.742
Saturation	2.543	.933 (.243)	.000	2.528	.927 (.242)	.000
Faces	.939	-.063 (.108)	.563	.890	-.117 (.109)	.284
Happy	1.202	.184 (.115)	.109	1.257	.229 (.116)	.048
Sad	.878	-.131 (.152)	.390	.871	-.138 (.152)	.364
Surprise	.851	-.161 (.417)	.700	.911	-.093 (.415)	.822
Fear	1.337	.290 (.150)	.053	1.280	.247 (.151)	.102
Angry	1.329	.285 (.217)	.190	1.346	.297 (.220)	.177
Disgust	.409	-.893 (.630)	.157	.422	-.863 (.632)	.172
Verbal Controls						
Caption Hashtags	1.010	.010 (.014)	.460	1.007	.007 (.014)	.604
Caption Mentions	.853	-.159 (.080)	.048	.855	-.157 (.080)	.050
Caption Emoji	.946	-.056 (.024)	.020	.955	-.046 (.024)	.058
Valence	.658	-.419 (.227)	.064	.562	-.577 (.226)	.011
Arousal	1.635	.492 (.291)	.091	1.436	.362 (.289)	.211
Concreteness	.999	-.001 (.001)	.127	.998	-.002 (.001)	.030
Readability	.991	-.009 (.004)	.037	.991	-.009 (.004)	.028
Verbal Density	1.004	.004 (.025)	.885	1.020	.020 (.026)	.452
Verbal Similarity Index	.553	-.592 (.200)	.003	.556	-.587 (.203)	.004
Visual-Text Similarity	145.954	4.983 (.101)	.000	135.047	4.906 (.996)	.000
Assertiveness	.975	-.025 (.017)	.143	.976	-.024 (.017)	.158
Informative (cf. Commercial) Goal	.733	-.311 (.090)	.001	.755	-.281 (.088)	.001
Brand Congruity	.000	-27.053 (13.739)	.049	.000	-29.600 (13.726)	.031
Fixed Effects						
Time FE		Included			Included	
Brand FE		Included			Included	
Intercept	221'000'000	19.212 (8.471)	.023	1'190'000'000	20.894 (8.462)	.014
N		1,225			1225	
AIC		18,613.609			18,614.536	
Log likelihood		-9,218.805			-9,219.268	

Notes: IRR = incidence rate ratio. An IRR greater than (less than) 1 indicates a positive (negative) effect. Standard errors are in parentheses. TO = text overlay, VO = voice over

References (not in the main text)

Blanchard, Simon J., Nofar Duani, Aaron M. Garvey, Oded Netzer, Travis Tae Oh (2025), "New Tools, New Rules: A Practical Guide to Effective and Responsible GenAI Use for Surveys and Experiments Research," *Journal of Marketing*, 89(6), 1-21.