

UNDERSTANDING THE EFFECTIVENESS OF MULTIMODAL MESSAGES ON SOCIAL MEDIA ENGAGEMENT: THE ROLE OF NARRATING VOICE AND TEXT OVERLAYS IN GREEN CLAIM VIDEOS

ABSTRACT

Using videos to convey brand messages to consumers has become predominant in marketing communication. This raises questions about *what* video elements (e.g., narrating voice, text overlays) should be combined to engage consumers with multimodal messages. In a preliminary analysis of 5,749 TikTok videos by top European companies, we find that for videos including green (vs. non-green) claims using multiple modes can backfire. Specifically, using a narrating voice, text overlay, or a combination of both on top of the baseline elements (i.e., moving images and caption) is detrimental to engagement of videos that promote a green lifestyle or consumption.

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

INTRODUCTION

Multimodal messages are ubiquitous. On social media platforms such as TikTok 25 million videos are shared daily (Ch 2024). These videos often include moving images, audio, text overlays, and narrating voices, simultaneously.

Nevertheless, anecdotal evidence suggests that adding as many elements as possible to videos is becoming a common managerial practice across industries. Considering the urge and societal relevance of improving the effectiveness of sustainable communication (European Commission 2023; Federal Trade Commission 2022), the combination of different modes could be deemed as a viable strategy to clearly communicate the green attributes of a product, avoid greenwashing perceptions and foster social change. Sustainable communication refers to “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, and (3) presents a corporate image of environmental responsibility” (Banerjee, Gulas, and Iyer 1995, p. 22).

Past research falls short in investigating how multiple modes (i.e., text, image, audio) *jointly* affect consumer behavior (Grewal et al. 2022; Packard and Berger 2024), more so when the context of sustainability is considered. Research in said context, indeed, focuses mostly on text-based communication (e.g., Kronrod et al. 2023; Olsen, Slotegraaf, and Chandukala 2014; White, MacDonnell, and Dahl 2011) neglecting to focus on multiple modes. Similarly, marketing research on videos has explored the impact of *individual* modes on consumer responses, focusing only on text (e.g., Cascio Rizzo et al., 2023) or on audial features (e.g., Chang, Mukherjee and Chattopadhyay 2023) or visual features (e.g., Yang, Zhang, and Zhang, 2023). Little research investigates the combined use of different modes and the effect on consumer’s attitude

and behavior (Holiday et al. 2023; Zhou et al. 2021). In this paper we aim at filling this gap by investigating how the combination of visual, textual, and audio stimuli affects consumer engagement of social media videos.

METHODOLOGY

We selected the European companies from the 2023 Global Fortune 500 list with a verified TikTok account. We collected TikTok posts using the official TikTok Research API. The data scraping generated 5,749 videos by 52 companies across 21 industries from 1st January 2023 to 31st December 2023.

Measures and model estimation. We asked a research assistant to annotate the degree of multimodality in videos and whether the social media post showed a green advertisement (Banerjee, Gulas, and Iyer 1995). Specifically, 15.83% of the videos ($n = 910$) included a green claim. Videos featuring the baseline elements (i.e., moving images and caption) represent 24.44% of the sample ($n = 1,405$), 29.08% include a text overlay ($n = 1,672$), and 4.84% ($n = 278$) a narrating voice. Finally, videos presenting both a text overlay and narrating voice sum up to 41.64% ($n = 2,394$) of the total sample.

We treat engagement measures such as like, comments and shares separately given that they require different effort, are driven by different reasons and have different implications (Swani and Labrecque 2020). Given that our dependent measures were count variables and overdispersed, we run a negative binomial regression. However, the decision of a company to post a content including a green claim is not randomly assigned across the posts, rather can be a strategic and deliberate decision of the company, which may correlate with other factors that also influence engagement.

Following prior research (e.g., Li and Xie 2020), we use propensity score matching to alleviate endogeneity concerns about self-selection bias and adjust for the differences in the treatment and control group. In our study, the propensity score is the predicted probability to post a green claim conditional on the value of the covariates. Both observable and unobservable factors may lead to endogeneity and bias our inferences. Therefore, we included brand fixed effects to account for unobservable factors, together with posting time (e.g., weekday) and post features in our matching model. Specifically, we included textual features of the caption (i.e., arousal, valence, readability, concreteness, wordcount, number of hashtags, emoji and mentions), visual features¹ (i.e., colorfulness, saturation, presence of human faces across frames) and audial features (i.e., volume). We adopted a 1:1 nearest-neighbor matching algorithm without replacement and a caliper of .01 to match a green post with a non-green post. The resulting matched sample included 936 videos, half green and half non-green. After matching, the control and treatment groups were not significantly different from each other, despite some variables were still not perfectly balanced between treatment and control group ($B = 28.9\%$, $R = .96$).

Results. When videos include a green claim, using more modes decreases engagement. Specifically, using a text overlay ($b = -.871$, $SE = .204$, $z = -4.27$, $p = .000$), narrating voice ($b = -1.494$, $SE = .346$, $z = -4.31$, $p = .000$), or a combination of both (b

¹ To extract visual features, we used an open-source video mining tool from Schwenzow et al. (2021).

= -.585, SE =.190, z = -3.08, p = .002) on top of the baseline elements (i.e., moving images and caption) negatively affects the engagement of videos including green claims. Results remain consistent for any measure of engagement.

Table: results

	DV: Like	DV: Comment	DV: Share	DV: Overall Engagement
IV: MD				
Image+Caption+TO	.092 (.129)	.013 (.164)	.289* (.164)	.117 (.129)
Image+Caption+NV	.191 (.232)	.245 (.294)	.216 (.354)	.198 (.392)
Image+Caption+TO +NV	.064 (.132)	.120 (.170)	-.077 (.173)	.071 (.132)
GreenClaim:				
Yes	.371** (.165)	.945*** (.210)	.539** (.214)	.414** (.166)
Interaction: MD * GreenClaim				
Image+Caption+TO GreenClaim	* -.815*** (.203)	-1.184*** (.260)	-.661** (.259)	-.871*** (.204)
Image+Caption+NV GreenClaim	* -1.46*** (.347)	-1.389*** (.438)	-1.467*** (.475)	-1.494*** (.346)
Image+Caption+TO +NV GreenClaim	* -.523*** (.189)	-1.006*** (.242)	-.544** (.247)	-.585*** (.190)
Controls:				
Log(Views)	.590*** (.017)	.495*** (.020)	-1.000*** (.028)	.587*** (.0165)
Video Duration	-.001 (.001)	-.004* (.002)	.007*** (.002)	-.000 (.001)
Audio				
Volume	.005 (.006)	-.003 (.008)	-.006 (.008)	.004 (.006)
Visual				
Colorfulness	-.263 (.372)	-.970** (.480)	-.636 (.503)	-.245 (.371)
Saturation	.436 (.288)	.506 (.369)	.498 (.385)	.432 (.287)
Faces_ratio	.167 (.128)	.490*** (.169)	-.360** (.176)	.194 (.128)
Caption				
Valence	-.387 (.280)	-.407 (.357)	.422 (.380)	-.394 (.281)
Arousal	.280 (.316)	1.010** (.408)	-.346 (.423)	.334 (.315)
Concreteness	.000*** (.000)	.000** (.000)	.000* (.000)	.000*** (.000)
Readability	.010 (.008)	.032*** (.010)	-.004 (.010)	.012 (.008)
Wordcount	-.004* (.002)	-.003 (.003)	-.009 (.004)	-.004*** (.002)
# Hashtags	-.043** (.020)	-.029 (.026)	.043 (.026)	-.040** (.020)
# Mentions	.009 (.097)	-.272** (.122)	-.402*** (.135)	-.002 (.098)
# Emoji	-.030 (.029)	.086** (.038)	.014 (.037)	-.025 (.029)
Time FE	Included	Included	Included	Included
Brand FE	Included	Included	Included	Included
Intercept	2.339 (.797)	-.136 (.968)	-7.870*** (1.070)	2.32*** (.792)
N	936	936	936	936
Log likelihood	-7,323.666	-4,103.036	-3,046.085	-7389.395

Note: .- * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

The effect is more negative when only a narrating voice or a text overlay is used. A possible explanation might be that using a text overlay, and a narrating voice simultaneously could make the video easier to process particularly when the information replicates across modes (Ceylan, Diehl and Proserpio 2024). Thus, our findings suggest that companies on TikTok should carefully implement text overlays or narrating voices when communicating sustainability through videos. Consumers generally expect that companies aim to persuade them when promoting their products

and activities (Friestad and Wright 1994). It is reasonable to think that consumers tend to become more skeptical when companies make higher efforts when promoting sustainability of their products and activities – in other words, when they use multiple modes to share information. This might explain our findings on green claim videos given that 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; Sen et al. 2024; White et al. 2019).

GENERAL DISCUSSION

This paper investigates how two key video elements such as the text overlays and narrating voice should be used on top of the baseline elements of a video (i.e., moving image and caption) to increase engagement on social media. Our findings highlight that videos including green (vs. non-green) claims should avoid using multiple modes. Specifically, consumers are less likely to engage with green videos that implement a narrating voice, text overlay, or a combination of both on top of the baseline elements (i.e., moving images and caption).

This work makes the following contributions. First, we answer the calls for research on multimodality (Grewal et al. 2022; Packard and Berger 2024) by considering the joint role of multiple modes that are presented simultaneously to consumers. Recent work on social media communication has started to examine the effect of the interplay among multiple modes (i.e., image, text, audio) on consumer engagement (Holiday et al. 2023; Kanuri, Hughes, and Hodges 2024; Lapresta-Romero et al. 2024; Villarroel et al. 2019). In comparison to previous literature (e.g., Li and Xie 2020; Villarroel et al., 2019), we move beyond the interplay between still images and written text and contribute to the literature on video marketing (e.g., Chang et al. 2023; Liu et al. 2018; Stuppy, Landwehr, and McGraw 2024) by examining how the presence of text overlays and narrating voices in videos shape consumer engagement of branded content. Specifically, while prior studies show how the interplay among modes can have positive consequences on engagement, we find that implementing a narrating voice, text overlay, or a combination of both on top of the baseline elements (i.e., moving images and caption) is detrimental to engagement of videos that specifically promote a green lifestyle or consumption.

Second, we contribute to the literature on sustainable communication by showing that using multiple modes can have differential effects in terms of consumer engagement for brand messages that include a green claim. Despite the importance for brands of communicating sustainability on social media (Fernandez, Hartmann, and Apaolaza 2022), literature falls short in investigating how brands can effectively use sustainability cues to promote green products and lifestyles through an effective combination of multiple modes (Bharadwaj, Naik, and Nath 2022). A burgeoning stream of research focused on how to use visual cues to make advertising look “green” (e.g., Segev, Fernandes and Hong 2016) and how to frame textual information to foster change in sustainable behavior (e.g., Olsen et al. 2014; White et al. 2011; White et al. 2019). This work extends research on how to design videos by unveiling contingencies within sustainable communication that suggest a careful use of multiple modes, or

social media engagement for green-oriented messages might likely suffer. For non-green claims multimodal communication benefits from the use of multiple modes, this way confirming existing research that shows a positive effect of the simultaneous use of multiple modes in brand messages (e.g., Li and Xie 2020; Ceylan et al. 2024; Ceylan, Diehl, and Wood 2024).

Third, from a managerial standpoint, considering the concerns of public policy makers and governmental agencies that are engaged with sustainable communication, we offer actionable insights in how content managers should combine multiple modes to boost engagement.

In follow-up studies, we aim at establishing an explanation of *why* these negative effects of sustainable multimodal communication materialize. Some potential processes that are being tested are related to perceptions of redundancy across modes (MacInnis, Moorman, and Jaworski 1991), persuasion knowledge (Friestad and Wright 1994) and greenwashing perceptions (Chen and Chang 2013). This work represents the first attempt to understand the intricacies of multimodality in different types of communication, with a strong focus on how to improve the effectiveness of green claim communication that is at the core of growing societal and governmental concern.

SELECTED REFERENCES

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

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.

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.

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.

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

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.