Digital In-Context Experiments

Brand Safety: Pre-Registration

Thursday Aug 15, 2024, 10:38 GMT+2

Brand safety is a critical concern in digital advertising, particularly on social media platforms where automated systems place ads in dynamic, user-generated content environments. Misplacements can lead to significant reputational damage, with studies showing that about 70% of brands take brand safety seriously (GumGum Inc. 2017; Ahmad et al. 2024). Despite efforts to prevent ad placement alongside overtly harmful content, misplacements adjacent to disasters, divisive political content, and misinformation remain prevalent. Many decision-makers are unaware of their ads appearing on misinformation websites.

To illustrate the unique capabilities of DICE, we propose a simple study that extends beyond altering individual posts to modifying entire feeds: Unlike traditional online platform studies, we hold the ad copy and creative constant while manipulating the surrounding context between-subjects. Importantly, this study design is uniquely feasible within the DICE paradigm due to its precise control over the contextual environment—a capability not available in other experimental methodologies such as vignette or online platform studies. This level of control is crucial when examining brand safety, a phenomenon inherently defined by an advertisement's context. By manipulating the surrounding content while keeping the ad constant, we can directly investigate how context impacts brand perceptions, offering insights into brand safety that would be challenging to obtain through alternative research approaches.

We test the intuitive hypothesis that an inappropriate (compared to a more general and appropriate) context negatively affects brand attitudes.

1 Experimental Design

Our study focuses on scenarios where airlines promote travel destinations through targeted advertising, placing ads in contexts that align with specific destinations. Given that major airlines serve numerous destinations globally, these ad placements are typically managed through automated programmatic systems. We leverage this automated placement approach to create two hypothetical scenarios featuring KLM Royal Dutch Airlines (KLM) promoting flights to

Brazil. Prior to running this study, Brazil was experiencing severe flooding that claimed at least 95 lives (Buschschlüter 2024). To simulate real-world conditions, we scraped real tweets from that period and assembled them to two distinct Twitter feeds: one covering the natural disaster and another featuring more general content, including coverage of Madonna's free concert in Rio de Janeiro.

For the illustrative character of this study, we assume that automated placement systems primarily target the keyword "Brazil" without considering nuanced contextual factors. This assumption allows us to simulate how the same advertisement might appear in markedly different contexts on a social media platform. Consequently, we place an identical fictitious sponsored post by KLM, promoting flights to Brazil, into both Twitter feeds. The advertisement features a creative (as shown in Figure 1) as well as copy that read: *"Brazil's wild beauty calls! Experience nature like never before. Book your breathtaking adventure with KLM."* While this messaging would typically be considered appropriate for tourism promotion, it appears strikingly insensitive when juxtaposed against news of a natural disaster.

Figure 1: KLM Ad Creative

2 Implementation

We implement the two cell between-subjects design by creating a csv file that contains two times twenty rows (i.e., twenty tweets for each condition). Whereas all other columns are unique, two of these rows represent one and the same apart from their assignment to either one of the two conditions. To specify the sponsored posts, we set `sponsored` to 1, provide a landing page in the `target` column to which participants are directed when clicking on the ad. In addition, we set its `sequence` parameter to 5 to guarantee that it is displayed in fifth position of the feed. We do not specify that parameter for any other row such that DICE orders the remaining tweets randomly between-subject. Finally, we add a `source` column that provides URLs to the tweets we scraped. Even though this column is not required (as DICE does not evaluate it) we consider such a column useful for documentation purposes. The described csv file, whose structure we illustrate in Table 1, will then be uploaded to Github such that we can pass the corresponding URL to the DICE app.

Table 1: CSV Excerpt

doc_id	text	username	condition	sponsored	target	sequence
1	Madonna breaks the record for biggest audience…	chart data	appro…	0		
2	Saudades do Rio didn't want to leave…	diplo	appro…	0		
3	50 million people watched on TV Madonna…	Madonna Daily	appro…	0		
4	Chelsea really wanted Real Madrid-bound…	Nizaar Kinsella	appro…	0		
5	Brazil's wild beauty calls! Experience nature…	KLM	appro…	1	[KLM url]	5
…	…	…	…	…		
25	Brazil's wild beauty calls! Experience nature…	KLM	inappro…	1	[KLM url]	5
…	…	…	…	…		
40	i mentioned this on another tweet! if you can help…	Evil Scientist	inappro…	0		

3 Measures and Procedure

Participants read instructions and browse one of two twitter feeds (appropriate vs. inappropriate context) in which we place the KLM ad, before they are directed to a Qualtrics survey.

The two feeds consist of 20 real tweets each where the focal ad (by KLM) is placedin fifth position.[1] In the Qualtrics survey, we elicit whether participants recall a brand advertising in the feed—first uncued and then cued (i.e., participants saw a list of a diverse range of brands and had to indicate whether they recall seeing them). Next, participants evaluate the target brand on three seven-point scales presented in a random order (1 = "Negative/Unfavorable/Dislike" and 7 = "Positive/Favorable/Like"), which we will average into a single measure. Finally, participants will describe their experience interacting with the feed in an open text field, indicate whether they were aware of the flooding, provide demographic information, and read a debriefing before they are redirected to Prolific.

4 Primary Analysis

Our primary interest lies in the effect the context has on our three-item brand attitude measure. Hence, we compute an average of these items and compare means using a simple OLS regression:

$$y_i = \alpha + \beta_1 D_i + \epsilon_i$$

where y_i denotes the averaged brand attitude measure and where D_i is treatment dummy:

$$D_i = \begin{cases} 1 & \text{if } i \text{ was exposed to inapproriate context,} \\ 0 & \text{if } i \text{ was exposed to approriate context.} \end{cases}$$

β_1 is the coefficient of interest. We expect the inappropriate context to have a negative effect on the brand attitude (i.e. $\beta_1 < 0$).

5 Secondary Analysis

We expect the main effect described above to be heterogeneous. Specifically, we expect the the effect to be stronger (i.e. $|\beta_1|$ to be larger) for those participants who recall seeing an ad for the target brand, compared to those, who do not recall seeing it.

Furthermore, we expect a positive correlation between dwell time (that is, the time the sponsored post was visible to the participant) and recall: the more time a participant allocates to the sponsored post, the more likely it is that she will recall seeing it.

Finally, we expect the dwell time to moderate the main effect as indicated by β_3 in the following OLS regression: The more time participants allocate to the sponsored post, the stronger its effect on brand attitude.

[1] You can browse the flooding-related, inappropriate feed here and the more general, appropriate feed here.

$$y_i = \alpha + \beta_1 D_i + \beta_2 \text{ dwell time}_i + \beta_3 D_i \times \text{ dwell time}_i + \epsilon_i$$

6 Population

We will recruit participants from Prolific who meet the following criteria:

- Approval Rate >= 99%
- First Language == 'English'
- Location == 'USA'

7 Sample Size

We recruit 1000 participants. To this end, we create databases containing 1400 rows.

8 Exclusion Criteria

We will only consider complete observations, that is, data from participants who browsed through the feed, answered the qualtrics survey and who were redirected to Prolific with a functional completion code.

Because we gather process data, such as dwell time, we have tools to assess the data quality (see, e.g., Cuskley and Sulik 2024)–at least during the exposure to the social media feed. If these data reveal inattentive participants, for instance, we may exclude them too but label the resulting analyses as exploratory.

References

Ahmad, Wajeeha, Ananya Sen, Charles Eesley, and Erik Brynjolfsson. 2024. "Companies Inadvertently Fund Online Misinformation Despite Consumer Backlash." *Nature* 630: 123–31. https://doi.org/10.1038/s41586-024-07404-1.

Buschschlüter, Vanessa. 2024. "Brazil Floods: 'We've Never Experienced Anything Like It'." BBC News. https://web.archive.org/web/20240805170342/https://www.bbc.com/news/articles/cle07g0zzqeo.

Cuskley, Christine, and Justin Sulik. 2024. "The Burden for High-Quality Online Data Collection Lies with Researchers, Not Recruitment Platforms." *Perspectives on Psychological Science* 0 (0): 17456916241242734. https://doi.org/10.1177/17456916241242734.

GumGum Inc. 2017. "The New Brand Safety Crisis: A Fractured Environment." GumGum. https://web.archive.org/web/20220317063148/https://insights.gumgum.com/hubfs/Brand_Safety_GumGum.pdf.