

Effects of Explicit Sponsorship Disclosure on User Engagement in Social Media Influencer Marketing

Zike Cao*, Rodrigo Belo†

Abstract

Social media influencer marketing has grown substantially in the last decade and is a major advertising channel for many brands. Social media influencers weave sponsored posts with organic content in their feeds, which raises concerns among regulators and consumer advocates that users may not be able to clearly distinguish between sponsored and organic influencer content. Thus, regulators often mandate the explicit disclosure of sponsored content. However, there is little empirical evidence based on field data about the effects of explicit sponsorship disclosure. Therefore, we empirically investigate the effects of explicitly disclosing sponsorship in influencers' content on users' engagement using a large-scale field dataset collected from Facebook and Instagram. Our empirical results suggest that explicit sponsorship disclosure increases user awareness of the advertising nature and earns users' favorability by enhancing the transparency about the sponsored content. We further design two online experiments to corroborate our empirical results and directly test the underlying mechanisms. Our findings have novel and important implications for marketers, influencers, social media platforms, and regulators in the influencer marketing industry.

Keywords: social media; sponsorship disclosure; transparency; online experiment; influencer marketing; Instagram; Facebook

1 Introduction

The last decade has witnessed the unprecedented growth of social media influencers, viz., individuals who can attract a large number of followers on social media either because they are known for their offline activities (e.g., actors and athletes) or because of the original content they create and share on social media in a specific area (e.g., fitness, beauty and fashion, video games, etc.). The dramatic increase in the number and variations of online

*School of Management, Zhejiang University, email: caozike@zju.edu.cn

†Nova School of Business and Economics, Universidade Nova de Lisboa, email: rodrigo.belo@novasbe.pt

products that lead consumers to rely on intermediaries, such as social media influencers, for product discovery and recommendations (Ershov and Mitchell 2020; Mitchell 2021). Marketers have shown an increasing interest in this phenomenon and now allocate a considerable portion of their advertising budget to this new form of digital marketing, called “influencer marketing” (eMarketer 2022). In influencer marketing, marketers collaborate with social media influencers, who post sponsored content on their social media accounts to promote a brand. We use the term *sponsored content* to refer to such advertising content throughout the paper.

Broadly, influencer marketing is a form of *native advertising*, an advertising format wherein ads are designed to match the form and style of the organic content into which they are integrated (Sahni and Nair 2020b). Native advertising encompasses a variety of ad formats, such as keyword search ads, in-feed ads, recommendation widgets, advertorials (i.e., ads in the style of an editorial or objective journalistic article), advergames (i.e., games featuring brands as part of the game), and social media influencers’ sponsored posts (Campbell and Grimm 2019; Evans et al. 2019). Like many other native ads, social media influencers’ sponsored posts are designed to be as unintrusive to the user experience as possible. Hence, social media influencers typically mix organic posts (unsponsored) with posts endorsing companies or products (sponsored), making it hard for their followers to distinguish between the two content types (Mitchell 2021).

Although the effectiveness of influencer marketing (and other native advertising formats) is likely to be superior to that of traditional display ads (Sharethrough 2015), consumers may also be less likely to recognize ads in influencer marketing. In other words, the superior performance of native ads may be due to consumer deception or because consumers wrongly treat ads as organic content (Federal Trade Commission 2018; Sahni

and Nair 2020b). As such, social media influencer marketing is now under regulatory focus worldwide, with regulators mandating explicit sponsorship disclosure to ensure a clear separation between sponsored and organic content to raise consumers’ awareness of the nature of the content they are exposed to (Campbell et al. 2013). For example, under German law, social media influencers’ sponsored posts should be fully and clearly disclosed (Ershov and Mitchell 2020). In the US, the Endorsement Guides of the Federal Trade Commission (FTC) state that any “material connection” (e.g., business or family relationships, monetary payment, or the gift of a free product) should be “clearly and conspicuously disclosed” (Federal Trade Commission 2018). The FTC’s premise is that informing consumers of the advertising nature of an influencer post endorsing a product may “affect the weight or credibility that consumers give to the endorsement” (Federal Trade Commission 2017).

Nevertheless, not all influencers comply with this regulation. For example, Mathur et al. (2018) found that only about 10% of the marketing content creators on YouTube and Pinterest include any disclosure in their sponsored content. Moreover, influencers could use an array of disclosure strategies that vary considerably in how clear and explicit they are in disclosing (Federal Trade Commission 2018; Mathur et al. 2018). The FTC regularly sends warning letters to social media influencers who do not fully disclose sponsorship in their content (Federal Trade Commission 2020). In Germany, too, there have been several high-profile lawsuits against influencers for failing to prominently disclose sponsorship (Luong 2019).

Clearly, regulators and many influencers have opposing views on the necessity of explicit sponsorship disclosure. To resolve this conflict, it is crucial to understand how social media users in real scenarios react to explicit sponsorship disclosure. The direction of social media users’ reactions to an explicit disclosure of sponsorship (i.e., positive, negative, or nil)

provides important implications for regulators’ policy-making and influencers’/advertiser’ real practices. To the best of our knowledge, no empirical study has investigated how the explicit disclosure of sponsorship affects users’ *revealed* behavior based on field data. Most studies on this topic are survey-based and thus rely on *reported* preferences (e.g., Boerman et al. 2017; Campbell et al. 2013; Evans et al. 2019; Wojdyski and Evans 2016). This study fills this gap by compiling a large-scale field dataset from two major social media platforms—Instagram and Facebook—on users’ engagement with influencers’ posts (i.e., likes and comments, the most observable user behaviors on social media). Formally, we seek to investigate the following research question:

How does explicit sponsorship disclosure of influencers’ sponsored content on social media affect users’ engagement, in terms of likes and comments, with the content?

Drawing from the literature, explicit sponsorship disclosure exerts two contrasting effects on how users react to sponsored content. Explicit sponsorship disclosure raises users’ awareness of the advertising nature of the content by clearly labeling the content as sponsored. Users who cannot recognize advertising content without an explicit disclosure may treat the advertising content as if it is organic and act accordingly (Campbell et al. 2013; Darke and Ritchie 2007). Thus, explicitly disclosing content sponsorship can make such users aware of its advertising nature, which leads them to engage less with the content, as users generally prefer organic content over sponsored content (Boerman et al. 2017; Campbell et al. 2013). Therefore, explicit disclosure could have a *negative* effect on user engagement due to users’ increased awareness of the advertising nature of the sponsored influencer content. We call this the *awareness effect* of explicit sponsorship disclosure on user engagement.

However, in addition to sensitizing users to the advertising nature, explicit sponsorship

disclosure also makes the information about paid sponsorship transparent and more accessible. The enhanced transparency about advertising may have a *positive* effect on users' reactions and behavior for multiple reasons. From an economic perspective, explicit disclosure makes the information about brand–influencer collaboration more salient, which can be interpreted as a positive *signal* of the influencer's quality (or status) according to the signaling model (Sahni and Nair 2020a,b). Therefore, users could be more willing to engage with explicitly-disclosed sponsored content due to the signaling effect. From a behavioral perspective, previous studies have shown that an explicit sponsorship disclosure may be favored by consumers because it enhances the marketing tactic's perceived credibility (Isaac and Grayson 2017), increases consumers' perceived credibility and trustworthiness of the marketing agent (Abendroth and Heyman 2013; Carl 2008; Evans et al. 2019; Wang et al. 2018; Wang and Wang 2019), and reduces perceptions of deception and manipulative intent (Forehand and Grier 2003). This effect is supposed to be more pronounced among users who are sophisticated enough to detect advertising purposes even in the absence of an explicit disclosure (Sahni and Nair 2020b). We call this positive effect the *transparency effect* of explicit sponsorship disclosure on user engagement.

Therefore, the net impact of explicit sponsorship disclosure on user engagement on social media is ultimately an empirical question, which carries important ramifications for relevant policy-making and business practices. Our empirical exercise uses a feature available on both platforms that allows influencers to disclose sponsorship by attaching a conspicuous tag to the header of the sponsored post. When an influencer tags a sponsored post, the post gets a header label that explicitly indicates the advertising nature of the post. Hereafter, we refer to sponsored posts that are disclosed in this manner as *header-disclosed* posts. Figure 1 shows an example of a header-disclosed sponsored post on

Instagram. We empirically assess how header-disclosing a sponsored post affects its level of user engagement.

We exploit the fact that many social media influencers are multi-homed (i.e., have multiple social media accounts) and frequently post *identical* content across multiple social media platforms but without always explicitly disclosing sponsorship on all platforms. We measure the effect of explicit sponsorship disclosure by comparing users' engagement for a given post with different disclosure statuses across platforms to causally identify the effect of explicit disclosure of sponsored content while controlling for the platform-specific level of engagement.

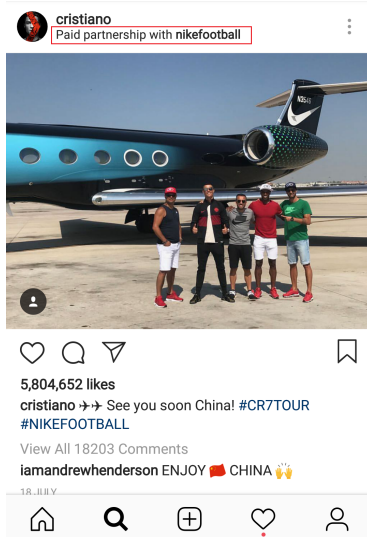


Figure 1: Example of a sponsored post with header-disclosed tag on Instagram

We find that sponsorship disclosure via the platform's header-disclosure tool significantly *increases* user engagement with a sponsored post. The positive effect is more pronounced and robust for users' likes than for comments. Our preferred estimates suggest that on average, a sponsored post receives 34% more likes and 14% more comments when sponsorship is explicitly disclosed using the platform's disclosure tool than a post without such header disclosure. The effects are both statistically and economically significant. These results are consistent with the hypothesis that explicit sponsorship disclosure

increases user engagement by enhancing information transparency.

To corroborate the empirical results based on the field data, we conduct two online experiments wherein the subjects are randomly allocated to different treatment conditions. By measuring subjects’ advertising awareness, perceived sponsorship transparency, and willingness to engage, we directly show header disclosure exerts two parallel but contrasting effects on user engagement, which are mediated by advertising awareness and perceived sponsorship transparency.

We contribute to the literature on social media influencer marketing (e.g., Boerman et al. 2017; Campbell et al. 2013; Ershov and Mitchell 2020; Mathur et al. 2018; Mitchell 2021) by documenting causal evidence of the effects of explicit sponsorship disclosure on user engagement based on field data. Our results uncover the nuanced effects of explicit sponsorship disclosure. Explicit sponsorship can simultaneously have positive and negative effects on users’ engagement behavior, which has not been documented in the literature. We also contribute to the literature on information transparency or disclosure (e.g., Calvo et al. 2020; Granados et al. 2010; Li and Wu 2018; Sahni and Nair 2020a,b; Zhou et al. 2018) by providing evidence of a novel and important positive effect of information disclosure on social media user behavior mediated by enhanced information transparency.

2 Context and Data

2.1 “Branded Content” on Facebook and Instagram

Launched in 2004, Facebook is the world’s largest social network and a major platform for various marketing campaigns. In April 2016, Facebook updated its advertising policy to allow verified entities (referred to as *Pages* by Facebook) to post “branded content,” which is defined as any post that “features a third-party product, brand, or sponsor” and is “typically posted by media companies, celebrities, or other influencers” (Facebook 2016). We use the term *sponsored content* to refer to this type of content. In a sponsored post, the corresponding brand or marketer is tagged by the influencer, who posts the content,

with the brand name appearing along with the influencer’s name. Since March 2017, the sponsored content tag includes the word “Paid” to disclose sponsorship more clearly.

Instagram—a mobile photo and video sharing platform launched in 2010—is one of the fastest-growing social networks. Instagram is considered a valuable social media marketing tool because of its visual nature and high levels of user engagement, and it is the leading platform for influencer marketing campaigns (eMarketer 2019). Although Facebook (now Meta) acquired Instagram in September 2012 and the two platforms share the same set of policies on influencer marketing,¹ Instagram did not have an official “branded content” tool until June 2017, when it introduced a “Paid partnership with [brand]” label shown above the post, similar to Facebook, that allows influencers to tag their business partners. Figure 1 in the Introduction section presents an example of a sponsored post on Instagram.

Before the header-disclosure tool came into existence, influencers on both platforms typically used relevant hashtags (e.g., #sponsored, #ad) in posts’ textual description for disclosure purposes. By offering a tool for disclosing sponsorship, the platforms can better facilitate and monetize influencer–business partnerships on their platforms. For example, both influencers and their business partners (if permitted by the influencers) can access the statistics on post-engagement metrics.

2.2 Data

We first identify the top 2,500 most followed verified Instagram accounts of individuals (i.e., excluding brands’ and organizations’ public accounts) using Heepsy’s directory of social media influencers ([https:// www.heepsy.com/](https://www.heepsy.com/)). Then we focus on the top influencers on Instagram given the prominence of the platform in social media influencer marketing. We search for all of the top 2,500 most followed Instagram accounts’ names on Facebook. For this, we use the Google Search API to search for each Instagram account on Face-

¹See <https://www.facebook.com/help/instagram/116947042301556/> (accessed Feb 13, 2023)

book with the name that is used on Instagram. We find 1,355 matched verified Facebook accounts. We then scrape all the posts created by the 1,355 influencers between June 1, 2017 and June 1, 2018 on both platforms within a week (July 13–20, 2018). We obtain a total of 810,168 posts on both platforms (347,291 on Facebook and 462,877 on Instagram), of which 0.78% (2,708) of Facebook posts are header-disclosed and 0.64% (2,960) of Instagram posts are header-disclosed.

Table 1: Individual-level summary statistics by platform (all posts)

Variable	Statistic	Instagram	Facebook	Difference	t-Value	p-Value
No. of posts	Avg	343.89	260.53	83.36	5.08	0
	SD	550.57	241.59			
	Min	1	1			
	Max	13,041	1,404			
	N	1,346	1,333			
No. of likes	Avg	209,584.80	13,027.56	196,557.30	22.79	0
	SD	315,204.30	27,344.62			
	Min	300.81	0.04			
	Max	3,961,128.00	504,667.60			
	N	1,346	1,333			
No. of comments	Avg	2,813.37	529.87	2,283.51	7.98	0
	SD	10,427.40	1,177.95			
	Min	0.04	0			
	Max	271,714.40	14,366.50			
	N	1,346	1,333			
Used branded content tool (dummy)	Avg	0.339	0.340	-0.001	0.06	0.95
	SD	0.47	0.47			
	Min	0	0			
	Max	1	1			
	N	1,346	1,333			
Fraction of header-disclosed posts	Avg	0.008	0.014	-0.006	3.75	0.0002
	SD	0.02	0.06			
	Min	0	0			
	Max	0.33	1			
	N	1,346	1,333			
Fraction of hashtag-disclosed posts	Avg	0.01	0.01	-0.004	-1.99	0.05
	SD	0.03	0.06			
	Min	0	0			
	Max	0.26	1			
	N	1,346	1,333			
No. of words	Avg	20.05	19.34	0.72	0.94	0.35
	SD	16.16	22.75			
	Min	0	0			
	Max	259.40	625.29			
	N	1,346	1,333			
No. of hashtags	Avg	1.45	0.84	0.61	5.28	0
	SD	1.80	0.99			
	Min	0	0			
	Max	14.90	5.54			
	N	545	131			
No. of @ mentions	Avg	0.89	0.25	0.64	11.08	0
	SD	0.72	0.56			
	Min	0	0			
	Max	6.98	5.36			
	N	545	131			
No. of followers (millions)	Avg	8.41	6.67	1.74	3.72	0.0002
	SD	13.18	10.98			
	Min	2.10	0.01			
	Max	144	120.77			
	N	1,346	1,333			

Table 1 presents the summary statistics at the influencer level for the full sample of 810,168 posts. Of the 1,355 individual influencer accounts, we are able to scrape the posted

content of 1,346 accounts on Instagram and 1,333 on Facebook. On average, these influencers post more often and receive more user engagement on Instagram than on Facebook. On average, their posts receive over 10 times as many likes and 5 times as many comments on Instagram than on Facebook. The summary statistics show that on average, influencers are equally likely to use the branded-content tool on both platforms, as on each platform, about 34% of the influencers use the branded-content tool for at least one post. However, branded content with the “paid” tag (i.e., header-disclosed posts) constitutes a slightly larger proportion of an influencer’s posts on Facebook (1.4%) than on Instagram (0.8%).

We match each influencer’s posts on both platforms. An Instagram post is matched with a Facebook post if they are both posted by the same influencer, contain identical (non-blank) textual description, and are of the same media type (photo or video). We obtain 135,886 matched observations, or 67,943 distinct posts by 1,072 (out of the 1,355) influencers. Out of the 67,943 posts on Instagram and Facebook, 219 (0.32%) are header-disclosed on Instagram and 164 (0.24%) are header-disclosed on Facebook. Moreover, 89 (0.13%) posts are header-disclosed on both platforms. Most of the posts (99.57%) are not header-disclosed on either platform.

We also identify posts disclosing sponsorship using hashtags. We use the set of hashtags typically used on social media as discussed in the FTC’s endorsement guides (Federal Trade Commission 2018) and label a post as a *hashtag-disclosed* sponsored post if at least one of the relevant hashtags is present in the textual description.

Moreover, posts that are neither header-disclosed nor hashtag-disclosed can be of two types: organic (i.e., unsponsored) or sponsored but undisclosed. Distinguishing between these two types of posts is quite challenging, as we do not have access to the truth about whether a post is sponsored. We address this issue using two strategies. First, we use

a proxy for post sponsorship—whether the media content includes brand logos. In other words, a post with an image or video that has product logo(s) embedded within is considered a sponsored post. For this, we use Google’s publicly available paid image analysis service—Cloud Vision API (<https://cloud.google.com/vision/>)—to detect product logo(s). The results indicate that 11.87% of the media content in our sample has embedded product logo(s).

Second, we use another proxy for post sponsorship—whether the text in the posts contains references to business accounts on Instagram using an @-mention. We classify accounts whose Instagram profiles have information about their “area of business” as business accounts. While we believe this is a good proxy for a sponsored post, we are aware that influencers may not always tag the corresponding business account(s) in a sponsored post. Furthermore, not all business accounts provide information about their areas of business, which prevents us from identifying all (undisclosed) sponsored posts. In any case, the accounts identified as business accounts are labeled correctly, which is the most important aspect for our estimation strategy.

3 Empirical Strategy and Results

3.1 Main Results

Our main goal is to estimate and compare the aggregate effect of explicitly disclosing sponsored content on user engagement with that of no disclosure. We start with a pooled ordinary least squares specification using the full sample of collected posts (i.e., the 810,168 posts on both platforms) and then gradually add further controls and restrict our sample to account for potential heterogeneity and endogeneity. Our baseline specification is as follows.

$$\begin{aligned}
 \text{Engagement}_{ijk} = & \beta_1 \text{Header-disclosed}_{ijk} + \beta_2 \text{Hashtag-disclosed}_{ij} + \\
 & \beta_3 \text{Sponsored}_{ij} + X_{ij}\alpha + \sigma_{jk} + \delta_k + \epsilon_{ijk},
 \end{aligned} \tag{1}$$

where i denotes a post, j denotes an influencer, k denotes a platform (Facebook or Instagram), and $Engagement_{ijk}$ denotes an engagement metric (e.g., number of likes or comments) of post i posted by influencer j on platform k . $Header-disclosed_{ijk}$ is a dummy variable indicating whether post i posted by influencer j is header-disclosed on platform k (note that the header-disclosure status of the same post may be different on the two platforms, so $Header-disclosed_{ijk}$ has k in the subscript); $Hashtag-disclosed_{ij}$ is a dummy variable indicating whether post i posted by influencer j is hashtag-disclosed but not header-disclosed (if a post is both header-disclosed and hashtag-disclosed, it is classified as a header-disclosed post); and $Sponsored_{ij}$ is a dummy variable indicating whether post i by influencer j is sponsored. As mentioned earlier, we identify a post as sponsored if it is header-disclosed on at least one of the platforms, is hashtag-disclosed, includes product logo(s) as indicated by the image analysis, or references business accounts on Instagram using an @-mention. Thus, β_1 corresponds to the average difference in engagement levels between header-disclosed and undisclosed sponsored posts, β_2 corresponds to the average difference in engagement levels between hashtag-disclosed and undisclosed sponsored posts, and β_3 corresponds to the average difference in engagement levels between undisclosed sponsored posts and organic posts.

We add some additional control variables for post-specific attributes in our data, denoted as X_{ij} , which are the logarithm of the number of words ($Log(words)$), the logarithm of the number of @mentions ($Log(@mentions)$), the logarithm of the number of total hashtags ($Log(hashtags)$), and, in some specifications, post fixed effects. δ_k represents platform fixed effects and σ_{jk} represents platform-specific influencer fixed effects, which control for unobserved confounding factors related to differences between influencers on the same platform and differences between platforms for the same influencers. ϵ_{ijk} cap-

tures idiosyncratic random errors. Standard errors are clustered at the influencer level to account for correlations in user engagement in posts from the same influencer. We specify all of the continuous variables in logarithms, as the dependent variables are skewed and our focus is on relative change. Moreover, some of the posts have no engagement; therefore, we add 1 to the dependent variables to avoid a logarithm of 0.

Column (1) in Table 2 shows the estimates of our baseline specification, which measures engagement as the number of likes on a post and uses all of the observations in our sample. On average, long posts, and posts with fewer mentions and hashtags elicit higher engagement than short posts, and posts with more mentions and hashtags. Moreover, Instagram posts elicit considerably higher engagement than Facebook posts. Our coefficients of interest in Equation (1), β_1 , β_2 and β_3 , suggest that undisclosed sponsored posts on average have lower levels of user likes than organic posts, implying that users in general favor organic content than sponsored content. Moreover, header disclosure has a positive effect on user engagement as measured by the number of likes compared to no disclosure: header-disclosed posts elicit roughly double the level of engagement ($e^{0.701} \approx 2$, $p < 0.01$) of undisclosed sponsored posts. Moreover, hashtag-disclosed posts also appear to elicit statistically significantly higher engagement than undisclosed sponsored posts, but with smaller magnitude ($e^{0.223} \approx 1.25$, $p < 0.05$). This might be because hashtag disclosure is less explicit and clear than header disclosure.

Although the results in Column (1) consider post-specific observed information, such as post length, mentions, and hashtags, it is possible that other unobserved factors may influence sponsorship disclosure and user engagement. For example, popular influencers may be more likely to disclose post sponsorship, leading to a positive correlation between sponsorship disclosure and user engagement. Column (2) includes *platform-specific influencer fixed effects*, which enable us to control for the popularity of an influencer on a

Table 2: Effects of explicit disclosure on user engagement: Likes

VARIABLES	Log(Likes)						
	(1) All	(2) All	(3) All	(4) Spons.	(5) Matched	(6) Matched & Spons.	(7) Matched & Spons.
Header disclosed	0.701*** (0.141)	-0.0816* (0.0483)	0.148** (0.0698)	0.156** (0.0621)	0.348*** (0.0587)	0.296*** (0.0597)	
Hashtag disclosed	0.223** (0.107)	-0.0806** (0.0371)					
Sponsored	-0.129*** (0.0354)	-0.111*** (0.0117)					
Log(words)	0.107*** (0.0280)	0.0546*** (0.0151)					
Log(mentions)	-0.313*** (0.0567)	-0.104*** (0.0159)					
Log(hashtags)	-0.0961** (0.0456)	0.0239 (0.0196)					
Instagram	3.724*** (0.0697)						
Header disclosed (FB)							0.305*** (0.0837)
Header disclosed (IG)							0.295*** (0.0670)
Platform-specific influencer FE	No	Yes	Yes	Yes	Yes	Yes	Yes
Post FE	No	No	Yes	Yes	Yes	Yes	Yes
Observations	810,168	810,168	810,168	147,073	135,592	51,620	51,620
R-squared	0.468	0.762	0.998	0.995	0.977	0.978	0.978

Robust standard errors clustered by influencers are in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

specific platform, along with other platform-specific influencer factors. Here, the results suggest either header disclosure or hashtag disclosure appears to negatively affect user engagement.

However, one concern with these estimates is that posts from the same influencer on the same platform may be heterogeneous in terms of engaging users, so the inherent heterogeneity in the posts may bias our estimates. Social media influencers often post *identical* content on both platforms, and the same post from the same influencer may be disclosed using the disclosure tool on one platform but not on the other. We use this fact to address heterogeneity. Figure 2 shows an example of a post that is header-disclosed on Instagram but not on Facebook. This allows us to *estimate the effect of explicit sponsorship disclosure using the platform's disclosure tool on user engagement*. This approach, which compares the same items on two websites, has been commonly used in the literature (e.g.,

Chevalier and Mayzlin 2006; Mayzlin et al. 2014).

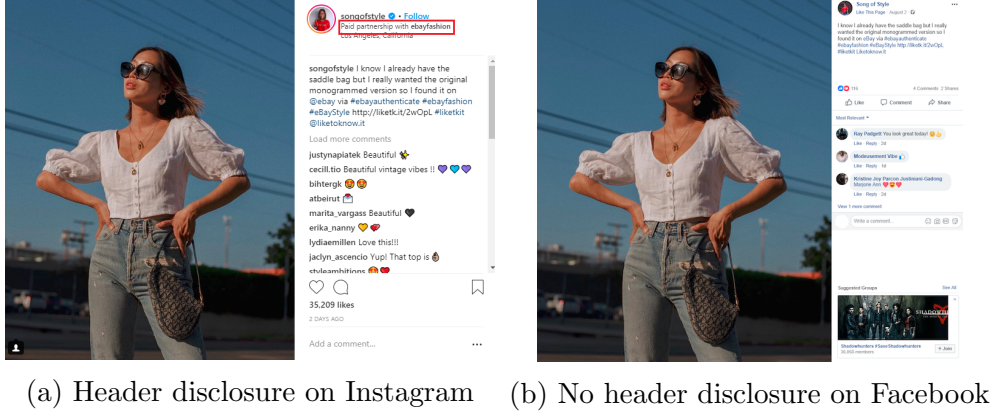


Figure 2: Header disclosure on Instagram versus Facebook

In Column (3), we further add post-specific fixed effects, which allow us to measure the effects of header-disclosure *within* the same post, taking advantage of the fact that header-disclosure is platform-specific—some posts are header-disclosed on one platform but not on the other. The key assumption is that for a given sponsored post, the platforms on which the post is header-disclosed are exogenously determined. Note that by including post fixed effects, we cannot identify the effects of hashtag disclosure or obtain coefficients for any other post-specific covariates. On the positive side, we can identify the effect of header-disclosure by comparing the differences in the level of engagement for the same post. The results indicate that header-disclosing a post leads to an average increase of approximately 16% in engagement as measured by the number of likes ($p < 0.05$).

Although our fixed effects specification applied to the full sample in Column (3) yields consistent estimates for the effect of header-disclosing a post, we can identify only a small fraction of posts that are sponsored, have been posted on both platforms, and have a different header-disclosure status on both platforms. This means that the data used to calculate the baseline values in the fixed-effect estimation include a wide range of posts (e.g., organic posts on only one platform) that are not directly comparable with the posts that are the focus of our analysis. To ensure a valid comparison, we run the fixed-effect

specification on the following subsamples: posts identified as sponsored, matched posts (or posts present on both platforms), and posts that are identified as both sponsored and matched.²

Columns (4), (5), and (6) in Table 2 show the results for the subsamples. Although the coefficient of interest is always positive and statistically significant, its magnitude and significance increase as the sample becomes more uniform, including the subsample of only the sponsored posts on both platforms. Overall, our preferred specification is the one in Column (6), in which only the matched and sponsored posts are used. In this specification, header-disclosing a post leads to an increase of approximately 34% ($p < 0.01$) in engagement, measured as the number of likes. Column (7) presents the separate estimates for the effects of header disclosure on Instagram and Facebook. The estimates are very close to each other (and to the aggregate), indicating that the disclosure effect is not platform-specific. We also repeat the analysis by measuring user engagement with users' comments. The effects follow the same direction as those on users' likes, although they are less significant (see Table A1 in Appendix A).

Overall, these results suggest that although consumers prefer organic influencer content to sponsored content, an explicit disclosure of sponsorship such as the header disclosure studied here can have positive effects on user engagement with the sponsored content compared to no such explicit disclosure.

²Header-disclosed posts could be boosted on both platforms, see <https://www.facebook.com/help/publisher/1045927788843932> (Facebook, accessed Feb 13, 2023) and <https://help.instagram.com/116947042301556> (Instagram, accessed Feb 13, 2023). The boosting functionality could artificially inflate our estimates for the effect of header-disclosure given that brands could pay to increase the exposure of their posts. Fortunately, the boosting functionality was not available on Instagram during our sampling window, and Facebook rolled out the boosting functionality in the middle of our sampling window. In our focal estimation strategy, we exclude 147 header-disclosed Facebook posts that were posted after the introduction of this functionality—on August 14, 2017—from the analysis. Nevertheless, our results remain qualitatively the same even if we retain these 147 posts in our sample, which indicates that there was likely no boosting of these posts. The detailed results are available from the authors upon request.

3.2 Sensitivity to Composition of Control Group

We conduct sensitivity analyses to examine whether the estimate of the header disclosure effect is robust to different definitions of sponsored posts. To do so, we start with our preferred specification—as shown in Column (6) of Table 2, replicated in Column (1) of Table 3—and adjust the definition of a sponsored post. As header-disclosure is our effect of interest, we consider header-disclosed posts as treated and assume that header-disclosed posts are sponsored. We then vary the definition of a sponsored post to also include hashtag-disclosed posts, posts with brand logos, posts with mentions of Instagram business accounts, and the combinations of any two of these. In other words, we vary the set of posts defined as control posts to assess the sensitivity of our preferred estimation strategy to different compositions of the control group. The results are presented in Columns (2)–(7) of Table 3, in which engagement is measured as users’ likes.

Table 3: Sensitivity to composition of control group: Likes

VARIABLES	Log(Likes)						
	(1) Control: Spons.	(2) Control: Hashtag-disclosed	(3) Control: With logos	(4) Control: @business	(5) Control: Hashtag-disclosed + with logos:	(6) Control: Hashtag-disclosed + @business	(7) Control: With logos + @business
Header disclosed	0.296*** (0.0597)	0.401*** (0.0772)	0.337*** (0.0578)	0.297*** (0.0651)	0.346*** (0.0534)	0.295*** (0.0648)	0.297*** (0.0600)
Platform-specific influencer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Post FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	51,620	834	16,748	39,880	17,218	40,124	51,410
R-squared	0.978	0.992	0.981	0.977	0.981	0.977	0.978

Robust standard errors clustered by influencers are in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

All of the estimates are highly consistent with the main results, with the header disclosure effect significant and positive on all definitions of sponsored posts and with similar magnitudes. We also repeat this sensitivity analysis with user engagement measured by users’ comments. The effects on users’ comments are also consistent, with an always-positive coefficient, but not always statistically significant (see Table A2 in Appendix A). These outcomes indicate that our main results are robust to different definitions of spon-

sorship and are not particularly driven by specific definitions.

We perform several additional robustness checks of our preferred specification by only considering influencers that are similarly popular across platforms, only considering posts that are posted approximately at the same time across platforms, and adding a proxy variable for trends in influencers’ follower counts on both platforms. The results consistently indicate a positive effect of header disclosure on user engagement, especially when measured as users’ likes (see Table A3 in Appendix A).

4 Online Experiments: Testing Underlying Mechanisms

Thus far, we have shown that header-disclosed sponsored posts elicit higher engagement (especially in terms of likes) from users than other undisclosed or less explicitly-disclosed sponsored posts. These results are confirmed by multiple robustness checks and are consistent with the existence of a positive effect—the *transparency effect*—that results in higher levels of engagement.

However, two concerns remain. First, despite our efforts, we cannot completely ensure that our estimate is not affected by endogeneity because of data limitations. For example, influencers’ self-selection bias remains a significant concern. Second, even though we show that header-disclosed posts increase engagement, we cannot pinpoint the nature of the mechanism at play—that is, we cannot separate the effect of advertising awareness from the effect of enhanced sponsorship transparency. We address these issues by conducting two online experiments to eliminate potential endogeneity and study the underlying mechanisms of the observed effects. Formally, we aim to test the theoretical model as depicted in Figure 3, as informed by the related prior literature (see the detailed discussion in the

Introduction Stimuli and Measurements

We focus on simulating the Instagram environment and testing the disclosure effect using a “Paid partnership with” header tag. We select one sponsored post from our field

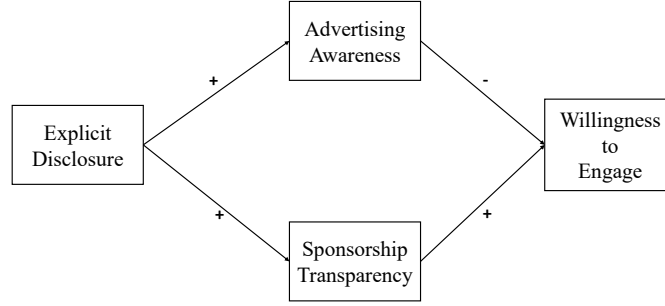


Figure 3: Theoretical model of explicit sponsorship disclosure on user engagement

data as the stimulus in the experiment. The influencer of our choice has the median number of followers on Instagram, as shown in our observational data. Figures 4a and 4b show the undisclosed and disclosed versions of the influencer’s post, respectively. We replicate the focal experiment using another post from the same influencer (a sponsored post that includes a product logo in the picture) to increase the generalizability of our findings. The results are reported in Appendix B.

We measure the following constructs (variables) in the experiment. We use one item (*“This post is a sponsored advertisement”*) to measure the extent to which subjects recognize the advertising nature (Boerman et al. 2017), three items (*“The creator of the post provided clear, comprehensible, or complete information about whether the post was sponsored”*) to measure the subjects’ perceived information transparency of the sponsorship provided by the influencer (Zhou et al. 2018), four items (*“I think this post is worth sharing with others”*; *“I will recommend this post to others”*; *“I would ‘Like’ this post”*; *“I would comment on this post”*) to measure the subjects’ willingness to engage with the post (Boerman et al. 2017). All items are measured on a 7-point Likert scale (1 = *strongly disagree* and 7 = *strongly agree*).³

³Moreover, we also measure the subjects’ perceived trustworthiness of the post (Boerman et al. 2017; Darke and Ritchie 2007; Ohanian 1990), and perceived status of the influencer (Dubois et al. 2012) to explore how explicit disclosure may affect these two outcomes. The results show explicit disclosure also exerts two parallel but contrasting effects on these two outcomes. The detailed results are available from



(a) Undisclosed version



(b) Header-disclosed version

Figure 4: The influencer’s post used as the experiment stimulus

4.2 Study 1

Study 1 has two conditions. In the first condition, labeled “No disclosure,” subjects are exposed to the undisclosed version of the post shown in Figure 4a, and in the second condition, labeled “Header disclosure,” subjects are exposed to the header-disclosed version of the post shown in Figure 4b.

Both conditions perform the same procedure, as follows. (1) The subjects are given a brief introduction about the experiment; (2) subjects are shown the stimulus; (3) subjects respond to items related to the willingness to engage, perceived post trustworthiness, and perceived influencer status (we randomize the order of all of the items for these three constructs to avoid potential order effects); (4) subjects respond to the item related to advertising awareness; (5) subjects respond to the items related to the perceived sponsorship transparency; (6) subjects answer questions on demographic characteristics (including gender, age, education level, employment status, Instagram usage, and familiarity with the influencers) and manipulation checks; and (7) the experiment ends. Note that the two conditions in Study 1 differ only in the stimulus that the subjects are exposed to.

the authors upon request.

Table 4: Comparisons of means across conditions in online experiments

VARIABLES	Study 1					
	Header disclosure = Yes			Header disclosure = No		
	N	Mean	[95% CI]	N	Mean	[95% CI] <i>t</i> -test (Y - N)
Advertising awareness	290	5.59	[5.42, 5.75]	298	4.82	[4.63, 5.01] $p < 0.01$
Perceived sponsorship transparency	290	4.10	[3.91 4.29]	298	3.07	[2.92 3.22] $p < 0.01$
Willingness to engage	290	2.57	[2.41 2.72]	298	2.67	[2.51 2.84] $p = 0.36$
Perceived post trustworthiness	290	4.10	[3.96 4.24]	298	4.42	[4.28 4.55] $p < 0.01$
Perceived influencer status	290	4.39	[4.27 4.52]	298	4.44	[4.31 4.57] $p = 0.61$
VARIABLES	Study 2					
	Header disclosure = Yes			Header disclosure = No		
	N	Mean	[95% CI]	N	Mean	[95% CI] <i>t</i> -test (Y - N)
Advertising awareness	297	6.46	[6.34, 6.58]	289	6.39	[6.28, 6.51] $p = 0.43$
Perceived sponsorship transparency	297	4.36	[4.17 4.57]	289	3.07	[2.89 3.24] $p < 0.01$
Willingness to engage	297	2.51	[2.35 2.67]	289	2.36	[2.22 2.50] $p = 0.16$
Perceived post trustworthiness	297	3.84	[3.69 3.99]	289	3.63	[3.48 3.78] $p = 0.05$
Perceived influencer status	297	4.33	[4.20 4.45]	289	4.37	[4.24 4.50] $p = 0.66$

The items for advertising awareness and perceived sponsorship transparency are presented after other items to ensure that these questions do not alert the subjects to the advertising nature of the post and thus affect their evaluations. Furthermore, the item for advertising awareness is presented before the items for perceived sponsorship transparency as this prevents the possibility that the questions related to perceived sponsorship transparency alert the subjects to the advertising nature of the post, and thus affect their true level of advertising awareness. Moreover, we inform the subjects beforehand that they cannot go back to the previous pages during the experiment.

We use a between-subjects design and randomly allocate subjects to one (and only one) treatment group. We recruit 290 subjects for the “No disclosure” condition and 298 subjects for the “Header disclosure” condition from Prolific, an online survey platform. The collected responses pass the manipulation and randomization checks (detailed results are available from the authors upon request).

Table 4 presents the statistics of all of the measured variables across conditions. Figure 5a shows the means of advertising awareness, perceived sponsorship transparency, and willingness to engage for the two conditions of Study 1. Header disclosure increases both advertising awareness and perceived sponsorship transparency. We check whether the

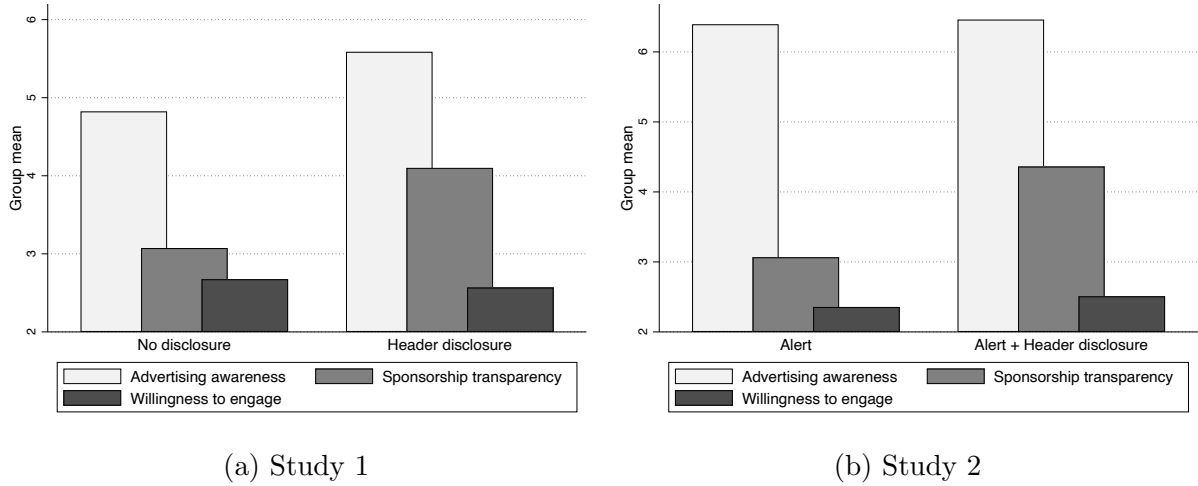


Figure 5: Group means by disclosure

differences between the two conditions are significant using one-way analysis of variance (ANOVA). The results confirm that advertising awareness is significantly higher in the “Header disclosure” condition than in the “No disclosure” condition ($M = 5.59$ vs. 4.82 ; $F(1, 586) = 36.66$, $p < 0.01$). The same holds for perceived sponsorship transparency ($M = 4.10$ vs. 3.07 ; $F(1, 586) = 69.11$, $p < 0.01$). Figure 5a also shows that willingness to engage is relatively similar in the two conditions. We confirm this with the one-way ANOVA—there is no significant difference between the two conditions ($M = 2.57$ vs. 2.67 ; $F(1, 586) = 0.89$, $p = 0.35$) regarding willingness to engage. The lack of a significant main effect on willingness to engage could be because there are multiple parallel mechanisms operating in opposite directions, canceling each other out (Hayes 2009; Lei et al. 2021).

We formally test this by conducting a mediation analysis using the PROCESS model of Hayes (2013). PROCESS is a bootstrapping method that does not require the normality assumption and can test parallel mediation effects in a between-subjects design. The results show that header disclosure has a significant and positive effect on advertising awareness ($\beta = 0.76$, $t(586) = 6.05$, $p < 0.01$) and perceived sponsorship transparency ($\beta = 1.03$, $t(586) = 8.31$, $p < 0.01$). Moreover, advertising awareness has a significant negative effect on willingness to engage ($\beta = -0.23$, $t(584) = -6.09$, $p < 0.01$), and

perceived sponsorship transparency has a significant positive effect on willingness to engage ($\beta = 0.18$, $t(584) = 4.65$, $p < 0.01$) (see Figure 6a). The indirect effect through advertising awareness is significant and negative (a * b path coefficient = -0.17 , 95% CI = $[-0.26, -0.10]$) and the indirect effect through perceived sponsorship transparency is significant and *positive* (a * b path coefficient = 0.18 , 95% CI = $[0.10, 0.28]$). These results fully support our theoretical model in Figure 3.

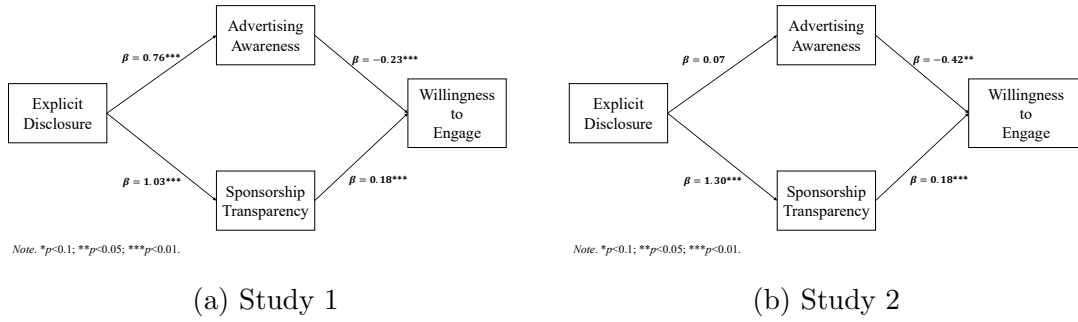


Figure 6: Estimated path coefficients for willingness to engage

4.3 Study 2

The results of Study 1 show that header disclosure significantly increases subjects' advertising awareness, which significantly decreases their willingness to engage, perceived post trustworthiness, and perceived influencer status. Recall that our main contribution in this study is the finding of the positive effect of header disclosure mediated by enhanced sponsorship transparency. In Study 2, we add an extra manipulation in the experiment to inhibit the negative mediation effect, which is expected to have little influence on the positive mediation effect. We expect to only detect a significant and positive mediation effect through perceived sponsorship transparency.

Study 2 also has two conditions. The stimuli remain the same (see Figures 4a and 4b). In the first condition, subjects see the undisclosed version, and in the second condition, subjects see the header-disclosed version of the post. The procedure in Study 2 is identical to that in Study 1 with one additional manipulation. The subjects are shown the following

message after seeing the stimulus and before answering the questions (i.e., between Step 2 and Step 3 of the Study 1 procedure)—“*Please note that the post you have just seen is a sponsored advertisement.*” This message is intended to alert all of the subjects in both conditions to the advertising nature of the post before they respond to the measurements, which is expected to reduce the negative effect of disclosure on engagement mediated by advertising awareness.

The condition with the undisclosed post is called “Alert” and the condition with header-disclosed post is called “Alert + Header disclosure.” We recruited 289 subjects for the “Alert” condition and 297 subjects for the “Alert + Header disclosure” condition from Prolific. The collected responses pass the manipulation and randomization checks.

Table 4 presents the statistics of all the measured variables for both conditions. Figure 5b shows the means of advertising awareness, perceived sponsorship transparency, and willingness to engage for both conditions of Study 2. Advertising awareness is at a similarly high level in both conditions, while perceived sponsorship transparency is much higher with header disclosure than with no disclosure and willingness to engage is slightly higher with header disclosure than with no disclosure.

The results based on the PROCESS method show that header disclosure does not significantly affect advertising awareness but still significantly increases perceived sponsorship transparency (see Figure 6b). The indirect effect through advertising awareness is *not* significant ($a * b$ path coefficient = -0.03, 95% CI = [-0.10, 0.04]) and the indirect effect through perceived sponsorship transparency remains significant and *positive* ($a * b$ path coefficient = 0.24, 95% CI = [0.15, 0.33]). These results are consistent with our *a priori* expectation. Furthermore, the ANOVA results show that the main effect of header disclosure on willingness to engage ($M = 2.51$ vs. 2.36 , $F(1, 584) = 1.99$, $p = 0.16$) is

positive and closer to statistical significance in Study 2, which are in stark contrast to the pattern in Study 1.

Taken together, these results provide some hints about why the positive header-disclosure effect is more salient in our field data than in the online experiments. In real life, influencers' sponsored content is mostly, if not all, exposed to their followers, who are familiar with the influencers. In contrast, in our experiments, most subjects are not familiar with the influencers featured in the stimuli (based on the subjects' self-reported familiarity with the influencer, most of them are not familiar with the influencer—the mean value of their responses is 2.24 on a 7-point Likert scale and 77% of them respond with 1 or 2). Familiarity with the influencers may enable followers to more easily recognize the advertising content, thus largely inhibiting the negative effect mediated by advertising awareness. Moreover, followers might interpret the explicit disclosure as a salient signal of the influencers' status or reputation more positively when they are familiar with the influencer than when they are not familiar. Therefore, we observe a more pronounced positive effect of header disclosure on user engagement in the field data than in the online experiments.

5 Discussion and Conclusion

Using large-scale field data from two major social media platforms (Facebook and Instagram), we empirically investigate how explicitly disclosing sponsored content from social media influencers affects users' engagement. To our knowledge, our research provides the first empirical evidence based on field data showing that explicitly disclosing sponsorship can significantly *increase* user engagement, especially in terms of the number of users' likes. We also provide evidence that this effect is driven by the increased information transparency about the sponsorship from randomized online experiments.

Our findings have important implications for regulators of social media influencer marketing or native advertising. Regulators (e.g., the FTC in the US) seem to focus their

policy-making and enforcement on the premise that consumers are at risk of being deceived and thus may wrongly consider sponsored content as organic if the sponsorship is not prominently and explicitly disclosed (Federal Trade Commission 2018, 2020). Our results show that in addition to raising users’ advertising awareness, being transparent can significantly *increase* the level of user engagement. This gives a novel perspective to regulators on better enforcing the regulation on prominent disclosures—instead of only emphasizing the unlawfulness of hiding advertising purposes, regulators could also highlight the positive consequences of explicit disclosures on user engagement. For example, in the letters sent out to influencers who do not prominently disclose sponsorship in their content⁴, the FTC can include insights from this study to educate influencers that being candid about sponsorship could potentially earn them more user engagement than being evasive about it.

The implications of our findings for marketing firms and influencers in native advertising are clear and important. They should avoid engaging in deceptive advertising and hiding the commercial nature of their ads because sponsorship transparency in sponsored content can earn user goodwill and higher levels of engagement. Moreover, adequate disclosure of sponsorship helps avoid potential legal sanctions. It is possible that sponsoring firms and influencers are not incentive-aligned in explicitly disclosing sponsorship. To avoid insufficient disclosure due to influencers’ carelessness or other ulterior motives, firms can mandate sponsorship to be disclosed in a sufficiently clear and explicit manner while developing contracts with influencers to push them to ensure sufficient disclosure diligently.⁵

Our findings also have novel implications for advertising platforms. It is beneficial for the platform to encourage or even mandate advertisers and influencers to explicitly disclose

⁴For sample warning letters sent by the FTC to influencers, see <https://www.ftc.gov/system/files/documents/cases/1823174teamiwarningletters.pdf> (accessed Feb 13, 2023))

⁵We thank one anonymous reviewer for this suggestion.

sponsored content posted on their platform. Following the example of Facebook and Instagram, the platforms could offer tools to facilitate prominent disclosures of sponsorship. The platform can further design features or functions to nudge content contributors to ensure prominent disclosures. For example, the platform could utilize artificial intelligence (AI) technologies (e.g., auto labeling of image or video) to detect potentially sponsored content and then send messages to the content contributors to require them to disclose the sponsorship sufficiently when they try to upload content to the platform. This would help the platform to better monitor and monetize the influencers' marketing activities on the platform and users' goodwill from the enhanced information transparency. The resulting increase in user engagement due to enhanced information transparency is ultimately good for the growth of the platform. Moreover, our empirical results show that a more salient disclosure format (i.e., header-disclosure) works better than a less salient one (i.e., hashtag disclosure) (see results in column (2) of Table 3 on page 18). This means when designing the disclosure format, the platform is advised to make it salient to help users quickly comprehend the advertising purposes embedded in the content.

References

- Abendroth, Lisa J, James E Heyman. 2013. Honesty is the best policy: The effects of disclosure in word-of-mouth marketing. *Journal of Marketing Communications* **19**(4) 245–257.
- Boerman, Sophie C, Lotte M Willemsen, Eva P Van Der Aa. 2017. "this post is sponsored": Effects of sponsorship disclosure on persuasion knowledge and electronic word of mouth in the context of facebook. *Journal of Interactive Marketing* **38** 82–92.
- Calvo, Eduard, Ruomeng Cui, Laura Wagner. 2020. Disclosing product availability in online retail. *Manufacturing & Service Operations Management* .
- Campbell, Colin, Pamela E Grimm. 2019. The challenges native advertising poses: Exploring potential federal trade commission responses and identifying research needs. *Journal of Public Policy & Marketing* **38**(1) 110–123.
- Campbell, Margaret C, Gina S Mohr, Peeter WJ Verlegh. 2013. Can disclosures lead consumers to resist covert persuasion? the important roles of disclosure timing and type of response. *Journal of Consumer Psychology* **23**(4) 483–495.
- Carl, Walter J. 2008. The role of disclosure in organized word-of-mouth marketing programs. *Journal of Marketing Communications* **14**(3) 225–241.

- Chevalier, Judith A., Dina Mayzlin. 2006. The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research* **43**(3) 345–354.
- Darke, Peter R, Robin JB Ritchie. 2007. The defensive consumer: Advertising deception, defensive processing, and distrust. *Journal of Marketing research* **44**(1) 114–127.
- Dubois, David, Derek D. Rucker, Adam D. Galinsky. 2012. Super size me: Product size as a signal of status. *Journal of Consumer Research* **38**(6) 1047–1062.
- eMarketer. 2019. Instagram leads as a global platform for influencer marketing. *eMarketer.com* Available at <https://www.emarketer.com/content/instagram-is-the-leading-platform-for-influencer-marketing> (Retrieved April 8, 2021).
- eMarketer. 2022. Influencer marketing sees steady adoption and room for growth. *eMarketer.com* Available at <https://www.emarketer.com/content/influencer-marketing-growth> (Retrieved June 24, 2022).
- Ershov, Daniel, Matthew Mitchell. 2020. The effects of influencer advertising disclosure regulations: Evidence from instagram. *Proceedings of the 21st ACM Conference on Economics and Computation (EC '20)*. ACM, 73–74. doi:10.1145/3391403.3399477.
- Evans, Nathaniel J, Bartosz W Wojdyski, Mariea Grubbs Hoy. 2019. How sponsorship transparency mitigates negative effects of advertising recognition. *International Journal of Advertising* **38**(3) 364–382.
- Facebook. 2016. Updates to our branded content policy. *Facebook for Media* Available at <https://www.facebook.com/facebookmedia/blog/updates-to-our-branded-content-policy> (Retrieved April 8, 2021).
- Federal Trade Commission. 2017. FTC staff reminds influencers and brands to clearly disclose relationship. *FTC.gov* Available at <https://www.ftc.gov/news-events/press-releases/2017/04/ftc-staff-reminds-influencers-brands-clearly-disclose> (Retrieved April 8, 2021).
- Federal Trade Commission. 2018. The FTC’s endorsement guides: What people are asking. *FTC.gov* Available at <https://www.ftc.gov/tips-advice/business-center/guidance/ftcs-endorsement-guides-what-people-are-asking> (Retrieved April 8, 2021).
- Federal Trade Commission. 2020. Tea marketer misled consumers, didn’t adequately disclose payments to well-known influencers, ftc alleges. *FTC.gov* Available at <https://www.ftc.gov/news-events/news/press-releases/2020/03/tea-marketer-misled-consumers-didnt-adequately-disclose-payments-well-known-influencers-ftc-alleges> (Retrieved June 23, 2022).
- Forehand, Mark R, Sonya Grier. 2003. When is honesty the best policy? the effect of stated company intent on consumer skepticism. *Journal of consumer psychology* **13**(3) 349–356.
- Granados, Nelson, Alok Gupta, Robert J Kauffman. 2010. Research commentary—information transparency in business-to-consumer markets: concepts, framework, and research agenda. *Information Systems Research* **21**(2) 207–226.
- Hayes, Andrew F. 2009. Beyond baron and kenny: Statistical mediation analysis in the new millennium. *Communication monographs* **76**(4) 408–420.
- Hayes, Andrew F. 2013. *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford publications.
- Isaac, Mathew S, Kent Grayson. 2017. Beyond skepticism: Can accessing persuasion knowledge bolster credibility? *Journal of Consumer Research* **43**(6) 895–912.

- Lei, Zhanfei, Dezhi Yin, Han Zhang. 2021. Focus within or on others: the impact of reviewers’ attentional focus on review helpfulness. *Information Systems Research* **32**(3) 801–819.
- Li, Xitong, Lynn Wu. 2018. Herding and social media word-of-mouth: Evidence from group. *MIS Quarterly* **42**(4) 1331–1351.
- Luong, Ashley. 2019. All that glitters is gold: the regulation of hidden advertisements and undisclosed sponsorships in the world of beauty social media influencers. *William & Mary Business Law Review* **11** 565–607.
- Mathur, Arunesh, Arvind Narayanan, Marshini Chetty. 2018. An empirical study of affiliate marketing disclosures on YouTube and Pinterest. *Workshop on Technology and Consumer Protection (ConPro)*.
- Mayzlin, Dina, Yaniv Dover, Judith Chevalier. 2014. Promotional reviews: An empirical investigation of online review manipulation. *American Economic Review* **104**(8) 2421–2455.
- Mitchell, Matthew. 2021. Free ad(vice): internet influencers and disclosure regulation. *RAND Journal of Economics* **52**(1) 3–21.
- Ohanian, Roobina. 1990. Construction and validation of a scale to measure celebrity endorsers’ perceived expertise, trustworthiness, and attractiveness. *Journal of Advertising* **19**(3) 39–52.
- Sahni, Navdeep S, Harikesh S Nair. 2020a. Does advertising serve as a signal? evidence from a field experiment in mobile search. *The Review of Economic Studies* **87**(3) 1529–1564.
- Sahni, Navdeep S, Harikesh S Nair. 2020b. Sponsorship disclosure and consumer deception: Experimental evidence from native advertising in mobile search. *Marketing Science* **39**(1).
- Sharethrough. 2015. Influencer marketing sees steady adoption and room for growth. *Sharethrough.com* Available at <https://www.sharethrough.com/blog/ad-effectiveness-study-native-ads-vs-banner-ads> (Retrieved June 24, 2022).
- Wang, Weiquan, May Wang. 2019. Effects of sponsorship disclosure on perceived integrity of biased recommendation agents: Psychological contract violation and knowledge-based trust perspectives. *Information Systems Research* **30**(2) 507–522.
- Wang, Weiquan, Jingjun Xu, May Wang. 2018. Effects of recommendation neutrality and sponsorship disclosure on trust vs. distrust in online recommendation agents: Moderating role of explanations for organic recommendations. *Management Science* **64**(11) 5198–5219.
- Wojdyski, Bartosz W, Nathaniel J Evans. 2016. Going native: Effects of disclosure position and language on the recognition and evaluation of online native advertising. *Journal of Advertising* **45**(2) 157–168.
- Zhou, Liying, Weiquan Wang, Jingjun David Xu, Tao Liu, Jibao Gu. 2018. Perceived information transparency in b2c e-commerce: An empirical investigation. *Information & Management* **55**(7) 912–927.

Appendix A. Additional Results Based on Field Data

Table A1 follows the same structure as Table 2 but measures user engagement with users’ comments. The effects follow the same direction as those on users’ likes, although

less significant. The estimate from our preferred specification in Column (6) indicates that header-disclosing a post leads to an increase of approximately 14% ($p < 0.1$) in users' comments.

Table A1: Effects of explicit disclosure on user engagement: Comments

VARIABLES	Log(Comments)						
	(1) All	(2) All	(3) All	(4) Spons.	(5) Matched	(6) Matched & Spons.	(7) Matched & Spons.
Header disclosed	0.558*** (0.135)	-0.0642 (0.0416)	0.0515 (0.0639)	0.0659 (0.0596)	0.167** (0.0688)	0.135* (0.0697)	
Hashtag disclosed	0.172* (0.0976)	-0.0614* (0.0357)					
Sponsored	-0.146*** (0.0376)	-0.129*** (0.0103)					
Log(words)	0.120*** (0.0302)	0.0595*** (0.0108)					
Log(mentions)	-0.433*** (0.0528)	-0.167*** (0.0152)					
Log(hashtags)	-0.229*** (0.0490)	-0.0261** (0.0130)					
Instagram	2.195*** (0.0632)						
Header disclosed (FB)							0.318* (0.183)
Header disclosed (IG)							0.113 (0.0768)
Platform-specific influencer FE	No	Yes	Yes	Yes	Yes	Yes	Yes
Post FE	No	No	Yes	Yes	Yes	Yes	Yes
Observations	810,168	810,168	810,168	147,073	135,592	51,620	51,620
R-squared	0.261	0.661	0.995	0.990	0.964	0.964	0.964

Robust standard errors clustered by influencers are in parentheses.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table A2 follows the same structure as Table 3 but measures user engagement with users' comments. The effects on users' comments are always positive, but not always significant.

Table A2: Sensitivity to composition of control group: Comments

VARIABLES	Log(Likes)						
	(1) Control: Spons.	(2) Control: Hashtag-disclosed	(3) Control: With logos	(4) Control: @business	(5) Control: Hashtag-disclosed + with logos:	(6) Control: Hashtag-disclosed + @business	(7) Control: With logos + @business
Header disclosed	0.135* (0.0697)	0.178* (0.0958)	0.160** (0.0696)	0.117 (0.0747)	0.162** (0.0654)	0.118 (0.0747)	0.134* (0.0697)
Platform-specific influencer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Post FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	51,620	834	16,748	39,880	17,218	40,124	51,410
R-squared	0.964	0.983	0.968	0.962	0.968	0.962	0.964

Robust standard errors clustered by influencers are in parentheses.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

We conduct several additional robustness checks of our preferred specification. First, we focus on the subsample consisting of matched sponsored posts posted by influencers

who are equally popular, with their ratios of the number of followers on the two platforms is between the 25th and 75th sample percentiles. As reported in columns (1) and (4) of Table A3, the estimates of the disclosure effects remain significantly positive.

Second, we focus on the matched sponsored posts that were posted on the two platforms within the median posting time difference (which is approximately 2 hours). As reported in Columns (2) and (5) of Table A3, the estimates of the disclosure effects remain consistently positive.

Third, we use the median number of users' likes of all the posts of an influencer on a platform in a given month as a proxy for influencers' follower count on the platform in that month. Note that the proxy measure is calculated from the full set of 810,168 posts. The results are reported in Columns (3) and (6) of Table A3. Reassuringly, the disclosure effects are still positive.

Table A3: Additional robustness checks

VARIABLES	Log(Likes)		Log(Comments)			
	(1)	(2)	(3)	(4)	(5)	(6)
	Similar influencer popularity across platforms	Similar posting time across platforms	Control for tends of followers' size	Similar influencer popularity across platforms	Similar posting time across platforms	Control for trends of followers' size
Header disclosed	0.370*** (0.0654)	0.234*** (0.0776)	0.184*** (0.0602)	0.215*** (0.0806)	0.139 (0.0914)	0.0602 (0.0696)
Monthly median user likes			0.418*** (0.0473)			0.282*** (0.0321)
Platform-specific influencer FE	Yes	Yes	Yes	Yes	Yes	Yes
Post FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	25,932	32,664	51,620	25,932	32,664	51,620
R-squared	0.974	0.980	0.982	0.961	0.966	0.966

Robust standard errors clustered by influencers are in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

Appendix B. Additional Experiment Results

As discussed in the main text, we replicate the experiment in Study 1 using another influencer post as the stimulus to increase the generalizability of our findings. Figures B1a and B1b show the undisclosed and disclosed versions of the influencer post, respectively. Note that in this influencer post, the picture embeds an explicit product logo.



(a) Undisclosed version



(b) Header-disclosed version

Figure B1: The influencer post used as experiment stimulus in Study 1 replication

We recruited 284 subjects for the “No disclosure” condition and 292 subjects for the “Header disclosure” condition from Prolific. The collected responses pass the manipulation and randomization checks. The results from the analyses based on the PROCESS method are as follows, which are consistent with the results reported in the main text. With willingness to engage as the outcome variable, the estimated path coefficients are shown in Figure B2. The indirect effect through advertising awareness is significant and negative ($a * b$ path coefficient = -0.29, 95% CI = [-0.39, -0.19]) and the indirect effect through perceived sponsorship transparency is significant and *positive* ($a * b$ path coefficient = 0.25, 95% CI = [0.15, 0.37]).

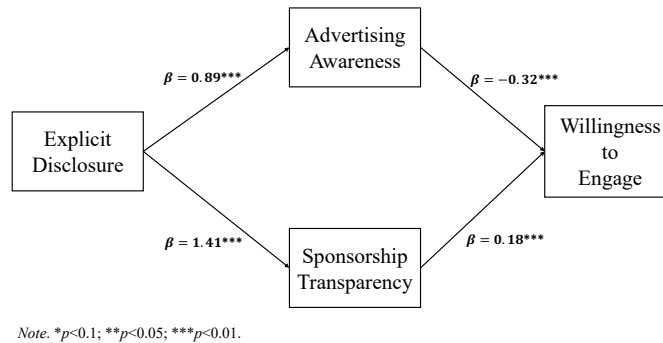


Figure B2: Estimated path coefficients for willingness to engage in Study 1 replication