f i ® s t m ¤ ñ d @ ¥
PEER-REVIEWED JOURNAL ON THE INTERNET

The impact of engagement metrics overwhelms the influence of online disinhibition

by Ruohan Wen and Asako Miura

Abstract

Clicking on the Like and Dislike buttons on social media has become an important channel for social engagement. This study examined the influence of online disinhibition and engagement metrics on Internet users' attitude expressions via clicking Like and Dislike. Initially, a preliminary study was conducted, wherein participants engaged with a fictitious thread and clicked Like or Dislike on each post. We recorded the Likes and Dislikes each post received, which served as engagement metrics. In Study 1, an updated thread was introduced, displaying these metrics alongside the Like and Dislike buttons. The results indicated that engagement metrics had a significant impact on attitude expressions, while online disinhibition had no significant direct or moderating effects. Study 2 explored scenarios with these metrics reversed, again finding that the influence of online disinhibition remained insignificant. These findings underscore the importance of considering situational factors when applying the online disinhibition effect to understand online behavior.

Contents

Introduction

Attitude expression in social media

Social media have become important platforms for social engagement in modern society. They not only present a wide range of opinions and discussions but also offer users easy ways to participate. Many platforms provide one-click interaction features, such as the Like button, which signals a positive response, and the Share button, which allows users to share content on their own pages. Hayes, *et al.* (2016) called these buttons paralinguistic digital affordances (PDAs), facilitating online communication and interaction without language.

This study focuses on PDAs, particularly the Like button. As a widely used feature, the Like button enables

users to interact with specific content conveniently and with minimal effort. Although previous studies have shown diverse motivations for clicking Like and its meaning (*e.g.*, Chin, *et al.*, 2015; Hayes, *et al.*, 2016; Sumner, *et al.*, 2018), we can generally consider a Like as indicating a positive attitude toward certain content.

However, when only a Like button is available, users' attitudes may be ambiguous. Not-clicking-Like might suggest a lack of interest or a lack of a distinct stance toward certain content. It could also indicate that the user holds a negative viewpoint, such as disagreement or hostility. To reduce this uncertainty in users' not-clicking-Like behavior, our study additionally considers situations that include the Dislike button. This button provides users with a way to express a distinctly negative attitude. The combination of both Like and Dislike buttons enables a clearer capture of participants' attitude expressions.

This study aims to explore the situational and individual factors that influence general Internet users to click Like or Dislike to express their attitude when encountering content on social media. As a situational factor, the study focuses on the display of the number of Likes and Dislikes — defined here as engagement metrics — and investigates how they shape individuals' attitude expression. In addition to using engagement metrics generated in real scenarios, we also investigate a condition with artificially modified metrics, in which the original numbers of Likes and Dislikes are reversed. As an individual factor, we consider users' characteristics of online disinhibition (Suler, 2004). To empirically examine the effects of engagement metrics and online disinhibition, we conducted three experiments. The results are exploratory in nature and offer initial insights into the mechanisms underlying attitude expression on social media.

Online disinhibition theory

People sometimes experience a more relaxed atmosphere on the Internet, feeling free from inhibitions of real life, which may lead them to say or do things they would not typically say or do offline. Suler (2004) termed this phenomenon the "online disinhibition effect." Stuart and Scott (2021) pointed out that online disinhibition is a mental state where Internet users experience reduced control over their behavior. Wen and Miura (2023) further explored this state in detail, arguing that individuals' cognitions regarding social phenomena — such as identity, social desirability, and social rules — constrain their behavior. However, in the online environment, due to characteristics like anonymity and invisibility, these cognitions change, reducing constraints on behavior. According to Wen and Miura (2023), online disinhibition is "a mental state where an individual's inhibition over their behavior weakens or disappears in an online environment."

One notable effect of online disinhibition is its role in facilitating the expression of hidden attitudes or behaviors, whether friendly or aggressive. Suler (2004) suggested that online disinhibition contributes to greater openness among individuals due to less restraint on behavior. Empirical research by Schouten, *et al.* (2007) found that higher online disinhibition encouraged greater online self-disclosure. Conversely, research has established a link between online disinhibition and aggressive behaviors, such as online trolling and cyberbullying (*e.g.*, Kurek, *et al.*, 2019; Udris, 2014; Stuart and Scott, 2021; Wright, *et al.*, 2019; Wright and Wachs, 2021). Based on these findings, we infer that higher online disinhibition might increase users' likelihood of clicking the Like and Dislike buttons to express their attitudes.

To capture the effect of online disinhibition more precisely, this study adopts a theoretical framework based on the motivation-based online disinhibition (MOD) model (Wen and Miura, 2023). The MOD model argues that individuals' various motivations determine their specific behaviors, but whether these behaviors are actually expressed in online environments is moderated by online disinhibition. For example, when individuals see a topic of interest on social media, they may feel motivated to join the discussion. However, such motivation does not always lead to actual participation. Individuals with low levels of online disinhibition tend to be more concerned about their actions. For example, they may give more thought to whether the discussion aligns with social norms or whether clicking Like on inappropriate content could cause trouble. These concerns can inhibit the translation of motivation into real action. In contrast, individuals with higher levels of online disinhibition experience fewer such concerns and are therefore more likely to act on their motivations.

The MOD model provides a more comprehensive insight into Internet users' behavior. This study aims to examine how online disinhibition influences attitude expression according to this model. We used an external stimulus to create conditions with stronger motivations for attitude expression and examined whether online disinhibition moderates the translation from motivation into behavior.

The current study

Examination of the engagement metrics and online disinhibition

This study examines how engagement metrics influence individuals' motivations for attitude expression. Many social media platforms display the number of Likes and Dislikes specific content has received. In public discussions, engagement metrics serve as cues to other users' attitudes, exerting a strong social influence that shapes subsequent interactions.

Empirical research supports this effect. Muchnik, *et al.* (2013) demonstrated that posts on Reddit receiving pre-assigned positive feedback were more likely to gain additional positive feedback. Matsui (2021) found that content with higher engagement metrics elicited more user interaction than content with lower metrics. Building on these findings, we hypothesize that users are more motivated to express their attitudes by clicking Like or Dislike when engagement metrics are displayed. Accordingly, we propose the following hypothesis:

> *H1*: Individuals have a stronger motivation to express their attitudes when reading a discussion with displayed engagement metrics.

Next, we examined the effect of online disinhibition. According to the MOD model, online disinhibition moderates the translation from motivation into behavior. Specifically, even if individuals develop a strong motivation for attitude expression in discussions where engagement metrics are displayed, those with lower levels of online disinhibition still tend to regulate their actions. They may hesitate to participate in radical or unfriendly discussions that deviate from social desirability; they may also worry that their expressed attitudes could violate social norms. Such concern leads to a lower likelihood of taking action. In contrast, individuals with higher levels of online disinhibition are more likely to express their attitudes due to reduced cognitive constraints. Based on this, we propose the following hypothesis:

> *H2*: Online disinhibition moderates the impact of engagement metric displays.

To test these hypotheses, we designed an experiment simulating the experience of reading a thread on anonymous bulletin boards. The thread featured an initial post and multiple replies, with participants able to click Like or Dislike buttons on each reply. We operationally defined *attitude expression* as the total number of Likes and Dislikes clicked by participants while reading a thread.

We conducted two separate surveys. In the preliminary study, participants read a version of the thread where replies featured only Like and Dislike buttons, but no engagement metrics were displayed. We recorded the number of Likes and Dislikes each reply received and used these as engagement metrics in the subsequent study. Additionally, attitude expression in the preliminary study served as the baseline for comparison.

In Study 1, participants viewed an updated version of the thread where engagement metrics were displayed next to the Like and Dislike buttons. We tested the hypotheses by investigating the effects of engagement metrics and online disinhibition across these two situations.

Exploration of majority and minority attitude expressions

Since the engagement metrics of Likes and Dislikes were visible on each post, participants could determine whether their attitude (Like or Dislike) toward each post aligned with the majority or minority positions. This study explored the effect of online disinhibition on the expression of majority and minority attitudes.

Traditional social psychological theories suggest that the presence of engagement metrics increases motivation to align with majority opinions. For example, conformity theory (Deutsch and Gerard, 1955) argues that majority positions exert both normative and informational influences, encouraging individuals to conform. Similarly, social impact theory (Latané, 1981) suggests that positions with a larger number of supporters have stronger social influence. These insights imply a higher likelihood of majority attitude expression when engagement metrics are displayed in a thread.

Nevertheless, minority positions do not necessarily disappear. Online communities are extremely diverse in various stances and attitudes (Hogg and Szabo, 2009). Even those holding minority views can find a space on social media. According to online disinhibition theory (Suler, 2004), individuals — including those in the minority — may feel empowered to express their views without typical social pressures when experiencing disinhibition. This effect may facilitate the expression of minority attitudes.

Given the complexity of users' motivations, this study explored whether online disinhibition correlates with majority and minority attitude expressions. We operationally defined majority (minority) attitude expressions as the number of attitude expressions aligning with the majority (minority) position. This led to the following research question:

> *RQ1*: What is the relationship between online disinhibition and majority (minority) attitude expressions when engagement metrics are displayed in a thread?

Exploration of the impact of reversed engagement metrics

Additionally, we further explored the impact of the values of engagement metrics on attitude expression. In Study 1, the displayed engagement metrics were derived from general Internet users. As a result, the majority of participants in that study may have perceived these metrics as similar to their own values or preferences when reading posts. This raises a key question: do engagement metrics influence behavior simply because they are present, or because they reflect the majority's values? To address this, Study 2 conducted a new survey by adapting the thread from Study 1 with one significant modification: the numbers beside each Like and Dislike button were reversed. Namely, the number next to the Like button indicated the count of Dislikes from the preliminary study, and conversely for the Dislike button. Under this setup, we investigate the following research question:

> *RQ2*: Does the impact remain consistent when the displayed engagement metrics are reversed?

Exploration of the spiral of silence phenomenon

Finally, we explored how support for majority and minority positions changed in Studies 1 and 2. The spiral of silence theory (Noelle–Neumann, 1974) posits that among two opposing positions, individuals who perceive their stance as aligned with the majority are more likely to express themselves openly, while those who perceive their position as unpopular tend to remain silent. This dynamic can widen the gap between majority and minority opinions. Based on this theory, we propose the following research question:

> *RQ3*: Compared with the preliminary study, does the spiral of silence phenomenon occur in Studies 1 and 2?

Preliminary study

Methods and participants

To expose participants to a disinhibitory context and provide them with varied content, we created experimental materials based on a highly topical and controversial political discussion. In March 2023, the Japanese government introduced a policy to cull cows due to milk overproduction. Around the same time, news about a high school using edible cricket powder in school meals sparked significant controversy on the Japanese Internet. As eating crickets or insects is generally considered unacceptable by most Japanese people, the incident quickly triggered strong dissatisfaction with the government. On the anonymous Japanese forum *5ch*, many users used this meme to criticize the government and politicians.

We constructed the experimental thread using content drawn from actual discussions on *5ch*. In the preliminary study, we created a thread consisting of one initial post introducing the topic and 24 replies representing a range of opinions. The initial post introduced the cow-culling policy in Japan. Some replies focused on the initial topic, discussing the policy's merits and demerits. Others referenced this cricket-eating meme to satirize the government for promoting unusual foods while reducing milk production.

We used the "hotspot" function in Qualtrics to create more authentic experiences. We drew long images of the thread, with each reply post containing interactive Like and Dislike buttons. Participants could click on the thumb icons to activate the hotspots to express Likes or Dislikes on the post. Additionally, the participants could click on any activated hotspot again to withdraw Likes or Dislikes.

To prevent the problem of satisficing on online surveys (Krosnick, 1991; Miura and Kobayashi, 2019), where participants might quickly skip the process without reading the post in detail, we added a Done button to each post. When participants had responded with a Like or Dislike, clicking the Done button was not required. However, if they had not responded with a Like or Dislike, they were required to click on the Done button to indicate they had completed reading the post. Clicking multiple buttons (*e.g.*, Like + Dislike or Like + Done) on the same post would result in an error. Consequently, the number of attitude expressions per participant across the 24 posts could range from 0 to 24. In subsequent studies, we adopted a similar approach, with only minor adjustments made to the visual effects. Figure 1 illustrates the posts used in the present study; complete materials are available at https://osf.io/6um3y/?view_only=80f1ef703bc94039bf730999531b609c.

Figure 1: Samples of posts in the present study.

InfoPlatforms usage rate. The participants were required to answer one filler question about using online bulletin boards or news sites, with response options ranging from 1 to 10. Hereinafter, it will be referred to as the InfoPlatforms usage rate.

Introduction of cricket-eating event and reading thread. We quoted a part of a news article from the *Nikkei* (2022) to introduce the cricket-eating event. Subsequently, we informed the participants that they would then be required to read a thread. We provided the following instructions to participants: “There are Like and Dislike buttons below each post. Feel free to click them if you wish to. However, if you choose not to, click the Done button to signify that you have read this post. Each post requires one of the Like, Dislike, or Done buttons to be clicked.” We thoroughly explained the following situations that might cause errors: clicking Like + Dislike, Like + Done, Dislike + Done, Like + Dislike + Done, and no button clicks. To familiarize them with the process, participants were given two practice opportunities: one to click the Like or Dislike button, and another to click Done. The participants then read the fictitious thread we had created.

Direct question scale (DQS) (Maniaci and Rogge, 2014). After reading the thread, we asked participants why they clicked Like, Dislike, or only clicked Done. Among these items, a DQS item, which directly required participants to select ”disagree” was included, and the respondents who answered this item incorrectly were excluded from the analysis.

Japanese version of ten item personality inventory (Oshio, *et al.*, 2012). To explore the relationship with personality, we measured the Big Five using the Japanese version of 10-item personality inventory, on a seven-point scale ranging from “completely disagree” to “strongly agree.” The results are beyond the scope of this study and are not discussed further here.

Multidimensional measure of online disinhibition (Wen and Miura, 2024). We employed the multidimensional measure of online disinhibition (MMOD) to assess participants' online disinhibition levels. MMOD measures online disinhibition from three factors: unique perspective on online environment (five items), meaning that percept online environment as a special world differs from the real world; change of alienation cognition (four items), meaning reduced social distance; and change of relationship cognition (three items), meaning reduced of interpersonal constrain. Participants were asked to answer 12 items with response options ranging from one (strongly disagree) to six (strongly agree), including an "I don't know" option (which was coded as a missing value).

Demographic questions. Finally, the participants were required to answer demographic questions.

We commissioned the crowdsourcing company CrowdWorks Co., Ltd. to recruit general Japanese Internet users aged 18 to 70, without restricting gender or education level. Data were collected on 16 March 2023. We specified in advance that the survey would be conducted anonymously and would not involve any questions violating the participants't privacy. Next, we obtained the participants't consent before proceeding with the survey; the participants could stop or withdraw at any time during the survey. The survey lasted approximately 12 minutes, and each participant received 170 JP¥ as a reward for completing it.

Results

A total of 414 individuals completed the survey. We performed list-wise deletion of the data with incorrect answers to the DQS item and the data with missing values in the MMOD items and demographic questions. This process left us with 368 valid responses. The average age was 42.12 (SD = 10.12) years, and 53.0 percent were female. Furthermore, 53.7 percent of the participants held a bachelor's degree or higher.

We quantified the number of Likes and Dislikes each participant clicked, as well as the total number of attitude expressions. Initially, a descriptive statistical analysis was conducted. Table S1 presents the correlation coefficients and Cronbach's α for both attitude expressions and MMOD. MMOD exhibited significant correlations with medium effect sizes for attitude expressions ($r = 0.21$, $p < 0.01$) and clicking Like ($r = 0.25$, $p < 0.01$). The correlation between MMOD and clicking Dislike was not significant ($r = 0.09$, $p = 0.08$), which might be attributed to participants' lesser familiarity with the Dislike function. These findings suggest that individuals with higher levels of online disinhibition are more inclined to engage with interaction buttons than click Done.

Finally, we counted the number of Likes and Dislikes each reply post received from participants. On re-evaluation, we decided to exclude two posts featuring extreme personal attacks from subsequent studies. The Likes and Dislikes received by the 22 remaining posts are detailed in Table 1, serving as the engagement metrics for Studies 1 and 2. Additionally, the majority and minority positions are determined by the number size of Likes and Dislikes as shown in Table 1,. For a certain post, the position with the larger number is defined as the majority, while the position with the smaller number is defined as the minority.

Table 1: The engagement metrics of each post.

Post ID	Like	Dislike	Post ID	Like	Dislike
Post 2	23	237	Post 13	21	135
Post 3	94	57	Post 14	28	144
Post 4	37	156	Post 15	66	74
Post 5	37	143	Post 16	164	5

Post 6	254	12	Post 17	124	62
Post 7	216	27	Post 18	102	86
Post 8	38	181	Post 19	18	129
Post 9	51	57	Post 20	95	24
Post 10	87	71	Post 21	56	81
Post 11	195	18	Post 22	26	168
Post 12	190	20	Post 23	146	25

Study 1

Methods and participants

We added the engagement metrics next to the Like and Dislike buttons in each post (refer to Figure 1). The questionnaire was largely consistent with that used in the preliminary study, with some improvements. To assess the effect of online disinhibition on attitude expressions, participants answered the MMOD before reading the thread. Additionally, we posited that participants' attitudes toward cricket eating might influence their likelihood of attitude expression. Thus, we created four items (*e.g.*, "I am reluctant to eat crickets," which was reverse scored) to assess participants' attitudes toward cricket eating. Hereinafter, the average of these four items will be referred to as attitude toward cricket eating.

Although we incorporated these improvements, which might have had a slight influence on participants' behaviors, we believe that participants' attitude expressions were primarily determined by elements within the thread-reading scene itself (*i.e.*, the content and engagement metrics of the posts). Participants spent tens of seconds on an uninterrupted single page reading the thread and clicking buttons, which reduced the likelihood that their behaviors were significantly influenced by the nuanced improvements made during the phase preceding the thread-reading experiment. Given these considerations, we did not conduct a control experiment by randomly dividing participants into groups to compare responses with and without engagement metrics. Instead, we merged the data from the preliminary study and Study 1, treating the different studies as a between-participant factor for analysis. This approach allowed us to retain a larger dataset for analysis while acknowledging that there are some trade-offs in our methodology.

The approach to recruiting participants was identical to that in the preliminary study. Individuals who participated in the preliminary study were excluded. The survey was conducted on 17 June 2023, and lasted approximately 10.5 minutes. Each participant received 140 JP¥ as a reward for completing it.

Results

A total of 301 individuals completed the survey. Using the same data exclusion criteria as in the preliminary study, 282 valid responses remained. We counted the number of Likes and Dislikes each participant clicked, and performed a descriptive analysis, and Table 2 presents the correlation coefficients, descriptive statistics, and Cronbach's α.

Table 2: Results of the correlation analysis.

	Sum	Like	Dislike	MMOD	Cricket-eating attitude	Age	Gender
Like	.82**						
Dislike	.90**	.50**					
MMOD	.05	.15*	-.03				
Cricket-eating attitude	-.04	-.03	-.03	.01			
Age	.22**	.15*	.22**	-.17**	.09		
Gender	-.14*	-.10	-.13*	-.07	-.09	-.12	
Education	-.12	-.11	-.10	.06	.03	-.11	-.05
M	14.37	7.18	7.19	3.44	3.16	41.46	
SD	7.44	3.68	4.87	0.50	1.28	10.32	
α				0.64	0.86		

Note: Sum: number of attitude expressions, MMOD: multidimensional measure of online disinhibition, Gender (0 = male, 1 = female), Education (0 = Less than a bachelor's degree, 1 = bachelor's degree or higher). * $p < 0.05$, ** $p < 0.01$.

To test *H1* and *H2*, we combined the data from the preliminary study and Study 1 for analysis, with a total sample size of 650. Multiple regression analysis was conducted with attitude expressions serving as the dependent variables. Independent variables included the MMOD, engagement metrics, the interaction between MMOD and engagement metrics, InfoPlatforms usage rate, age, gender, and education levels, with the outcomes detailed in Table 3. The regression model was significant ($p < 0.01$) with an R^2 of 0.16. The coefficient for engagement metrics was significant ($p < 0.01$) with a large effect size. These findings indicate that participants in Study 1 were more inclined to click Like and Dislike to express their attitude than those in the preliminary study, thus supporting *H1*.

Table 3: Standardized coefficients of the regression model of attitude expressions.

	β	95% CI	Standard error	*p*-value
(Intercept)	-0.15	(-0.30, 0.00)	0.08	0.05
MMOD	0.16	(0.07, 0.26)	0.05	<0.01
Engagement metrics (= 1)	0.62	(0.48, 0.77)	0.07	<0.01
MMOD* Engagement		(-0.24,		

metrics	-0.10	0.05)	0.07	0.19
InfoPlatforms usage rate	0.15	(0.08, 0.22)	0.04	<0.01
Age	0.11	(0.03, 0.18)	0.04	<0.01
Gender (= 1)	-0.13	(-0.28, 0.01)	0.07	0.07
Education (= 1)	-0.08	(-0.24, 0.05)	0.07	0.27
Note: MMOD: multidimensional measure of online disinhibition, Engagement metrics (0 = Not displayed, 1 = Displayed), Gender (0 = male, 1 = female), Education (0 = Less than a bachelor's degree, 1 = bachelor's degree or higher).				

MMOD showed a significant ($p < 0.01$) but small effect, indicating that online disinhibition significantly promotes attitude expressions to some extent. However, the interaction effect between MMOD and engagement metrics was not significant ($p = 0.19$), and the negative coefficient even suggests a trend contrary to the MOD model's predictions. These findings suggest no significant difference in the effects of online disinhibition across threads with and without engagement metrics, thus not supporting *H2*. Compared to the model without the interaction effect, including the interaction effect resulted in a ΔR^2 of 0.002 and a corresponding f^2 of 0.002. Post hoc power analysis using G*Power 3.1.9.2 (Faul, *et al.*, 2009) revealed a low power of 0.21 to detect the interaction effect. Exploratory analyses using the three individual factors of MMOD instead of the total score yielded similar results (Table S2).

To investigate *RQ1*, we counted the number of attitude expressions and majority and minority attitude expressions and performed correlation analyses using MMOD, majority, and minority attitude expressions. The correlation coefficients between MMOD and the majority ($r = 0.6$, $p = 0.34$) and minority ($r = 0.2$, $p = 0.78$) are not significant. These suggest that higher levels of online disinhibition did not significantly influence the expressions of majority or minority opinions.

Discussion

In this study, we investigated the effects of online disinhibition with and without the presence of engagement metrics. Displaying engagement metrics created a highly interactive environment that significantly stimulated participant positivity, leading to a stronger motivation to attitude expression. While online disinhibition did significantly prompt attitude expression, its effect size was relatively small compared to engagement metrics. Critically, online disinhibition was not observed to have a stronger effect under the condition of stronger motivations created by engagement metrics, contrary to predictions of the MOD model. We attribute this result to the profound impact of the engagement metrics as situational factors that overwhelmed the effect of online disinhibition and made it negligible.

The introduction of engagement metrics might have qualitatively changed the nature of attitude expression. Without these metrics, participants' expressions were mainly influenced by individual factors, such as preferences toward the posts. However, with these metrics visible, participants were able to recognize the majority and minority positions based on the distribution of Likes and Dislikes. According to the conformity theory (Deutsch and Gerard, 1955), these preference cues likely motivated participants to conform to the majority. Consequently, participants were more inclined to select the numerically dominant option between Like and Dislike rather than opting for Done.

Moreover, since the engagement metrics were collected from general users, the numbers of Likes and Dislikes likely reflected the general public's opinion of the posts. As a result, most of the participants in Study 1 may have felt that the engagement metrics generally matched their own views. In this situation, following others' opinions and agreeing with the majority did not seem to break social rules or go against their personal preferences. Consequently, attitude expression in this context is influenced more by individuals' perceived consensus and safety from the situation, rather than by the state of online disinhibition.

This study further refined the existing MOD model, which posits that online disinhibition moderates the conversion of motivation into behavior under strong stimuli. Our findings suggest that when encouraged behavior is perceived as comfortable and aligns with situational norms, it does not trigger individuals' cognitive inhibition processes. Consequently, the behavior is more likely to manifest, even among those with lower levels of online disinhibition. Therefore, to prevent overgeneralizing online disinhibition theory's applicability, it is crucial to determine whether the behavior in question might engage individuals' social cognition.

Study 2

Methods and participants

Study 2 explored the situation where the engagement metrics were reversed. The questionnaire in Study 2 was identical to Study 1, except the engagement metrics of each post in the thread were reversed (refer to Figure 1). Similar to the reasoning applied in Study 1, we believed that the factors influencing attitude expression were primarily determined by the thread-reading scene itself. Therefore, we did not conduct a controlled experiment. Instead, we treated Study 1 and Study 2 as a between-participant factor for analysis.

The approach to recruiting participants was identical to the preliminary study. Individuals who participated in the preliminary study and Study 1 were excluded. The survey was conducted on 17 August 2023, and lasted approximately 10.5 minutes. Each participant received 140 JP¥ as a reward for completing it.

Results

A total of 151 individuals completed the survey. Using the same data exclusion criteria as in the preliminary study, 144 valid data remained. We performed descriptive analysis, and Table S3 presents the correlation coefficients, descriptive statistics, and Cronbach's α. To investigate *RQ2*, we combined the data from Studies 1 and 2, with a total sample size of 426. Multiple regression analyses were conducted with attitude expressions, majority and minority attitude expressions serving as the dependent variables. Independent variables included the MMOD, engagement metrics, the interaction between MMOD and engagement metrics, cricket-eating attitude, InfoPlatforms usage rate, age, gender, and education levels, with the outcomes detailed in Table 4. The three regression models were all significant ($p < 0.01$) with R^2 values of 0.16, 0.40, and 0.38, and corresponding f^2 values of 0.19, 0.67, and 0.61, respectively.

Table 4: Standardized coefficients of the regression analysis of majority and minority attitude expressions.

	Attitude expressions		Majority		Minority		

Independent variables	β (Stand error)	95% CI	β (Stand error)	95% CI	β (Stand error)	95% CI	VIF
(Intercept)	0.49 (0.10)**	(0.30, 0.68)	0.56 (0.08)**	(0.40, 0.72)	-0.05 (0.08)	(-0.21, 0.11)	
MMOD	0.06 (0.06)	(-0.05, 0.18)	0.07 (0.05)	(-0.03, 0.17)	0.00 (0.05)	(-0.10, 0.10)	1.74
Engagement metrics	-0.50 (0.10)**	(-0.69, -0.32)	-1.24 (0.08)**	(-1.40, -1.08)	1.15 (0.08)**	(0.99, 1.31)	1.01
MMOD * Engagement metrics	0.03 (0.10)	(-0.15, 0.21)	0.05 (0.08)	(-0.10, 0.20)	-0.03 (0.08)	(-0.18, 0.12)	1.70
Cricket-eating attitude	-0.09 (0.05)	(-0.18, 0.00)	-0.05 (0.04)	(-0.13, 0.02)	-0.08 (0.04)	(-0.15, 0.00)	1.04
InfoPlatforms usage rate	0.13 (0.05)**	(0.04, 0.22)	0.11 (0.04)**	(0.04, 0.19)	0.05 (0.04)	(-0.03, 0.13)	1.02
Age	0.16 (0.05)**	(0.07, 0.25)	0.13 (0.04)**	(0.05, 0.21)	0.07 (0.04)	(-0.01, 0.15)	1.08
Gender (=1)	-0.40 (0.09)**	(-0.57, -0.21)	-0.13 (0.08)	(-0.29, 0.02)	-0.49 (0.08)**	(-0.65, -0.34)	1.04
Education (=1)	-0.17 (0.09)	(-0.35, 0.01)	-0.11 (0.08)	(-0.27, 0.03)	-0.11 (0.08)	(-0.26, 0.05)	1.03
Note: The numbers in brackets indicate standard errors. MMOD: multidimensional measure of online disinhibition, Engagement metrics (1 = Displayed, 2 = Reversed Displayed), Gender (0 = male, 1 = female), Education (0 = Less than a bachelor's degree, 1 = bachelor's degree or higher). * $p < 0.05$, ** $p < 0.01$.							

The coefficients of engagement metrics in all three models were significant ($p < 0.01$) with considerable effect size, particularly in the models for majority and minority attitudes. These findings suggested that participants in Study 2 were less inclined to express attitudes through Likes and Dislikes.

Nonetheless, the coefficients for MMOD and the interaction effects between MMOD and engagement metrics were not significant, with *p*-values for MMOD of 0.25, 0.15, and 0.93, and for the interaction effects of 0.74, 0.51, and 0.70. Similar to the results of Study 1, we interpret these results as indicating that the influence of online disinhibition is negligible compared to the impact of engagement metrics. When participants encountered engagement metrics that contrasted with their preferences, their attitude expressions were dominated by these metrics rather than their level of online disinhibition.

To investigate *RQ3*, we counted the number of Likes and Dislikes each post received across the three studies. To evaluate the development of polarization, we calculated the gap between the ratios of Likes and Dislikes (*i.e.*, (Likes — Dislikes)/sample size) for each study. This allowed us to compare the change of position gap from the preliminary study to Study 1 (or 2), as illustrated in Figure 2. When the colored bar (representing Study 1 or Study 2) and the gray bar (representing the preliminary study) point in the same direction, a longer colored bar indicates that the gap between majority and minority opinions has widened under the influence of engagement metrics. A shorter colored bar in the same direction suggests that the gap

has narrowed, but the majority and minority positions remain unchanged. If the two bars point in opposite directions, this indicates a reversal in the majority and minority positions compared to the preliminary study.

Additionally, Table S4 further presents the change in gap as quotients: the Like–Dislike gaps in Study 1 (or 2) divided by the gaps in the preliminary study. A quotient greater than one indicates increased polarization, a value between zero and one indicates reduced polarization, and a negative value indicates a reversal of the majority and minority positions.

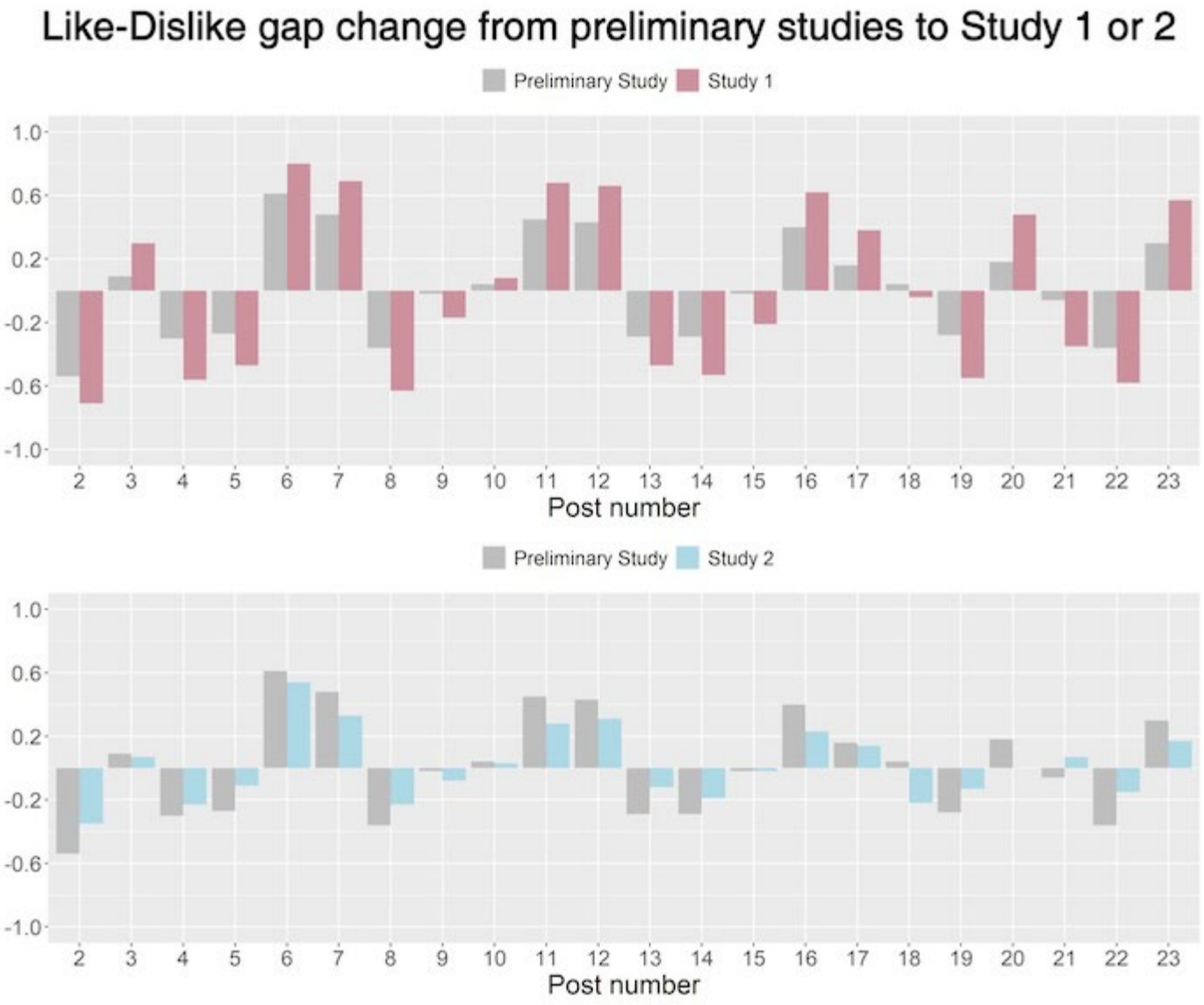

Figure 2: Like–Dislike gap change from preliminary studies to Study 1 or 2.

The results reveal that in Study 1, except for Post 18, an increased gap between Like and Dislike was

observed in all posts. The majority position gained more support, while the minority position received less support, indicating the occurrence of a spiral of silence. Conversely, in Study 2, except for Posts 9, 15, 18, and 21, a decreased gap between Like and Dislike was observed for the remaining posts. This shift occurred because the minority position, labeled as the majority, gained more support, while the majority position, labeled as the minority, gained less support. However, for most posts, the minority position still received less support than the majority position, indicating that a spiral of silence did not occur in Study 2.

Regarding the four exceptional outcomes in Study 2, these might be attributed to the relatively small gap between Likes and Dislikes for these posts in the preliminary study, making them more susceptible to potential sampling bias. It can be inferred that if the engagement metrics were updated in real time, the gaps in Study 1 would likely continue to widen. In Study 2, even though the initial distribution of positions was deliberately set to be almost opposite to the sample's preferences, it would still gradually converge to align with the group's preference over time.

Discussion

In Study 2, we reversed the values of engagement metrics that were manipulated based on real data in Study 1, creating a scenario that diverges from usual circumstances. The results shed light on how people behave when they stumble into an online community where users possess different values. In this context, attitude expression was generally restrained, regardless of whether participants were experiencing online disinhibition. According to the MOD model, this result might be attributed to a lack of motivation to engage. Joinson (2003) posited that individuals tend to selectively engage with content that they are interested in on the Internet. Similarly, Neubaum and Krämer (2018) implied that individuals tend to reduce their expressions in online communities perceived as having a hostile opinion climate. Therefore, when participants perceived their positions as unpopular or not dominant in this survey, their motivation to engage decreased, thereby leading to the irrelevance between attitude expression and online disinhibition.

Moreover, while the labeled majority (actual minority) received increased support and the labeled minority (actual majority) received decreased support in Study 2 compared with the preliminary study, the actual minority seldom surpassed the actual majority in most posts. This indicates that the conformity pressure implied by engagement metrics has a limited effect: When the displayed information violates participants' values, many do not alter their stance to join the apparent majority. However, lacking data on participants' initial attitudes toward each post, we cannot determine whether some participants may have shifted their positions to align with the majority. Further research is necessary to clarify these questions.

General discussion and conclusions

Contributions

The present study's most significant contribution is highlighting the critical role of situational factors in understanding the effect of online disinhibition on online behavior. We examined the influence of online disinhibition and engagement metrics by incorporating three fictitious threads with different settings. The findings revealed that the introduction of engagement metrics overwhelmed the effect of online disinhibition. In Study 1, the engagement metrics that catered to the sample's preferences fostered a comfortable environment for expressing attitudes, where most participants found there was no more need for inhibiting expression. In Study 2, direct opposition between engagement metrics and the sample's preferences led to a significant loss of motivation to express attitudes, thereby marginalizing the role of online disinhibition. These findings suggest that future studies investigating the effects of online disinhibition should fully consider the impact of situational factors.

It is worth noting that the investigation of *RQ3* revealed a social issue of growing concern: the echo

chamber effect (Sunstein, 2001), a phenomenon in which a specific viewpoint or position is repeatedly amplified in a closed online environment. In Study 1, the position labeled as the majority gained more support, and its viewpoints were continuously amplified, leading to the formation of an echo chamber. While Criss, *et al.* (2021) hinted at a possible link between online disinhibition and the emergence of echo chambers, the lack of correlation between majority attitude expression and online disinhibition in Study 1 indicates that such dynamics can emerge independently of online disinhibition. This finding may be attributed to the "Done" button setting in our experiments, which required participants to click one button for each post, resembling a form of compulsory participation (including abstaining). When individuals are obligated to vote, naturally occurring engagement metrics (voting information) may facilitate the formation of echo chambers. While compulsory participation has been considered a means to enhance turnout, political participation, and democracy (Singh, 2021), our results further suggest that it may also significantly intensify polarization when engagement metrics are visible.

Limitations

First, the most significant limitation is the insufficient analytical rigor of our analytical methods, which stems from its exploratory nature. Considering factors such as the order of questionnaires or visual design elements were unlikely to have a substantial impact on attitude expression, we did not strictly implement controlled experiments by randomly assigning participants to two groups. Instead, we combined data from different experimental situations and treated the display of engagement metrics as between-participant factors for analysis. This approach allowed us to include more data for exploratory analysis efficiently but came at the cost of methodological rigor. The findings revealed that the presence and form of engagement metrics were indeed overwhelmingly influential in the statistical models. Thus, we reported these interesting and noteworthy results. However, we acknowledge that while the influence of certain factors, such as timing differences between surveys, the sequence of questionnaires, and visual adjustments, may have been minor, their presence cannot be entirely ruled out. Therefore, more rigorously controlled experiments are necessary in future research.

Second, the examination of how online disinhibition moderates the translation from motivation into behavior was not thoroughly addressed. Based on the MOD model, we introduced engagement metrics to serve as a stimulus, encouraging participants to interact with the posts. However, these manipulations led either to motivation for safe behavior or to a lack of motivation among participants, failing to test the moderation role of online disinhibition. Future research should seek a more precise approach to conceptualize motivation and investigate the role of online disinhibition.

Finally, to prevent the issue of satisficing among online surveys (Krosnick, 1991; Miura and Kobayashi, 2019), the present study introduced a Done button into the experimental situation that does not exist in real situations, which may reduce the ecological validity of our findings. This setting may have led participants to feel forced into participating, resulting in an excessive use of PDAs. In Studies 1 and 2, the average rate of attitude expressions was as high as 59 percent, significantly surpassing expectations. Future research should strive for settings of greater ecological validity that also ensure effective participation.

About the authors

Ruohan Wen is a Ph.D. student at Osaka University, Graduate School of Human Sciences, Social Psychology Lab.
For correspondence regarding this paper, please direct comments to runningwz [at] gmail [dot] com.

Asako Miura is a professor at Osaka University, Graduate School of Human Sciences, Social Psychology Lab.
E-mail: miura [dot] asako [dot] hus [at] osaka-u [dot] ac [dot] jp

Acknowledgments

This study was supported by the JSPS KAKENHI Grant Number JP23KJ1483 and approved by the University's Ethics Committee (approval date: 7 March 2023, approval number: HB022-120).

References

C.Y. Chin, H.P. Lu, and C.M. Wu, 2015. "Facebook users' motivation for clicking the 'like' button," *Social Behavior and Personality*, volume 43, number 4, pp. 579–592.
doi: https://doi.org/10.2224/sbp.2015.43.4.579, accessed 6 June 2025.

S. Criss, E.K. Michaels, K. Solomon, A.M. Allen, and T.T. Nguyen, 2021. "Twitter fingers and echo chambers: Exploring expressions and experiences of online racism using Twitter," *Journal of Racial and Ethnic Health Disparities*, volume 8, pp. 1,322–1,331.
doi: https://doi.org/10.1007/s40615-020-00894-5, accessed 6 June 2025.

M. Deutsch and H.B. Gerard, 1955. "A study of normative and informational social influences upon individual judgment," *Journal of Abnormal and Social Psychology*, volume 51, number 3, pp. 629–636.
doi: https://doi.org/10.1037/h0046408, accessed 6 June 2025.

F. Faul, E. Erdfelder, A. Buchner, and A.-G. Lang, 2009. "Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses," *Behavior Research Methods*, volume 41, number 4, pp. 1,149–1,160.
doi: https://doi.org/10.3758/BRM.41.4.1149, accessed 6 June 2025.

R.A. Hayes, C.T. Carr, and D.Y. Wohn, 2016. "One click, many meanings: Interpreting paralinguistic digital affordances in social media," *Journal of Broadcasting & Electronic Media*, volume 60, number 1, pp. 171–187.
doi: https://doi.org/10.1080/08838151.2015.1127248, accessed 6 June 2025.

T. Hogg and G Szabo, 2009. "Diversity of user activity and content quality in online communities," *Proceedings of the International AAAI Conference on Web and Social Media*, volume 3, number 1, pp. 58–65.
doi: https://doi.org/10.1609/icwsm.v3i1.13940, accessed 6 June 2025.

A.N. Joinson, 2003. *Understanding the psychology of Internet behaviour: Virtual worlds, real lives*. New York: Palgrave Macmillan.

J.A. Krosnick, 1991. "Response strategies for coping with the cognitive demands of attitude measures in surveys," *Applied Cognitive Psychology*, volume 5, number 3, pp. 213–236.
doi: https://doi.org/10.1002/acp.2350050305, accessed 6 June 2025.

A. Kurek, P.E. Jose, and J. Stuart, 2019. "'I did it for the LULZ': How the dark personality predicts online disinhibition and aggressive online behavior in adolescence," *Computers in Human Behavior*, volume 98, pp. 31–40.
doi: https://doi.org/10.1016/j.chb.2019.03.027, accessed 6 June 2025.

B. Latané, 1981. "The psychology of social impact," *American Psychologist*, volume 36, number 4, pp. 343–356.
doi: https://doi.org/10.1037/0003-066X.36.4.343, accessed 6 June 2025.

M.R. Maniaci and R.D. Rogge, 2014. "Caring about carelessness: Participant inattention and its effects on

research," *Journal of Research in Personality*, volume 48, pp. 61–83.
doi: https://doi.org/10.1016/j.jrp.2013.09.008, accessed 6 June 2025.

A. Matsui, 2021. "Does the number of likes and views matter? An experimental study of the influence of SNS majority users," *Japan Marketing Review*, volume 2, number 1, pp. 30–37.
doi: https://doi.org/10.7222/marketingreview.2021.004, accessed 6 June 2025.

A. Miura and T. Kobayashi, 2019. "Survey satisficing biases the estimation of moderation effects," *Japanese Psychological Research*, volume 61, number 3, pp. 204–210.
doi: https://doi.org/10.1111/jpr.12223, accessed 6 June 2025.

L. Muchnik, S. Aral, and S.J. Taylor, 2013. "Social influence bias: A randomized experiment," *Science*, volume 341, number 6146 (9 August), pp. 647–651.
doi: https://doi.org/10.1126/science.1240466, accessed 6 June 2025.

G. Neubaum and N.C. Krämer, 2018. "What do we fear? Expected sanctions for expressing minority opinions in offline and online communication," *Communication Research*, volume 45, number 2, pp. 139–164.
doi: https://doi.org/10.1177/0093650215623837, accessed 6 June 2025.

Nikkei, 2022. "Introduction of edible cricket powder in school lunches — A nationwide first in Tokushima," at https://www.nikkei.com/article/DGXZQOCC24BFE0U2A121C2000000/, accessed 6 June 2025.

E. Noelle-Neumann, 1974. "The spiral of silence: A theory of public opinion," *Journal of Communication*, volume 24, number 2, pp. 43–51.
doi: https://doi.org/10.1111/j.1460-2466.1974.tb00367.x, accessed 6 June 2025.

A. Oshio, S. Abe, and P. Cutrone, 2012. "Development, reliability and validity of the Japanese version of Ten Item Personality Inventory (TIPI-J)," *Japanese Journal of Personality*, volume 21, number 1, pp. 40–52.
doi: https://doi.org/10.2132/personality.21.40, accessed 6 June 2025.

A.P. Schouten, P.M. Valkenburg, and J. Peter, 2007. "Precursors and underlying processes of adolescents' online self-disclosure: Developing and testing an 'Internet-attribute-perception' model," *Media Psychology*, volume 10, number 2, pp. 292–315.
doi: https://doi.org/10.1080/15213260701375686, accessed 6 June 2025.

S.P. Singh, 2021. *Beyond turnout: How compulsory voting shapes citizens and political parties*. Oxford: Oxford University Press.
doi: https://doi.org/10.1093/oso/9780198832928.001.0001, accessed 6 June 2025.

J. Stuart and R. Scott, 2021. "The measure of online disinhibition (MOD): Assessing perceptions of reductions in restraint in the online environment," *Computers in Human Behavior*, volume 114, 106534.
doi: https://doi.org/10.1016/j.chb.2020.106534, accessed 6 June 2025.

J. Suler, 2004. "The online disinhibition effect," *Cyberpsychology & Behavior*, volume 7, number 3, pp. 321–326.
doi: https://doi.org/10.1089/1094931041291295, accessed 6 June 2025.

E.M. Sumner, L. Ruge-Jones, and D. Alcorn, 2018. "A functional approach to the Facebook Like button: An exploration of meaning, interpersonal functionality, and potential alternative response buttons," *New Media & Society*, volume 20, number 4, pp. 1,451–1,469.
doi: https://doi.org/10.1177/1461444817697917, accessed 6 June 2025.

C.R. Sunstein, 2001. *Republic.com*. Princeton, N.J.: Princeton University Press.

R. Udris, 2014. “Cyberbullying among high school students in Japan: Development and validation of the Online Disinhibition Scale,” *Computers in Human Behavior*, volume 41, pp. 253–261.
doi: https://doi.org/10.1016/j.chb.2014.09.036, accessed 6 June 2025.

R. Wen and A. Miura, 2024. “Development of the multi-dimensional measure of online disinhibition and examination of its validity and reliability,” *Osaka Human Sciences*, volume 10, pp. 19–38.
doi: https://doi.org/10.18910/94827, accessed 6 June 2025.

R. Wen and A. Miura, 2023. “Online disinhibition: Reconsideration of the construct and proposal of a new model,” *Osaka Human Sciences*, volume 9, pp. 63–80.
doi: https://doi.org/10.18910/90710, accessed 6 June 2025.

M.F. Wright and S. Wachs, 2021. “Does empathy and toxic online disinhibition moderate the longitudinal association between witnessing and perpetrating homophobic cyberbullying?” *International Journal of Bullying Prevention*, volume 3, number 1, pp. 66–74.
doi: https://doi.org/10.1007/s42380-019-00042-6, accessed 6 June 2025.

M.F. Wright, B.D. Harper, and S. Wachs, 2019. “The associations between cyberbullying and callous-unemotional traits among adolescents: The moderating effect of online disinhibition,” *Personality and Individual Differences*, volume 140, pp. 41–45.
doi: https://doi.org/10.1016/j.paid.2018.04.001, accessed 6 June 2025.

Appendix

Table S1: Results of the correlations analysis.

	MMOD	F1	F2	F3	Sum	Like	Dislike
F1	0.82**						
F2	0.65**	0.26**					
F3	0.48**	0.23**	-0.06				
Sum	0.21**	0.12*	0.16**	0.15**			
Like	0.25**	0.13*	0.21**	0.18**	0.75**		
Dislike	0.09	0.06	0.06	0.07	0.83**	0.26**	
M	3.40	2.60	3.61	4.48	10.64	5.32	5.31
SD	0.53	0.76	0.82	0.80	6.98	4.01	4.74
α	0.71	0.72	0.69	0.69			

Note: MMOD: multidimensional measure of online disinhibition, F1: unique perspective on online environment, F2: change of alienation cognition, F3: change of relationship cognition, Sum: attitude expressions. * $p < 0.05$, ** $p < 0.01$.

Table S2: Exploratory analyses using 3-factor of MMOD.						
	Factor = F1		**Factor = F2**		**Factor = F3**	
	β	*p*-value	β	*p*-value	β	*p*-value
(Intercept)	−0.15	0.05	−0.15	0.05	−0.16	0.03
MMOD Factor	0.10	0.03	0.11	0.02	0.13	0.01
Engagement metrics (= 1)	0.62	<0.01	0.61	<0.01	0.63	<0.01
MMOD Factor * Engagement metrics	−0.00	0.97	−0.09	0.25	−0.15	0.04
InfoPlatforms usage rate	0.16	<0.01	0.15	<0.01	0.16	<0.01
Age	0.11	<0.01	0.10	<0.01	0.10	<0.01
Gender (= 1)	−0.13	0.08	−0.14	0.07	−0.14	0.06
Education (= 1)	−0.09	0.23	−0.08	0.28	−0.07	0.34
Note: MMOD: multidimensional measure of online disinhibition, F1: Unique perspective on online environment, F2: change of alienation cognition, F3: change of relationship cognition, Engagement metrics (0 = not displayed, 1 = displayed), Gender (0 = male, 1 = female), Education level (0 = Less than a bachelor's degree, 1 = bachelor's degree or higher).						

Table S3: Results of the correlations analysis.							
	MMOD	**Cricket-eating attitude**	**Sum**	**Like**	**Dislike**	**Majority**	**Minority**
Cricket-eating attitude	-0.09						
Sum	0.12	-0.15					
Like	0.14	-0.18*	0.82**				
Dislike	0.07	-0.09	0.89**	0.46**			
Majority	0.23**	-0.20*	0.77**	0.65**	0.66**		
Minority	0.00	-0.06	0.86**	0.69**	0.78**	0.34**	

M	3.61	3.98	10.53	5.42	5.11	3.55	6.98
SD	0.54	0.40	7.73	4.02	5.01	4.15	5.27
α	0.74	0.87					

Note: MMOD: multidimensional measure of online disinhibition, Sum: number of attitude expressions, Majority: number of majority attitude expressions, Minority: number of minority attitude expressions. * $p < 0.05$, ** $p < 0.01$.

Table S4: Like and Dislike ratio in the three studies and the change of attitude gap.

Preliminary Study				**Study 1**				**Study 2**			
Post ID	Like (%)	Dislike (%)	Gap	Like (%)	Dislike (%)	Gap	**Gap change**	Like (%)	Dislike (%)	Gap	**Gap change**
2	0.06	0.60	-0.54	0.03	0.74	-0.71	**1.32**	0.13	0.48	-0.35	**0.66**
3	0.24	0.14	0.09	0.45	0.15	0.30	**3.27**	0.28	0.22	0.07	**0.75**
4	0.09	0.39	-0.30	0.06	0.62	-0.56	**1.87**	0.13	0.36	-0.23	**0.76**
5	0.09	0.36	-0.27	0.08	0.55	-0.47	**1.75**	0.15	0.26	-0.11	**0.42**
6	0.64	0.03	0.61	0.83	0.04	0.80	**1.31**	0.62	0.08	0.54	**0.89**
7	0.54	0.07	0.48	0.74	0.05	0.69	**1.45**	0.44	0.12	0.33	**0.69**
8	0.10	0.46	-0.36	0.04	0.67	-0.63	**1.75**	0.15	0.38	-0.23	**0.64**
9	0.13	0.14	-0.02	0.16	0.32	-0.17	**11.03**	0.15	0.23	-0.08	**5.51**
10	0.22	0.18	0.04	0.33	0.26	0.08	**1.94**	0.23	0.19	0.03	**0.86**
11	0.49	0.05	0.45	0.73	0.05	0.68	**1.53**	0.41	0.13	0.28	**0.62**
12	0.48	0.05	0.43	0.70	0.05	0.66	**1.53**	0.44	0.13	0.31	**0.71**
13	0.05	0.34	-0.29	0.06	0.53	-0.47	**1.64**	0.15	0.26	-0.12	**0.41**
14	0.07	0.36	-0.29	0.05	0.58	-0.53	**1.82**	0.11	0.30	-0.19	**0.64**
15	0.17	0.19	-0.02	0.17	0.38	-0.21	**10.56**	0.18	0.20	-0.02	**1.03**
16	0.41	0.01	0.40	0.65	0.02	0.62	**1.55**	0.34	0.11	0.23	**0.57**
17	0.31	0.16	0.16	0.50	0.12	0.38	**2.43**	0.30	0.16	0.14	**0.89**
18	0.26	0.22	0.04	0.27	0.31	-0.04	**-0.88**	0.16	0.38	-0.22	**-5.34**
19	0.05	0.32	-0.28	0.02	0.58	-0.55	**1.98**	0.15	0.28	-0.13	**0.47**
20	0.24	0.06	0.18	0.54	0.06	0.48	**2.68**	0.17	0.17	0.00	**0.00**
21	0.14	0.20	-0.06	0.10	0.44	-0.35	**5.52**	0.26	0.19	0.07	**-1.10**
22	0.07	0.42	-0.36	0.04	0.61	-0.58	**1.62**	0.15	0.31	-0.15	**0.43**
23	0.37	0.06	0.30	0.64	0.07	0.57	**1.86**	0.32	0.15	0.17	**0.55**

Editorial history

Received 11 February 2025; revised 5 May 2025; accepted 6 June 2025.

The impact of engagement metrics overwhelms the influence of online disinhibition
by Ruohan Wen and Asako Miura.
First Monday, volume 30, number 7 (July 2025).
doi: https://dx.doi.org/10.5210/fm.v30i7.14146