



Is “Good vibes only” really good? Investigating perceptions of toxic positivity on social media

new media & society

1–32

© The Author(s) 2025

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/14614448251396941

journals.sagepub.com/home/nms**Stacy Siqi Wong**

Nanyang Technological University, Singapore

Aretha L. H. Wan

Nanyang Technological University, Singapore

Zijian Lew 

Nanyang Technological University, Singapore

Abstract

On social media, posting positive content should generate positive emotions via emotional contagion. Yet, emerging discussions regarding toxic positivity (TP) suggest that positivity can backfire, showing the limits of emotional contagion. Through the perspective of neoliberal self-help ideology, this research investigated how people understand and perceive TP. Study 1 found, via focus groups, that social media posts containing TP involve two message characteristics—overgeneralization and commanding words—and two psychological processes—ignoring negativity and perceived poster privilege. Additionally, post ephemerality mitigates the negative effects of TP. Study 2 experimentally tested these findings. It found that the effect of post positivity (low positivity / high positivity / toxic positivity message characteristics) on post liking was mediated by *ignoring negativity* and *perceived privilege*. However, ephemerality did not moderate the aforementioned mediation relationships. Therefore, the concepts associated with TP—overgeneralization, commanding words, ignoring negativity, and perceived privilege—can be understood as boundary conditions for online emotional contagion.

Corresponding author:

Zijian Lew, Wee Kim Wee School of Communication and Information, Nanyang Technological University, 31 Nanyang Link, Singapore 637718, Singapore.

Email: zlew@ntu.edu.sg

Keywords

Emotional contagion, neoliberalism, self-presentation, social media, toxic positivity

Among various types of toxic online behaviors (e.g. trolling, flaming), toxic positivity (TP) is unique as it is associated with good intentions, and is not considered uncivil (Goodman, 2022). Phrases such as “you’ll be fine” or “everything happens for a reason” seem positive, but scholars have argued that they make people feel lousy (Goodman, 2022; Upadhyay et al., 2022). This phenomenon—producing negative cognitive and affective outcomes via superficially positive messages—has been labeled TP (Shipp and Hall, 2024). Although TP can occur face-to-face, the trend of actively seeking out or sending positive messages that supposedly engender positive emotions (e.g. #hopecore) is happening on a grander scale online than offline (Fawehinmi, 2024). The potential for the sharing of such messages is also greater online than offline, due to the high visibility of online messages (Treem and Leonardi, 2013). Therefore, this research focuses on how messages on social media may contribute toward TP.

Researchers have posited several reasons why people may perceive a positive message as toxic. First, messages are toxic if they encourage people to suppress negative emotions while displaying positive emotions (Lew and Flanagin, 2025; Shipp and Hall, 2024). Second, toxic messages superficially convey optimism, but to unrealistic extents (Lew and Flanagin, 2025; Shipp and Hall, 2024). Indeed, Upadhyay et al. (2022) argued that calls for gratitude—amid death and loss of jobs—during the COVID-19 pandemic were toxic, especially when no end to the pandemic was foreseeable then. A third reason is maladaptive social comparison: people may use positivity to make themselves appear better or make others appear worse (Salopek and Eastin, 2024). Indeed, interviews and surveys have found that TP is associated with negative outcomes, such as unhealthy upward social comparisons (Salopek and Eastin, 2024), poorer social relationships with people who send messages containing TP (Goodman, 2022), and decreased well-being for audiences of messages containing TP (Shipp and Hall, 2024).

Researchers have studied online TP from different perspectives. For example, Lecompte-Van Poucke (2022) undertook discourse analysis, Upadhyay et al. (2022) used natural language processing to detect TP, Lew and Flanagin (2025) studied how TP messages backfired on message senders, and Salopek and Eastin (2024) examined senders’ motivations for TP. We adopt the perspective that some people genuinely want to uplift others emotionally, but fail for various reasons (Goodman, 2022; Shipp and Hall, 2024). After all, positive social media messages are often meant to be uplifting or motivational, yet they produce TP when they fail in their goal due to audience factors such as perceived lack of empathy from those providing such motivations (Putra et al., 2023). Therefore, we argue that a holistic understanding of why TP occurs requires TP to be viewed not as a standalone concept, but one that is understood alongside actual positivity. We argue that TP is a result of a failed attempt at genuine positivity—in some cases, a positive message is interpreted positively, while in other cases, audiences form their own perceptions of the same positive message and ascribe negativity to it. From this perspective, sharing positive social media posts can spread genuine positivity, but various factors can restrict the contagion of positive emotions, resulting in TP.

Through our studies, we aimed to investigate what factors contribute to an audience's perception of whether a social media post contributes to TP. Study 1 made use of focus group discussions to understand what TP means to people and what factors elicit TP. Study 2 experimentally tested these factors to determine if they indeed produce perceptions of TP from positive posts—as a way to triangulate our findings from Study 1. Our studies found that post characteristics such as overgeneralization and commanding words elicit perceptions of TP through encouraging audiences to ignore negativity and through audience's perception of a sender's privilege. We also draw upon social media affordances to argue that perceptions of TP are less severe when posts are ephemeral (i.e. they automatically disappear after a short time) than when they are persistent (Treem and Leonardi, 2013). This is because audiences understand that ephemeral posts are meant to be trivial and therefore do not treat such posts too seriously (Bayer et al., 2016; Xu et al., 2016). However, the influence of ephemerality on TP perceptions was supported only in Study 1.

Toxic positivity and neoliberalism

To understand how audiences may react toward TP, it is important to understand TP in the wider sociocultural context. Lecompte-Van Poucke (2022) studied TP as a discursive construct arising from neoliberal positive thinking ideology, which has promoted the self-help industry to such an extent that success appears to depend on one's ability to metaphorically pull oneself up by one's own bootstraps. The study found that TP discourse is rooted in the idea that positivity is the state that everyone should aim for, and that individuals have full control over themselves to achieve positivity. TP is then often used by message senders, who have substantial power over audiences, to influence what audience members should do to be more positive, regardless of the audiences' personal circumstances (Lecompte-Van Poucke, 2022). Similarly, neoliberal self-help thinking results in the expectation that happiness is attainable because everyone has the power to attain it, but continuous personal growth and happiness is often not easily achieved—and the seemingly compulsory pursuit of happiness therefore results in negative outcomes such as stress or anxiety (Becker et al., 2021). Additionally, Becker et al. (2021) pointed out that neoliberalism promotes the perspective that personal growth is only achievable through competition with others, adding to the stress when pursuing this expectation of happiness. The overall result of neoliberal self-help ideology, then, is the creation of a culture where positivity and confidence have been accorded a premium (Orgad and Gill, 2022), but it is also toxic because the messages that prop up this ideology are contrived—designed to meet the expectations surrounding this culture—and not a result of emotions organically experienced.

Amid such a culture, Althusser's (1971) concept of interpellation can be used to understand why TP has emerged. Social media interpellate people into a culture of positivity—the positive messages on social media call (or in Althusser's terminology, hail) people into adopting a neoliberal self-help ideology, and many people answer that call. That is, people internalize the need to always be positive, influenced by positive psychology and the social demand that they perform in ways aligned with the social expectations of positivity. But psychologically, the need to constantly dig deeper within ourselves—even in times of struggle—to be more positive is untenable. The neoliberal self-help

ideology of compelling oneself to be more positive through sheer willpower thus hits its limit, revealing a condition that scholars have described as TP: “the overgeneralization of a positive state of mind that encourages using positivity to suppress and displace any acknowledgement of stress and negativity” (Upadhyay et al., 2022: 64), or more succinctly, the platitudinous display of positivity (Lew and Flanagin, 2025). In the next sections, we explain why the sharing of positive messages—an ostensibly good thing—can turn toxic, leading to these negative outcomes.

Positive posts and emotional contagion on social media

Sharing positive messages on social media can be a good thing due to emotional contagion, defined as the non-conscious process of emotion transfer from one person to another (Kramer et al., 2014). Presumably, people exposed to positive messages will feel positive, and perhaps create positive messages themselves. Mimicry and social appraisal have been suggested as possible explanations for emotional contagion (Parkinson, 2011). Specifically, the mimicry of emotional expressions during social interactions may arouse similar emotions through facial feedback. Meanwhile, social appraisal is the process through which individuals use others’ emotions as information to make inferences regarding an object or situation. Therefore, as two parties interact, their emotions would converge.

Empirical studies have shown that emotional contagion can occur in digital environments via social appraisal, although mimicry is unlikely when nonverbal information is lacking (e.g. Hancock et al., 2008; Kramer et al., 2014). Considering that more positive and arousing posts are shared more frequently online (Berger and Milkman, 2012), audiences may be receptive to these posts due to the transfer of positive emotions. This also encourages posters (i.e. people who make posts on social media) to upload posts evoking positive emotions, such as happy moments or motivational quotes, creating an environment that encourages the viewing and/or sharing of positive experiences. Moreover, the rise of platform capitalism further exacerbates this issue, with social media companies seeking to attract more users or to gain a larger market share (Srnicsek, 2017). In the platform economy, social media act as intermediaries between users, making use of user data to refine algorithms that monetize the connection between users (Liang et al., 2022). Put differently, social media algorithms are designed to get people hooked, and thus repeatedly show feel-good messages to people who would like such posts. In this way, positivity has become commercialized and perpetuated in what is akin to a giant echo-chamber of positivity—and emotional contagion is the psychological mechanism that facilitates the propagation of positive emotions on social media.

However, emotional contagion may fail and lead to negative outcomes. The upcoming section presents two possible factors that may inhibit the contagion of positive emotions and lead to TP: emotional suppression and face-threatening acts. Emotional suppression was identified from previous research (e.g. Lew and Flanagin, 2025; Shipp and Hall, 2024) that highlighted its cardinal role in producing TP. Yet, past research on TP in the context of social media has examined emotional suppression only quantitatively or theoretically (e.g. Lew and Flanagin, 2025; Salopek and Eastin, 2024). We therefore decided to revisit this concept qualitatively, via focus groups, in Study 1. We

examined face-threatening acts (Brown and Levinson, 1987) because a substantial number of TP messages appear so didactic that they seem to threaten people's negative face, that is, disregard people's need for autonomy. Goodman (2022) gave several examples, such as "Just be happy/positive" or "Never give up." Such messages may even signal superiority, which could threaten people's positive face, that is, their need to feel valued. Hence, the failure of emotional contagion and perceptions of toxicity may be interpreted as a face threat.

Emotional suppression

The contagion of positive emotions may be hindered when there is a conflict between an audience's emotional state and the positive content of a post, potentially leading to TP. Emotional suppression is "the conscious inhibition of one's own emotional expressive behavior while emotionally aroused" (Gross and Levenson, 1993: 970). It is an ineffective way to regulate negative emotion as it makes people aware of their own inauthenticity, resulting in worse well-being and poorer social functioning (English and John, 2013; Gross and John, 2003).

At a broader level, emotional suppression can derive from and reinforce norms of positivity on social media that may inhibit the communication of authentic social information. Due to the norm of positive messages on social media, users are constantly exposed to positivity, regardless of whether their internal state matches that of the positivity shown to them in their feeds (see Swart, 2021). This deepens the disparity between their internal states and the positive messages they are exposed to, potentially leading to TP.

In sum, positive social media posts may encourage audiences to suppress negative emotions, potentially resulting in deleterious effects among audiences. This study therefore explores the question of how positive social media posts encourage the suppression of negative emotions, potentially resulting in deleterious effects among audiences.

Face-threatening acts

The contagion of positive emotions may also fail because positive posts potentially threaten audiences' face. According to Brown and Levinson (1987), people have a negative face and a positive face. People's negative face is threatened when others impinge on their autonomy, perhaps by giving them directive commands. People's positive face is threatened when others ignore or disapprove of their desires (Brown and Levinson, 1987).

Social media posts that have directive overtones may engender TP because they threaten audiences' negative face. Goodman (2022) gives a few examples: when one has career struggles, a toxic comment may be "Your attitude determines how successful you will be," or when one is upset, a toxic comment may be "You have so much to be grateful for and be happy about. Focus on that." These comments contain directives: to have a positive attitude or to focus on being grateful. These directives may threaten audiences' negative face, making the comments well-intentioned but irritating. Social media posts that contain similar directives may therefore lead to TP.

Toxic posts may also convey that the poster is superior to audiences in some way, thus threatening audiences' positive face. For example, one could log on to LinkedIn to read posts about people dishing out career advice: how they found career success after initial setbacks and if you were willing to work as hard as they did, you too could find great success (see Orgad and Gill, 2022). From the perspective of face-threatening acts, such messages reflect a threat to audiences' positive face desires to appear competent because they are bragging or patronizing (Hastings and Bell, 2018). Hence, posts that contain similar comments may engender TP.

In all, unrealistically positive posts can threaten audiences' positive or negative face. Instead of spreading positivity via emotional contagion, they backfire, ironically spreading negative feelings. However, the potential harm of TP may be limited under one condition: if posts were ephemeral, rather than persistent.

Ephemerality

We examined ephemerality because it is a unique social media affordance that influences audiences' perceptions of posters' communicatory goals (Bayer et al., 2016; Zhao et al., 2013), which may affect how audiences perceive TP in a message. Whereas social media traditionally contain persistent content, newer platforms such as Instagram Stories and Snapchat have introduced ephemeral content that have a fixed expiration time, after which the content cannot be publicly accessed (Bayer et al., 2016). For example, Instagram stories are deleted after 24 hours and no longer viewable to followers after that (Xu et al., 2016). Ephemerality (or its opposite, persistence) can therefore be viewed as an affordance—or action possibility—that is unique to social media (Treem and Leonardi, 2013).

Such time-limited content provides a way for posters to manage how long their content is seen by audiences. People also have different goals for using social media, including (a) the management of current self-presentation needs and (b) the archiving of past data that pertains to long-term self-presentation and identity (Zhao et al., 2013). Ephemeral social media posts cater to the former, which emphasizes posters' needs to express themselves in a spur-of-the-moment fashion. Indeed, ephemeral posts are more associated with trivial, mundane content and with lower salience of the posters' own self-presentations than persistent posts (Xu et al., 2016).

As ephemeral media are not typically taken seriously by posters or audiences, post ephemerality can limit the detrimental effects of TP. Audiences may not perceive positive posts shared via ephemeral social media as threats to their positive and negative faces or as instructions to suppress their emotions, which consequently reduces their likelihood of perceiving these positive posts as toxic. In other words, the earlier arguments on face-threatening acts and emotional suppression may apply only when posts are persistent and not ephemeral.

Given the foregoing, we proposed the following research questions: What makes a social media post toxic? Do emotional suppression, face-threatening acts, ephemerality—or any other factors—influence people's perceptions of TP? If so, how?

Study 1

Study 1 used focus groups to understand audience perspectives of positive motivational quotes on social media and sought to learn how positive messages can generate negative responses in audiences. Focus groups were chosen over other methods because they allowed participants to bounce ideas off each other. Crucially, focus group participants could show their own example of a social media post with TP, and the focus group could subsequently discuss why that post may be toxic. This suits the exploratory nature of Study 1.

Method

Participants were recruited from Nanyang Technological University, a public university in Singapore, via university email lists and text messaging channels. A short survey was sent out to find eligible participants who were interested in the study, who must also have fulfilled our criteria of being frequent social media users (participants must have used social media at least a few times a week). Students from various degree programs (e.g. Electrical Engineering, Business, Sociology, Computer Science) were recruited to obtain diverse perspectives. Participants were eligible as long as they were students who were willing to talk about TP. There were no other exclusion criteria.

Six 60-minute focus groups, comprising a total of 22 participants (17 female, 5 male), were conducted online on Zoom in December 2023. Participants' ages ranged from 21 to 26 years old. All participants indicated that they frequently used Instagram, and many of them used Tiktok and Facebook as well. We used Zoom as it allowed our participants to join the discussions from their own homes, making scheduling and recruitment easier while ensuring our participants' comfort. Discussions were recorded and transcribed. Each group had 3 to 5 participants, who were encouraged to share a post from their own social media feed that they thought contributed to TP. Each participant was paid \$15 in the local currency (approximately \$11 USD). See the Online Supplemental Materials for a summary of our participants' demographics and the posts they brought. Researchers also separately prepared several positive posts to obtain participants' thoughts regarding whether they perceived the posts as effective or as toxic (see Figure 1).

The first and second authors were present during all focus groups discussions. One moderated the discussion and the other handled administrative matters. The discussions were split into 3 sections: introductory questions to hear participants' understanding of TP, discussion of participants' self-sourced TP posts, and discussion of researchers' compiled positive/motivational posts. Participants were anonymized during transcription. Based on the transcripts, the first and second authors thematically analyzed each discussion independently. They then convened to reach a consensus on the identified themes and the categorization of quotes into each theme.

Results

The posts that participants brought to the discussion largely consisted of content such as motivational quotes situated on a simple background, realistic images with captions

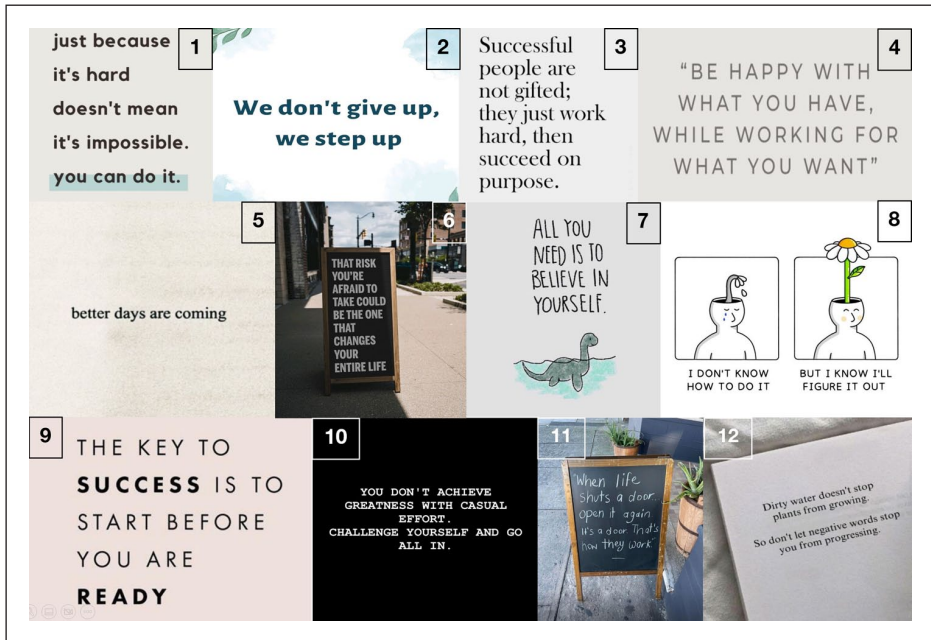


Figure 1. Posts shown to participants in Study 1.

containing life advice, personal anecdotes, and screenshots encouraging or portraying ideal behaviors (see Figure 2). Most of the posts emphasized the poster's encouragement toward audiences to pursue their ideals, such as perseverance, self-love, and achievement, while other posts included assurances and personal experiences of improvement.

In the first-cycle coding, we coded the transcripts while following Krueger and Casey's (2015) advice to look out for concepts that participants mentioned frequently, with intensity, with detail, with internal consistency, or with an indication that the concept was important. We also took note of a concept if multiple participants talked about it. In the second-cycle coding, we identified commonalities across our codes and abstracted the codes into several themes, which we present below.

Ignoring negativity. The majority of participants agreed that the disregard of negative emotions was a crucial aspect of TP. A range of subthemes was discussed, including the suppression, denial, and invalidation of negative emotions. Participants described this theme with substantial self-referencing (e.g. "I don't think it will help me feel better") and strong claims (e.g. "it becomes even more insidious and damaging"). P4 (Male, 22) understood TP as ". . . suppressing any negative feelings or expressions that other people are having, and then constantly encouraging them to only display positive emotions." Similarly, when asked what they understood of TP, P22 (Female, 22) said: "I feel that it's reframing your thoughts, but taken to a very large, very very large extent, such that you really just entirely neglect how you actually feel."

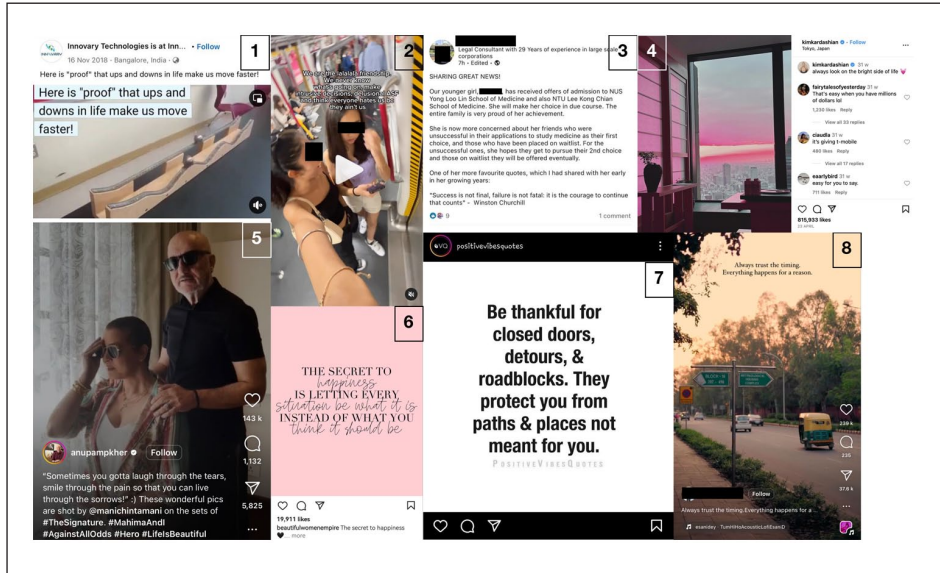


Figure 2. Sample TP posts submitted by participants in Study 1.

Posts encouraging audiences to be positive may also feel invalidating, leading audiences to feel unheard and perceiving the poster to be dismissive. P6 (Female, 22), who shared the post in Figure 2 Panel 1, defined a post with TP as:

It can be well-meaning, but at the same time it very much dismisses the behind-the-scenes struggles and nuances behind how the person is actually feeling, and it doesn't really encourage or comfort them, because they don't feel seen or heard in the process.

Some posts even encourage audiences to evade negativity entirely. P7 (Female, 22), in response to a post (see Figure 2 Panel 2) shared by P5 (Female, 22), explained that such posts were “like a form of escapism, where [people] constantly choose to not face the reality, which they deem to be negative, in order to maintain a perpetually sunny disposition.”

Therefore, ignoring negativity—the combination of suppression, invalidation, and evasion of negative emotions—by replacing negative emotions with positive perspectives or by trivializing them, were prominent characteristics of TP that almost all participants pointed out. This can be understood through the perspective of neoliberal self-help ideology: the belief that people are responsible for their own well-being (Lecompte-Van Poucke, 2022; Orgad and Gill, 2022). The unquestioned belief in this ideology encourages posters to, as they strive for positivity, push aside their negative feelings while interpreting negative emotions as a personal failure to maintain positivity. Combining this with how the suppression of negative emotions ironically fosters greater feelings of negativity (Gross and John, 2003), it is understandable why many participants perceived that ignoring negativity is a crucial part of TP.

Perceived privilege. Another recurring opinion of participants was the perceived privilege that the posters had, as their motivational posts were unrelatable to the general audience or unintentionally contained condescension. Participants showed great emotion when discussing this theme and described their displeasure at a post suggesting that people were successful for the sole reason that they worked hard (see Figure 1 Panel 3), for example, “You’re blaming me. You’re calling me lazy” (P9, Female, 25), “[It] looks down on people who did not achieve such success” (P2, Male, 24). When discussing another post with the quote, “You don’t achieve success with casual effort. Challenge yourself and go all in” (Figure 1 Panel 10), P13 (Female, 22) expressed how they understand the good intentions of the poster, but still perceived the poster as insensitive and unaware of their privilege:

I get the idea is to motivate a person to try even harder, but if I’m already having a hard time, and the other person is not even trying to understand how hard I’m trying. . . Instead, [they’re] calling my effort casual. . . if they think my effort is casual, then it probably means that they’re looking down on me.

Perceived privilege can also take the form of perceived bragging or humblebragging. P9 (Female, 25), when describing TP posts in general, explained that:

It’s about bragging and about how good they are, even if they talk about, “Oh, you know, I had like a really tough past and something,” but it’s in a way flexing about who [they] are, and at the same time also making [people] feel bad about [themselves].

Additionally, numerous participants mentioned that they perceive posters of posts with TP as engaging in humblebragging, subtly boasting about their achievements while downplaying their efforts. For example, when discussing a LinkedIn post by a father congratulating his daughter on being accepted into two universities while also mentioning how his daughter worries for those who did not get accepted (Figure 2 Panel 3), P4 (Male, 22) said: “It feels a bit like rubbing salt on people’s wounds.”

Perceived privilege can also be magnified if the poster is famous or powerful. P13 (Female, 22) brought to the focus group a post by Kim Kardashian with the caption: “Always look on the bright side of life” (Figure 2 Panel 4). P13 said that “we, as viewers and consumers, know that she lives a very privileged life. So when she says these kinds of things, it becomes a bit out of touch.” Relatedly, Kim Kardashian’s caption demonstrates the theme of ignoring negativity, as if people can compel themselves to be optimistic through sheer willpower.

Perceived privilege can thus be interpreted as a consequence of audiences experiencing face threats. Audiences’ positive face is threatened when they perceive a poster as giving unsolicited advice, and this is exacerbated when the poster engages in humblebragging or comes from a well-to-do background. Perceived privilege is also linked to Putra et al.’s (2023) finding that TP is associated with a lack of empathy, that is, audiences perceive posters to be privileged to the extent that they cannot understand or feel for ordinary individuals. As such, even if the poster has no ill intentions, audiences still interpret their posts as having TP.

Ephemerality. Participants also brought up how the ephemerality of a post influenced their opinion. P9 (Female, 25) explained that “[Instagram] Stories almost feel like a stream of consciousness. . . Whereas when you [upload something as] your post, it’s really saying this is almost something integral to me and my identity as a person.” P13 (Female, 22) also described a similar understanding of ephemerality: “Because Instagram stories are shorter and more fleeting. . . people usually post on Instagram stories, I assume, with the intention of it just being like a short moment that speaks to them at a point in time.” This suggests that audiences perceive persistent posts as more detrimental because persistent posts reflect a core value that the poster believes in. In contrast, there is a discounting effect for ephemeral posts because ephemeral posts are typically meant for trivial messages (Bayer et al., 2016). Audiences therefore do not perceive an ephemeral post with TP to be as detrimental as if a persistent post had TP.

Overgeneralization. The content of the post was frequently discussed by participants as well. In particular, posts containing statements with unrealistic standards that a poster foisted upon an audience contributed to the theme of overgeneralization. This theme had high specificity, as participants often mentioned the generalization or simplification of statements as the reasoning behind perceived toxicity. Participants emphasized that greater nuance and a variety of factors should be considered when posters give advice, rather than using baseless guarantees. P1 (Female, 22) described TP as:

Generalizing everyone to having problems that can be solved and thinking that it’s a one-dimensional or a one-way thing for everyone and that it will just work. . . [It is] toxic because it ignores the fact that bad things do happen, and they happen differently for everyone, and they are very real.

Thus, oversimplifying life and the problems that one may face was perceived as toxic to participants, even when the overall message of the post is a positive one meant for encouragement.

Perceptions of TP seemed to form if posts give advice and guarantee a positive outcome every time, which most participants do not believe in. Referring to the post in Figure 1 Panel 3 stating that hard work leads to success, P20 (Female, 22) said that it was highly toxic:

I think there are a lot of people in this world that work hard, but they are not considered as conventionally successful. I think a lot of success comes from having luck and also having money. So this post just seems to treat being successful as something very easy. Like you just work hard, and then you can do it. Which I guess might be motivational for some people to get off their ass and start working hard. But for me, I disagree.

The post that P20 brought as an example of a TP post also made overgeneralizations in a similar way (Figure 2 Panel 5). It stated, “Sometimes you gotta laugh through the tears, smile through the pain so that you can live through the sorrows,” offering an overly simplified aphorism toward the vicissitudes of life.

Additionally, P17 (Female, 21) brought to the focus group a post that said: “The secret to happiness is letting every situation be what it is instead of what you think it should be” (Figure 2 Panel 6). P15 (Female, 22) agreed that the post was toxic because of its overgeneralization, asking rhetorically: “How can you [reduce] the secret to happiness to become a one-size-fits-all approach, or something that can be summarized in a few lines?”

In this way, the theme of overgeneralization has some overlap with the theme of perceived privilege. But whereas the former stems from posters trying to give (unreasonably) universal advice, the latter stems from posters being unaware that they seem to be talking down to others from a privileged position. Overgeneralizations also show the extent to which posters may have been interpellated into holding unquestioned beliefs on neoliberal self-help ideology: Posters who propagate TP seem to assume that their successes and life experiences are generalizable to all other people, neglecting to consider that audiences may perceive their posts as toxic.

Commanding words. Besides overgeneralizations, commanding words were another feature of TP posts mentioned by our participants. Posts containing commanding words that give directives to audiences can come off as producing TP. P7 (Female, 22) explained:

If you keep repeatedly seeing “think happy thoughts,” “stay positive,” “love yourself,” it may start sounding like a command because you start drilling yourself to do so—like I need to love myself, I need to think happy thoughts, I need to stay positive. Especially if it’s like a command for you to do something, then it might sound toxic-positive, and I think a better way could be acknowledging what you have already done rather than emphasizing what you can do.

When discussing a post advising audiences to be grateful if they were obstructed from their goals because it was not meant for them (Figure 2 Panel 7), P19 (Female, 24) said:

... if you just got rejected from a role that you really want, and then you scroll and you see this, you feel that it doesn’t value-add because it may be encouraging us or guiding us to feel a certain way that we weren’t feeling at the point in time.

Similarly, P18 (Male, 24) said it would be better if posts did not make use of commanding words, such as those found in the post P18 brought to the focus group (Figure 2 Panel 8), which stated: “Always trust the timing. Everything happens for a reason.” P18 said that true encouragement should not sound commanding, that is, we need “the kind of encouragement that isn’t telling people upfront to go and do things, but encouraging them to continue to put in the effort.”

Commanding words are a good example of audiences’ negative face being threatened. Posters who adopt a forceful and demanding tone, telling audiences to adopt particular ways of thinking or to enact particular behaviors, are negatively perceived because commanding words threaten audiences’ autonomy. This leads to TP as audiences perceive these posters to be overbearing.

Discussion

Results show that TP is associated with the following factors: overgeneralization, commanding words, ignoring negative emotions, and perceived privilege. The first two factors are content-based characteristics while the third and fourth factors are psychological perceptions. Separately, a medium-based affordance—ephemerality—attenuates the negative effects of TP.

In terms of content-based characteristics, overgeneralizations and commanding words have been found to constitute TP posts. They potentially threaten audiences' negative face, as they appear to constrain audiences' autonomy when accompanied by strong advice. But more than that, these content-based characteristics feed into the neoliberal self-help ideology, with the idea that everyone must help themselves by being positive. This is aligned with Orgad and Gill's (2022) commentary on the neoliberal belief that people have "the ability to *choose* happiness over unhappiness" (italics in original, p. 13), which enables the self-help industry to monetize positive messages by selling books, posters, coaching sessions, courses, and workshops that optimistically guarantee the success of a self-help lifestyle. Orgad and Gill (2022) observed that happiness has even become an international competition, with several global happiness indices comparing the happiness levels of various countries. And it is within this system that overgeneralizations (e.g. "life will *never* give you more than you can handle") and commanding words (e.g. "*be grateful* for what you learned"; see Goodman, 2022) interpellate audiences to adopt, for themselves, an almost-mandatory belief in positive thinking—which we know, based on the focus group discussions, can be experienced as TP and not genuine positivity.

Two perceptual factors were frequently mentioned by participants as being TP-related: perceived ignoring of negative emotions and perceived privilege of a poster. Specifically, audiences are upset if they feel that a post encourages them to ignore their negative emotions in favor of exclusively positive emotions, or if they feel that the poster sent a positive message from a highly privileged position.¹ At the most basic level, meaning can indeed depend on content—for audiences who have internalized positive thinking ideology. For example, the message in Figure 1 Panel 3, "Successful people are not gifted; they just work hard, then succeed on purpose," can be (superficially) understood as a motivational message. But for some of our participants, meaning is not found in the words themselves, but in their interpretation. P2 (Male, 24) and P9 (Female, 25), among others, interpreted the message in Figure 1 Panel 3 not as a motivational message, but as one that reflects the poster's privilege. Thus, the message remains the same, but the interpretation changes depending on the audience.

In addition, the idea of perceived privilege is interesting because audiences may make interpretations of what a poster intended. For example, P10 (Female, 22) mentioned that the poster of Figure 1 Panel 10 ("You don't achieve greatness with casual effort, challenge yourself and go all in.") probably had good intentions, in that the poster probably wanted to motivate people. Hence, this post is a form of TP in part because the poster was judged to have good intentions. If, for example, the poster was judged to be trolling, then the post does not represent TP but, instead, some form of satire. Thus, meaning is intersubjective: Meaning is not found in the literal content of the post and not *exclusively*

found in audiences' interpretation of the post—but also in the relationship between the audience and the poster. The post therefore functions like the letter in Lacan's seminar on *The Purloined Letter* (Lacan, 1972), in which a message is important for how it structures the relationships between people and not for its content.

On social media, messages can be ephemeral (e.g. disappear after 24 hours, like Instagram Stories) or persistent (i.e. do not automatically disappear, like posts). Some participants associated TP more with persistent posts than with ephemeral posts, as they understood that ephemeral media are for things that are casual or mundane (see Xu et al., 2016). In other words, even when a message reflects TP, people are not likely to take it too seriously if it were also ephemeral. Therefore, ephemerality can be understood as a medium characteristic that influences perceptions of TP, alongside the content-based characteristics and perceptual factors mentioned previously.

However, Study 1 was not without its biases. Our advertisements clearly stated that the focus group discussions were about TP on social media. Participants were even instructed to bring their own examples of TP posts for discussion. This made sense given our research questions, but we cannot rule out the possibility that there are people who actually like positive messages that our focus group participants would consider as TP. But the people who like such positive messages were unlikely to have answered our advertisements for our focus groups.

Rationale for Study 2

Given this potential bias, we planned Study 2, which was an experiment that allowed us to deductively test our inductively-obtained findings from Study 1. Whereas the focus groups allowed us to have depth in our analyses of TP, the experiment—by dint of a larger sample—allowed us to potentially make more generalized assertions regarding TP.

Study 1 found that TP posts were associated with two message characteristics: commanding words and overgeneralization. They refer not to people's perceptions, but to concrete elements that make up a post. An example of commanding words is the instruction to "be thankful" (Figure 2 Panel 7) and an example of overgeneralization is the phrase "*everything* happens for a reason" (Figure 2 Panel 8, italics added). As such, we wrote commanding words and overgeneralizations into our experimental stimuli (fictitious Instagram posts/stories; see the Appendix). These message characteristics can therefore be regarded as the starting point for subsequent perceptions related to TP.

Study 1 also found that TP posts were associated with two perceptual variables: ignoring negativity and perceived privilege. These perceptual variables are not part of concrete message characteristics (i.e. the words that comprise a post), but rather, are a result of people's perceptions of those message characteristics. Put differently, commanding words and overgeneralizations in a post elicit in people the perception that the post is ignoring negativity and the poster is privileged.

Therefore, if we consider the extent to which people like a post as the dependent variable, then post positivity can be understood as the independent variable that we experimentally manipulate by introducing commanding words and overgeneralizations into the

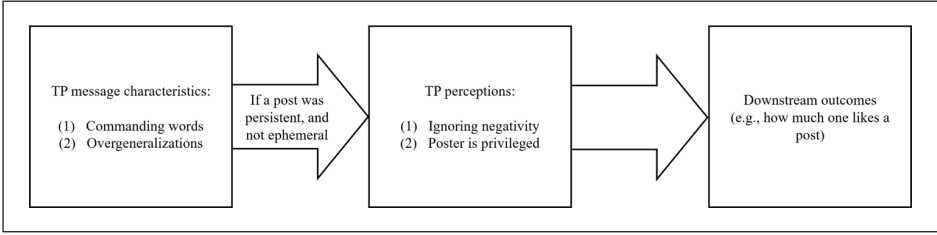


Figure 3. Relationships between TP variables identified in Study 1.

stimuli. The perceptual variables of ignoring negativity and perceived privilege then function as mediators, in that they explain why posts containing commanding words and overgeneralizations are less liked than posts that do not.

Ephemerality, which Study 1 found to attenuate TP, was tested as a moderator. More precisely, commanding words and overgeneralizations should foster a sense of ignoring negativity and a perception of poster privilege only if a post was persistent. If the post was ephemeral, then audiences should interpret the post as something that is trivial and not meant to be taken seriously (see Bayer et al., 2016)—and therefore audiences would not experience the pressure to ignore negativity or perceive that a poster is privileged.

The relationships described here are visualized in Figure 3.

Study 2

Study 2 was a 3 (positivity: high positivity with toxic message characteristics/high positivity/low positivity) × 2 (ephemerality: persistent/ephemeral posts) between-subjects experiment. In other words, the stimuli—fictitious Instagram screenshots that we created—had six possible variations: (1) ephemeral posts with TPMC,² (2) permanent posts with TPMC, (3) ephemeral High Positivity posts, (4) permanent High Positivity posts, (5) ephemeral Low Positivity posts, and (6) permanent Low Positivity posts. Each participant was randomly shown only one of the six conditions. See the Appendix for the full stimuli.

Based on Study 1’s findings, commanding words and overgeneralization (content-based TP characteristics that participants frequently mentioned) were used to create the TPMC posts. The High Positivity and Low Positivity posts were created without these characteristics, and differed in terms of tone, with the Low Positivity posts having the most neutral and matter-of-fact tone. As many participants in Study 1 brought motivational quotes as their examples of TPMC, we decided to create similar motivational quotes as the stimuli for Study 2. Liking was chosen as a direct measure of whether participants felt any negative emotions toward the posts they see. If participants viewing TPMC posts have less liking for the posts than participants viewing High Positivity or Low Positivity posts, then there is evidence for TP (i.e. the phenomenon that superficially positive messages elicit negative responses):

H1: TPMC posts engender less liking than High Positivity and Low Positivity posts.

To test the perceptual mechanisms associated with TP that were identified in Study 1:

H2: The relationship between positivity and liking (described in H1) is mediated by (a) perceived ignoring of negative emotions and (b) perceived privilege of the poster.

In other words, H2 predicts that TPMC posts—when compared with High Positivity or Low Positivity posts—engenders greater perceived ignoring of negative emotions and greater perceived privilege of the poster, which then leads to less liking for the posts.

As Study 1 found, the ephemerality of a post can influence people's perceptions of the seriousness of the post. Ephemeral posts are more associated with the mundane and the trivial, and therefore ephemerality may produce a discounting effect on audiences' TP-related perceptions. Hence:

H3: The ephemerality of a post moderates the mediation relationship (described in H2), such that ephemeral posts engender less (a) perceived ignoring of negative emotions and less (b) perceived privilege of the poster than permanent posts.

Method

Sample. Participants (59.5% female, 37.1% male, 0.9% non-binary, 2.6% prefer not to say) were students at Nanyang Technological University. They were recruited through university email lists and text messaging channels. Participants were between 18 and 29 years old ($M = 21.85$, $SD = 1.82$). Each participant was paid \$5 in the local currency upon questionnaire completion. Assuming a small-to-medium effect size, a minimum sample size of 318 participants were required to obtain 80% power. A total of $N = 348$ responses were used for analyses (10 responses that failed an attention check were excluded).

Procedure. Participants were first asked to affirm that they use social media at least once a week—those who did not were not allowed to participate. Next, they provided informed consent. The experimental stimuli were then presented. Participants were randomly assigned to one of the six conditions (see Appendix). A hidden timer prevented participants from progressing to the next page for 15s to ensure they read the stimuli. Three screenshots of posts belonging to the same condition were shown in order to reduce the chances of participants obtaining message-specific idiosyncratic effects. After viewing the stimuli, participants completed a post-test questionnaire. The experiment concluded with demographic questions.

Measures

Attention check. Participants were asked to identify the type of posts they saw through a multiple-choice question, choosing either “Instagram Stories” or “Instagram Posts.” Participants who failed this check were excluded from analyses.

Manipulation checks. The manipulation check for objective positivity read: “Rate how objectively negative/neutral/positive you think the words in these posts are (disregarding how much you like/dislike the posts)” (1 = *Very negative*, 7 = *Very positive*). The manipulation check for ephemerality was a seven-point semantic differential scale, where participants rated whether the posts they saw were “*Only temporarily accessible*–*Accessible to audiences for a long time*.” To ensure the stimuli created were in line with our desired manipulations of types of positivity, we used Study 1’s results—use of commanding words and overgeneralization—to create checks for the experimental conditions. Use of commanding words was measured using a 5-point scale (1 = *Not at all*, 5 = *Extremely*), where participants rated how “commanding,” “assertive,” and “forceful” the messages in the stimuli were (Cronbach’s $\alpha = .82$). Overgeneralization was measured using a 7-point scale (1 = *Strongly disagree*, 7 = *Strongly agree*), where participants indicated how much they agreed with the statements “These messages oversimplify reality,” “These messages guarantee outcomes without basis,” and “These messages contain an unfounded sense of certainty” (Cronbach’s $\alpha = .69$).

Ignoring negativity. Participants were instructed to look at the three posts and “answer all the questions in this section according to your perception of them collectively.” Participants completed items adapted from the Multidimensional Experiential Avoidance Questionnaire (MEAQ; Gámez et al., 2011), specifically the “distraction and suppression” and “repression and denial” subscales, as well as items adapted from the Perceived Invalidation of Emotion Scale (PIES; Zielinski and Veilleux, 2018). Participants rated the items from 1 (*Strongly disagree*) to 7 (*Strongly agree*). Seven items were adapted from the distraction and suppression subscale, such as “The messages are trying to distract me from feeling something negative” and “The messages are trying to encourage me to put negative emotions out of my mind.” 13 items were adapted from the repression and denial subscale, such as “The messages encourage me not to own up to my problems” and “The messages encourage me not to be aware of my problems.” The 10 items adapted from PIES include, “The messages don’t take me seriously with regard to how I’m feeling” and “The messages make me feel that my emotions are unimportant.” All 30 items from the MEAQ and PIES were combined to form a composite scale ($\alpha = .93$). See the Online Supplemental Materials for the exact phrasing of all adapted items.

Perceived privilege. Perceived privilege of the poster was an original 7-point scale (1 = *Strongly disagree*, 7 = *Strongly agree*). Participants rated four items: “The message sender writes from a position of privilege,” “The message sender perceives themselves to be better than the average person,” “The message sender has different life experiences from the typical audience,” and “The message sender is engaging in humblebragging” (Cronbach’s $\alpha = .75$). The definition of humblebragging was provided as “making a statement intended to highlight positive qualities while attempting to appear humble, by disguising it as an apology or complaint.”

Liking. An original seven-point semantic differential scale was used to rate liking toward the posts shown, assessed collectively. It had three items: *Strongly dislike*–*Strongly like*, *Strongly disapprove*–*Strongly approve*, and *Strongly against*–*Strongly in favor* (Cronbach’s $\alpha = .89$).

Results

Manipulation checks

Objective positivity. A one-way analysis of variance (ANOVA) tested whether participants perceived varying levels of objective positivity across the TPMC, High Positivity, and Low Positivity conditions, $F(2, 345) = 33.7, p < .001$. The Low Positivity condition showed the lowest rated objective positivity ($M_{LP} = 4.74, SD_{LP} = 1.33$) compared with the High Positivity condition ($M_{HP} = 5.70, SD_{HP} = 0.92$) and the TPMC condition ($M_{TPMC} = 5.75, SD_{TPMC} = 0.84$). Bonferroni post hoc tests affirmed that the Low Positivity condition was significantly different from the High Positivity ($p < .001$) and TPMC conditions ($p < .001$), while the High Positivity and TPMC conditions were not statistically different from each other ($p = 1.000$). The non-significant difference in objective positivity between High Positivity and TPMC did not detrimentally affect the testing of hypotheses (and was in fact desirable), as TPMC was differentiated through the use of commanding words and perceived overgeneralization, as reported below. In sum, participants rated the Low Positivity condition as containing words lower in positivity at a literal level than both the High Positivity and TPMC conditions.

Commanding words. A one-way ANOVA tested whether participants perceived varying levels of the use of commanding words across the TPMC, High Positivity, and Low Positivity conditions, $F(2, 345) = 6.37, p = .002$. The TPMC condition showed the highest rated usage of commanding words ($M_{TPMC} = 2.18, SD_{TPMC} = 0.99$) compared with the High Positivity condition ($M_{HP} = 1.84, SD_{HP} = 0.80$) and the Low Positivity condition ($M_{LP} = 1.84, SD_{LP} = 0.72$). Bonferroni post hoc tests showed that the TPMC condition was significantly different from the High Positivity ($p = .006$) and Low Positivity conditions ($p = .007$), while the High Positivity and Low Positivity conditions were not statistically different from each other ($p = 1.000$). Taken together, participants rated the TPMC condition as containing more commanding words than both the High Positivity and Low Positivity conditions.

Overgeneralization. A one-way ANOVA tested whether participants perceived varying levels of overgeneralization across the TPMC, High Positivity, and Low Positivity conditions, $F(2, 345) = 13.31, p < .001$. The TPMC condition had the highest perceived overgeneralization ($M_{TPMC} = 4.95, SD_{TPMC} = 1.10$) compared with the High Positivity condition ($M_{HP} = 4.38, SD_{HP} = 1.08$) and the Low Positivity condition ($M_{LP} = 4.22, SD_{LP} = 1.22$). Bonferroni post hoc tests showed that the TPMC condition was significantly different from the High Positivity ($p < .001$) and the Low Positivity conditions ($p < .001$), while the difference between the High Positivity and Low Positivity conditions was not statistically significant ($p = .822$). Therefore, participants rated the TPMC condition as containing more overgeneralizations than both the High Positivity and Low Positivity conditions.

Ephemerality. A one-way ANOVA tested whether participants perceived varying levels of post accessibility across the ephemeral and permanent conditions. Participants perceived that posts in the permanent condition were accessible for a longer time

Table 1. Total effects of positivity on liking.

Comparison	LP-TPMC	HP-TPMC
<i>b</i>	-.23	.35
<i>SE</i>	.15	.15
<i>t</i>	-1.47	2.31
<i>p</i>	.14	.02

($M_{\text{permanent}} = 6.13$, $SD_{\text{permanent}} = 1.12$) than posts in the ephemeral condition ($M_{\text{ephemeral}} = 2.76$, $SD_{\text{ephemeral}} = 1.87$), $F(1, 346) = 404.80$, $p < .001$. The experimental manipulation for ephemerality worked as intended.

Simple mediation analyses

Two simple mediation (i.e. without the moderator of ephemerality) analyses were conducted using Hayes' (2022) PROCESS Model 4, which generated 5,000 bootstrapped samples. PROCESS was used to create dummy codes, with the TPMC condition as the reference category. The dummy coding procedure generated two comparisons: the first was the difference between the Low Positivity condition and the TPMC condition (notated as LP-TPMC) and the second was the difference between the High Positivity condition and the TPMC condition (notated as HP-TPMC). These two comparisons were entered as predictors into the models described below. All continuous predictors were mean-centered.

Total effects of positivity. Total effects of the LP-TPMC and HP-TPMC comparisons are shown in Table 1. The HP-TPMC comparison was significant, but the LP-TPMC comparison was not. These results indicate High Positivity messages engendered greater liking than TPMC messages, but Low Positivity messages did not. Hence, H1 was only partially supported.

Ignoring negativity. The results of the mediation analysis—with ignoring negativity as the mediator—is shown in Figure 4. The overall model predicting ignoring negativity was significant, $R^2 = .035$, $F(2, 345) = 6.25$, $p = .002$. Compared with TPMC posts, Low Positivity posts and High Positivity posts both engendered lesser sentiment of ignoring negativity. The path from ignoring negativity to liking was significant, $R^2 = .091$, $F(3, 344) = 11.50$, $p < .001$. See Table 2 for the indirect effects. Results support H2(a): TPMC posts elicited a greater sense of ignoring negativity than Low Positivity or High Positivity posts, and this greater sense of ignoring negativity led to less liking toward the TPMC posts.

Perceived privilege. The results of the mediation analysis with perceived privilege as the mediator is shown in Figure 5. The overall model predicting perceived privilege was significant, $R^2 = .032$, $F(2, 345) = 5.60$, $p = .004$. Compared with TPMC posts, both Low Positivity posts and High Positivity posts engendered lesser perceived privilege.

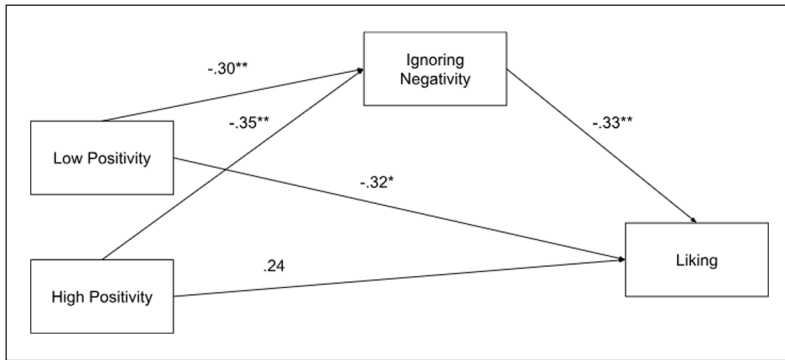


Figure 4. Mediation model for ignoring negativity as the mediator between positivity and liking.

** $p < .01$, * $p < .05$. Reference condition: TPMC.

Table 2. Indirect effects of positivity.

Mediator	Ignoring Negativity		Perceived Privilege	
	LP-TPMC	HP-TPMC	LP-TPMC	HP-TPMC
<i>b</i>	.10	.12	.10	.10
<i>SE</i> _{bootstrapped}	.04	.05	.04	.04
<i>CI</i> _{bootstrapped}	[.03, .19]	[.04, .21]	[.03, .19]	[.03, .19]

The path from perceived privilege to liking was significant, $R^2 = .089$, $F(3, 344) = 11.24$, $p < .001$. See Table 2 for the indirect effects. Results support H2(b): TPMC posts elicited a greater perception that the poster is privileged than Low Positivity or High Positivity posts, and this greater perceived privilege led to less liking toward the TPMC posts.

Moderated mediation analyses

Two moderated mediation analyses were performed using Hayes’ (2022) PROCESS Model 7 (5,000 bootstrapped samples). Like before, positivity was the independent variable and liking toward the posts was the dependent variable—but now ephemerality was added to the model as a moderator that influenced the relationship between the independent variable and the mediator. The first moderated mediation analysis modeled ignoring negativity as the mediator, and the second modeled perceived privilege as the mediator.

The overall model predicting ignoring negativity and moderated by ephemerality was significant, $R^2 = .044$, $F(5, 342) = 3.16$, $p = .008$. See Figure 6. Indirect effects are presented in Table 3. Ephemerality did not moderate the effect of positivity on liking that was mediated through ignoring negativity. When considering permanent and ephemeral

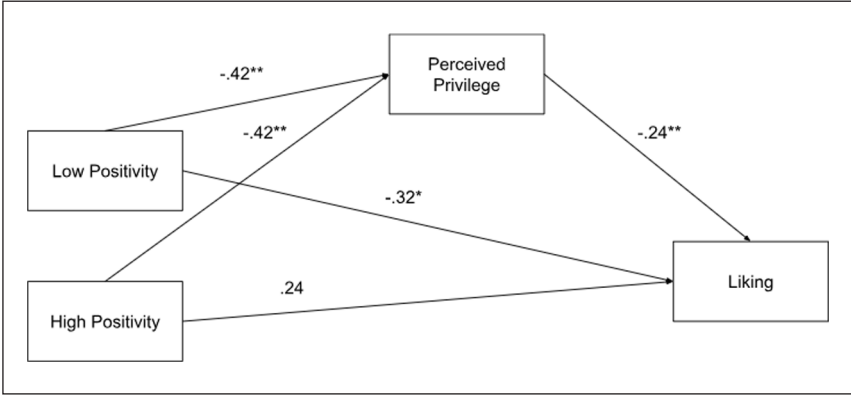


Figure 5. Mediation model for perceived privilege as the mediator between positivity and liking.

** $p < .01$, * $p < .05$. Reference condition: TPMC.

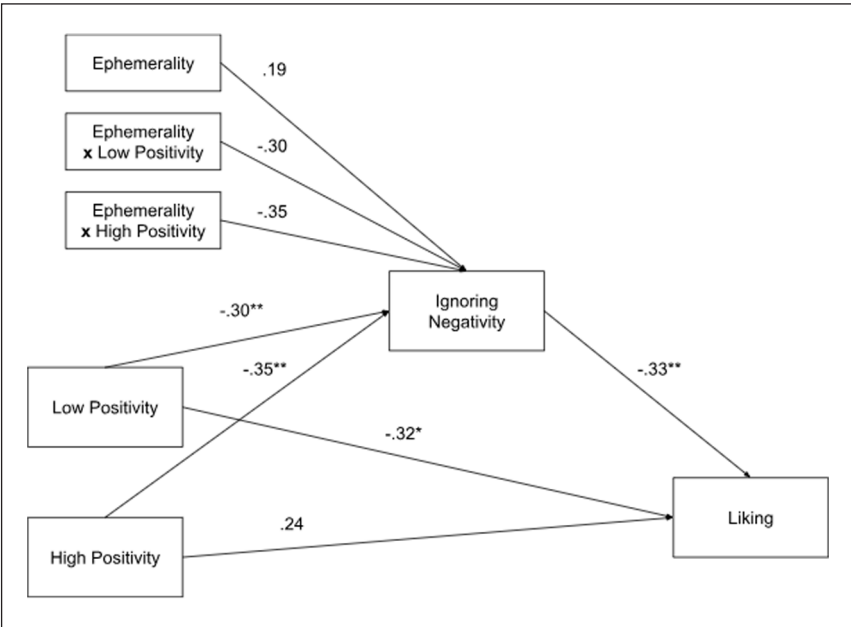


Figure 6. Mediated moderation model for ignoring negativity as the mediator between positivity and liking, and ephemerality as the moderator.

** $p < .01$, * $p < .05$. Reference condition: TPMC.

posts in isolation from each other, only permanent posts produced a significant mediation effect between positivity and liking through ignoring negativity for both the LP-TPMC and HP-TPMC comparisons.

Table 3. Indirect effects of positivity on liking with ignoring negativity as mediator.

Comparison	LP-TPMC			HP-TPMC		
	Ephemeral	Permanent	Index of Moderated Mediation	Ephemeral	Permanent	Index of Moderated Mediation
<i>b</i>	.05	.15	.10	.06	.17	.12
<i>SE</i> _{bootstrapped}	.05	.06	.05	.05	.07	.08
<i>CI</i> _{bootstrapped}	[-.04, .15]	[.04, .29]	[-.04, .26]	[-.03, .16]	[.05, .33]	[-.02, .29]

The overall model predicting perceived privilege and moderated by ephemerality was also significant, $R^2 = .035$, $F(5, 342) = 2.47$, $p = .03$. See Figure 7. Indirect effects of the LP-TPMC and HP-TPMC comparison are presented in Table 4. Ephemerality did not moderate the effect of positivity on liking that was mediated through perceived privilege. Like before, when considering permanent and ephemeral posts in isolation from each other, only permanent posts produced a significant mediation effect between positivity and liking through perceived privilege for both the LP-TPMC and HP-TPMC comparisons.

Therefore, the moderated mediation models were not statistically significant; H3 was unsupported.

Discussion

Study 2 aimed to deductively test the findings of Study 1. The HP-TPMC comparison showed a significant total effect on liking, but not the LP-TPMC comparison, resulting in a partial support for H1. Considering that there were no significant differences in reported objective positivity between High Positivity and TPMC posts, these findings suggest that effects were driven by characteristics unique to TPMC posts (but were not in the Low Positivity or High Positivity posts): overgeneralization and commanding words. Put differently, although the TPMC posts were rated as more positive in a literal sense than the Low Positivity posts, both TPMC and Low Positivity posts appear to be equally disliked when compared with High Positivity posts. Therefore, merely displaying positivity is insufficient to actually engender positivity or emotional contagion.

Ignoring negativity. In the simple mediation analysis, ignoring negativity mediated the relationship between positivity and liking, for both the LP-TPMC and HP-TPMC comparisons, supporting H2(a). Both Low Positivity and High Positivity posts engendered lower perceptions of ignoring negativity as compared with TPMC posts. This result corroborates Study 1's findings regarding the characteristics of TP (i.e. overgeneralization and commanding words) and the mechanism of ignoring negativity. The greater the perception of these characteristics in a positive post, the greater the perception that the post encourages ignoring negativity. This result complements existing literature regarding the negative effects of emotional suppression (e.g. English and John, 2013), which showed

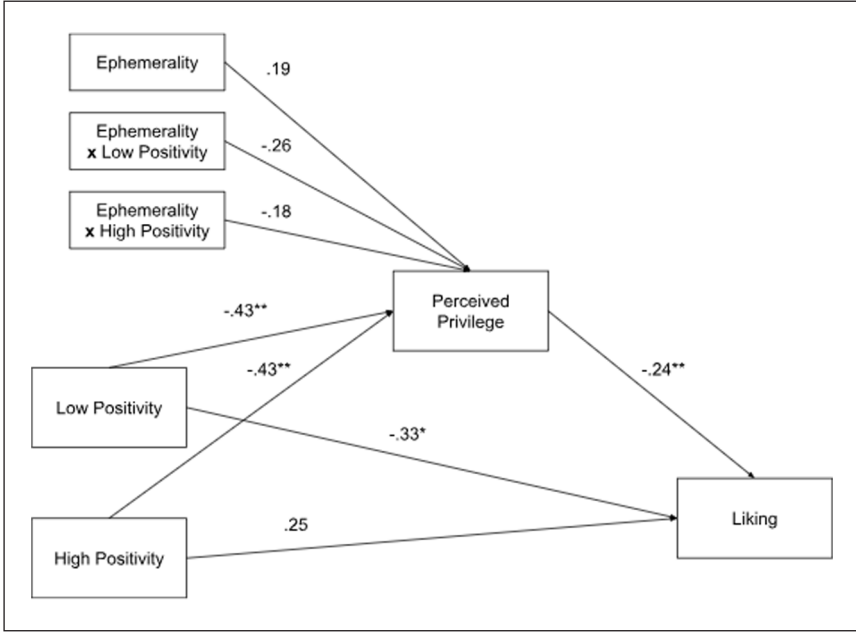


Figure 7. Mediated moderation model for perceived privilege as the mediator between positivity and liking, and ephemerality as the moderator.
 ** $p < .01$, * $p < .05$. Reference condition: TPMC.

Table 4. Indirect effects of positivity on liking with perceived privilege as mediator.

Comparison	LP-TPMC			HP-TPMC		
	Ephemeral	Permanent	Index of Moderated Mediation	Ephemeral	Permanent	Index of Moderated Mediation
<i>b</i>	.07	.13	.06	.08	.12	.04
<i>SE</i> _{bootstrapped}	.05	.06	.07	.05	.06	.07
<i>CI</i> _{bootstrapped}	[-.03, .18]	[.04, .26]	[-.07, .22]	[-.02, .19]	[.02, .26]	[-.09, .20]

that suppressing negative emotions is an ineffective way to regulate emotions that results in poorer social functioning.

Perceived privilege. Perceived privilege also mediated the relationship between positivity and liking, in both the LP-TPMC and HP-TPMC comparisons, supporting H2(b). TPMC posts engendered perceptions of greater poster privilege than Low Positivity or High Positivity posts. This is in line with previous literature on positive/negative face,

particularly for audiences who might perceive advice as a threat toward autonomy or that advice-givers perceive them as less competent (Brown and Levinson, 1987). This finding also highlights the importance of understanding TP through the lenses of intersubjectivity, or how a message structures the relationship between interactants. As part of their experience of TP, audiences infer that the poster is privileged, which aligns with the results of Study 1. Above and beyond the denotative and connotative meanings of a post, the assumed disparity in the relationship between a poster and an audience in terms of privilege is essential to understanding TP.

Ephemerality. For both the LP-TPMC and HP-TPMC comparisons, no moderated mediation effects were found, as the confidence intervals for the index of moderated mediation included zero (see Table 3 and Table 4). Ephemerality moderated neither the effect of positivity on ignoring negativity nor the effect of positivity on perceived privilege. Speculatively, it is possible that we did not find statistically significant results for ephemerality because participants viewed screenshots of Instagram posts (persistent) or Instagram Stories (ephemeral) on a web survey platform, rather than on Instagram itself. Future experiments dealing with ephemerality may consider more naturalistic treatments to avoid this issue.

Furthermore, the discounting effect of ephemerality may apply directly to the outcome of liking, instead of working through the mediators (as we tested). Is it possible that persistent TP posts engender less liking than Low Positivity or High Positivity posts—but ephemeral TP posts do not? This is commensurate with the idea that ephemerality mitigates the negative effects of TP because people infer that posters intend for ephemeral posts to be trivial. However, we did not measure whether our participants made this inference. Future research can test ephemerality with better measures or as a moderator between post positivity and liking (as opposed to being a part of moderated mediation analyses).

General discussion

Given the widespread positivity bias on social media (Waterloo et al., 2018) and the greater virality of positive content (Berger and Milkman, 2012), content creators should be more likely to post positive content than neutral or negative content. This behavior is plausibly driven by digital emotional contagion, which asserts that sharing positive content online results in a transferal of positive emotions from posters to audiences (Kramer et al., 2014). However, there is growing awareness that there is a limit to the contagion of positive emotions: TP (Goodman, 2022).

Across two studies, the findings suggest that content-based factors—namely overgeneralization and commanding words—engender TP through the mechanisms of ignoring negativity and perceived privilege. Collectively, they show how the process of digital emotional contagion fails, and positivity is not transferred from posters to audiences. Put differently, TP can be understood as a failure of the mechanisms theorized to engender emotional contagion (Goldenberg and Gross, 2020): there is a lack of mimicry, positive

emotion categories are not activated, and there is no social appraisal—in that audiences do not use the posters' emotions as a guide as to how they themselves should feel. Positivity—when accompanied by overgeneralization and the use of commanding words—can, instead, backfire and generate negative emotions in audiences.

From a cultural perspective, TP is more than just another case of a message having multiple possible interpretations or audiences attributing different intentions to a message sender. The self-help culture so prevalent in contemporary social media can be linked to neoliberalism, where “a market ethic works to reconstitute subjectivities, calling into being subjects who are self-motivating and entrepreneurial, who will make sense of their lives through discourses of freedom, responsibility, and choice” (Orgad and Gill, 2022: 16). Amid such an ideology, where successes are unquestioningly accepted as the outcome of one's own efforts in a meritocratic fashion, there is a cultural movement toward choosing to be positive—as if the act of exposing oneself to positive thinking messages can indeed make one feel more positive. We argue, and demonstrate through our studies, that it leads to TP.

Future research can consider analyzing TP through the lens of social defenses (Krantz, 2010; Menzies, 1960). According to Menzies (1960), social defenses are the methods of thinking or working that insulate people from their own anxieties, even though anxiety reduction is not the express aim of those methods. Menzies (1960) provides the example that healthcare workers may use professional detachment and ritualistic task performance to reduce their own anxieties regarding the burden of care for patients, including the terminally ill. Where TP is concerned, we cannot rule out the possibility that TP posts are posters' social defenses against their own anxieties. They may be a form of self-talk (Hardy, 2006) for posters to tell themselves they are optimistic, successful people—which entrenches the neoliberal notion that people should present themselves as capable of self-improvement by being positive. Even audiences may have their own anxieties, and when faced with posters who seem to have succeeded, they interpret posters as acting superior (especially in relation to the theme of perceived privilege), and therefore perceive certain posts to be toxic. In this sense, future research can explore how both the creation and interpretation of TP posts on social media are people's social defenses against their own anxieties.

Limitations

Although the results of Study 1 and Study 2 largely corroborated with each other, some limitations have to be addressed. One limitation is the focus on motivational posts in both studies. Although positive motivational posts that engender negative outcomes represent a substantial aspect of online TP, online TP is more than motivational posts. Future studies can focus on a different context, to examine whether posts other than those that are motivational can be linked to TP.

Several limitations pertain to the manipulations in Study 2. First, Study 1 did not ask participants for their views—beyond the most immediate psychological factors of ignoring negativity and perceived privilege—on the more distal outcomes of TP. Put

differently, Study 1 did not shed light on any specific outcome that could be tested in Study 2. Liking was therefore chosen as a broad measure of the effects of TP in Study 2. The use of liking as the outcome variable may not have captured the full effects of TP in the model. Future studies can address this issue by investigating other plausible distal outcomes of TP.

Additionally, Study 2 simultaneously manipulated both overgeneralization and commanding words. It remains to be seen whether either of the two factors is sufficient to elicit TP effects. Future studies can therefore consider manipulating either overgeneralization or commanding words, separately.

Another important point to note is that Study 2 was conducted experimentally through online surveys, which might elicit different reactions from participants as compared with real-world instances of TP. In real-world settings, audiences are likely to come across TP posts on their feed in otherwise mundane circumstances, which might contribute to feelings of TP if these posts were seen as unexpected and “intrusive” in some form. Future studies may choose to investigate this TP phenomenon in a more natural setting that mimics the real-world.

As scholars think about how social media can contribute toward well-being or a meaningful life, the present research makes it clear that merely displaying positivity is not an answer. The contagion of positive emotions may work under some circumstances, but we should recognize that inauthentic positivity can result in TP—harming, rather than uplifting, audiences.

Data availability

Study materials will be shared upon reasonable request made to the corresponding author.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded by the Singapore Ministry of Education AcRF Tier 1 Seed Grant (RS09/22) awarded to Zijian Lew

ORCID iD

Zijian Lew  <https://orcid.org/0000-0003-1769-7898>

Supplemental material

Supplemental material for this article is available online.

Notes

1. By “privilege,” we do not mean to say that a poster has advantages that they did not deserve or earn. Rather, we use the term “privilege” as a broad label to refer to any advantage (wealth, power, social standing, conventional markers of success, etc.) that a poster may appear to have and that an audience member may not have.
2. Considering that TP (according to Study 1) is rooted in both (a) message characteristics and (b) the perceptions those characteristics engender in people, we use the label “TPMC” here to refer to only (a).

References

- Althusser L (1971) *Lenin and philosophy and other essays* (B Brewster, Trans.). Monthly Review Press.
- Bayer JB, Ellison NB, Schoenebeck SY and Falk EB (2016) Sharing the small moments: Ephemeral social interaction on Snapchat. *Information, Communication & Society* 19(7): 956–977. <https://doi.org/10.1080/1369118X.2015.1084349>
- Becker JC, Hartwich L and Haslam SA (2021) Neoliberalism can reduce well-being by promoting a sense of social disconnection, competition, and loneliness. *British Journal of Social Psychology* 60(3): 947–965. <https://doi.org/10.1111/bjso.12438>
- Berger J and Milkman KL (2012) What makes online content viral? *Journal of Marketing Research* 49(2): 192–205. <https://doi.org/10.1509/jmr.10.0353>
- Brown P and Levinson SC (1987) *Politeness: Some universals in language usage*. Cambridge University Press.
- English T and John OP (2013) Understanding the social effects of emotion regulation: The mediating role of authenticity for individual differences in suppression. *Emotion* 13(2): 314–329. <https://doi.org/10.1037/a0029847>
- Fawehinmi Y (2024, February 19) What is TikTok's new 'hopecore' trend and is it exactly what we need right now? *The Independent*. <https://www.the-independent.com/life-style/tiktok-trend-hopecore-meaning-positivity-b2498452.html>
- Gámez W, Chmielewski M, Kotov R, Ruggero C and Watson D (2011) Development of a measure of experiential avoidance: The multidimensional experiential avoidance questionnaire. *Psychological Assessment* 23(3): 692–713. <https://doi.org/10.1037/a0023242>
- Goldenberg A and Gross JJ (2020) Digital emotion contagion. *Trends in Cognitive Sciences* 24(4): 316–328. <https://doi.org/10.1016/j.tics.2020.01.009>
- Goodman W (2022) *Toxic Positivity: Keeping it real in a world obsessed with being happy*. TarcherPerigee.
- Gross JJ and John OP (2003) Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology* 85(2): 348–362. <https://doi.org/10.1037/0022-3514.85.2.348>
- Gross JJ and Levenson RW (1993) Emotional suppression: Physiology, self-report, and expressive behavior. *Journal of Personality and Social Psychology* 64(6): 970–986. <https://doi.org/10.1037//0022-3514.64.6.970>
- Hancock JT, Gee K, Ciaccio K and Lin JMH (2008) I'm sad you're sad: emotional contagion in CMC. *Proceedings of the 2008 ACM conference on Computer supported cooperative work* 295–298. <https://doi.org/10.1145/1460563.1460611>
- Hardy J (2006) Speaking clearly: A critical review of the self-talk literature. *Psychology of Sport and Exercise* 7(1): 81–97. <https://doi.org/10.1016/j.psychsport.2005.04.002>
- Hastings SO and Bell GC (2018) Facing our heuristic limits: Expanding the terminology for types of positive face. *Communication Quarterly* 66(1): 96–110. <https://doi.org/10.1080/01463373.2017.1340317>
- Hayes AF (2022) *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (3rd ed.). The Guilford Press.
- Kramer AD, Guillory JE and Hancock JT (2014) Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences* 111(24): 8788–8790. <https://doi.org/10.1073/pnas.1320040111>
- Krantz J (2010) Social defenses and twenty-first century organizations. *British Journal of Psychotherapy* 26(2): 192–201. <https://doi.org/10.1111/j.1752-0118.2010.01173.x>

- Krueger RA and Casey MA (2015) *Focus groups: A practical guide for applied research*. Sage.
- Lacan J (1972) Seminar on “The Purloined Letter” (J Mehlman, Trans.). *Yale French Studies* 48: 39–72. <https://doi.org/10.2307/2929623>
- Lecompte-Van Poucke M (2022) ‘You got this!’: A critical discourse analysis of toxic positivity as a discursive construct on Facebook. *Applied Corpus Linguistics* 2(1): 100015. <https://doi.org/10.1016/j.acorp.2022.100015>
- Lew Z and Flanagan AJ (2025) Toxic positivity on social media: The drawbacks and benefits of sharing positive (but potentially platitudinous) messages online. *New Media & Society* 27(5): 2972–2995. <https://doi.org/10.1177/14614448231213944>
- Liang Y, Aroles J and Brandl B (2022) Charting platform capitalism: Definitions, concepts and ideologies. *New Technology, Work and Employment* 37(2): 308–327. <https://doi.org/10.1111/ntwe.12234>
- Menzies IEP (1960) A case-study in the functioning of social systems as a defence against anxiety: A report on a study of the nursing service of a general hospital. *Human Relations* 13(2): 95–121. <https://doi.org/10.1177/001872676001300201>
- Orgad S and Gill R (2022) *Confidence culture*. Duke University Press.
- Parkinson B (2011) Interpersonal emotion transfer: Contagion and social appraisal. *Social and Personality Psychology Compass* 5(7): 428–439. <https://doi.org/10.1111/j.1751-9004.2011.00365.x>
- Putra RP, Ramadhanti A, Sasanti A, Fadil A and Salsyabila N (2023) Toxic positivity in adolescents: An attitude of always being positive in every situation. *Journal of Psychology and Instruction* 7(1): 11–21.
- Salopek AH and Eastin MS (2024) Toxic positivity intentions: An image management approach to upward social comparison and false self-presentation. *Journal of Computer-Mediated Communication* 29(3): zmae003. <https://doi.org/10.1093/jcmc/zmae003>
- Shipp HG and Hall KC (2024) Analyzing the concept of toxic positivity for nursing: A dimensional analysis approach. *Journal of Advanced Nursing*. <https://doi.org/10.1111/jan.16057>
- Srnicek N (2017) *Platform capitalism*. Polity.
- Swart J (2021) Experiencing algorithms: How young people understand, feel about, and engage with algorithmic news selection on social media. *Social Media + Society* 7(2): 20563051211008828. <https://doi.org/10.1177/20563051211008828>
- Treem JW and Leonardi PM (2013) Social media use in organizations: Exploring the affordances of visibility, editability, persistence, and association. *Annals of the International Communication Association* 36(1): 143–189. <https://doi.org/10.1080/23808985.2013.11679130>
- Upadhyay IS, Srivatsa K and Mamidi R (2022) Towards toxic positivity detection. *Proceedings of the Tenth International Workshop on Natural Language Processing for Social Media* 75–82. <https://doi.org/10.18653/v1/2022.socialnlp-1.7>
- Waterloo SF, Baumgartner SE, Peter J and Valkenburg PM (2018) Norms of online expressions of emotion: Comparing Facebook, Twitter, Instagram, and WhatsApp. *New Media & Society* 20(5): 1813–1831. <https://doi.org/10.1177/1461444817707349>
- Xu B, Chang P, Welker CL, Bazarova NN and Cosley D (2016) Automatic archiving versus default deletion. *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* 1662–1675. <https://doi.org/10.1145/2818048.2819948>
- Zhao X, Salehi N, Naranjit S, Alwaalan S, Volda S and Cosley D (2013) The many faces of Facebook. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* 1–10. <https://doi.org/10.1145/2470654.2470656>

Zielinski MJ and Veilleux JC (2018) The perceived invalidation of emotion scale (PIES): Development and psychometric properties of a novel measure of current emotion invalidation. *Psychological Assessment* 30(11): 1454–1467. <https://doi.org/10.1037/pas0000584>

Author biographies

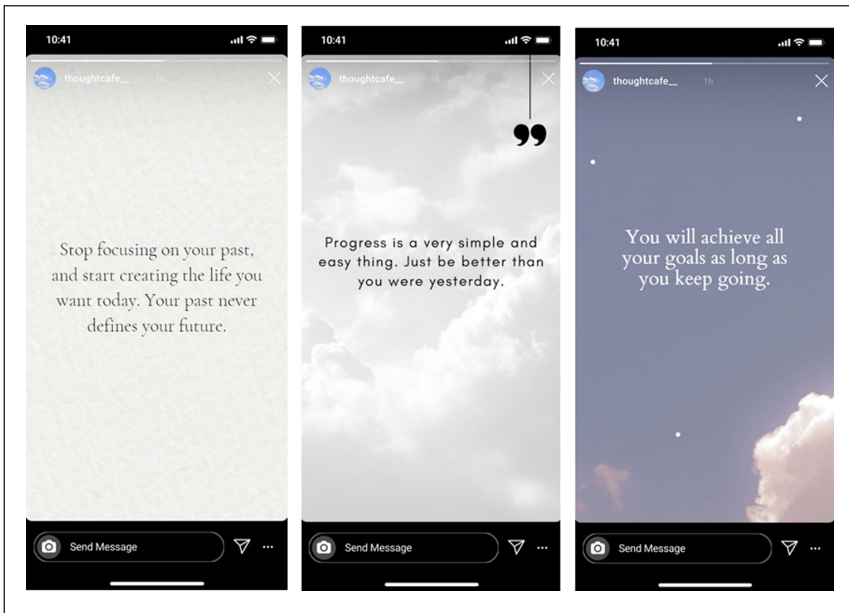
Stacy Siqi Wong graduated from Nanyang Technological University with a Bachelor of Social Sciences in Psychology and Media Analytics.

Aretha L. H. Wan graduated from Nanyang Technological University with a Bachelor of Social Sciences in Psychology and Media Analytics.

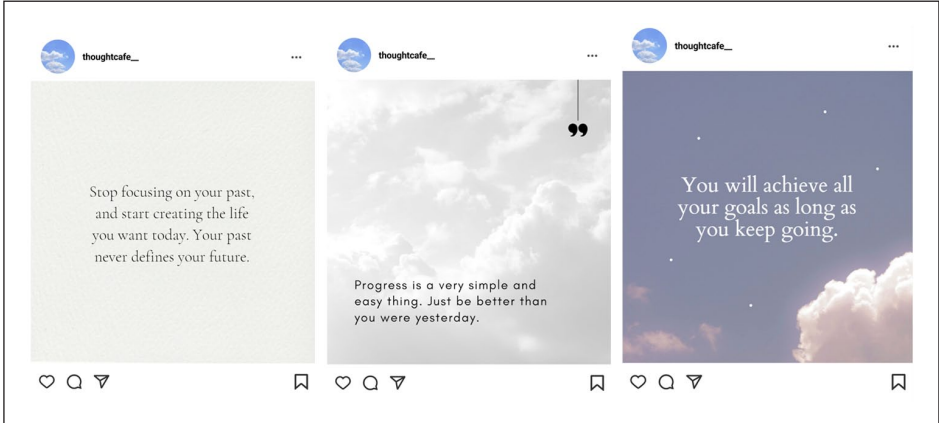
Zijian Lew (PhD, University of California, Santa Barbara) is an Assistant Professor at the Wee Kim Wee School of Communication and Information, Nanyang Technological University, Singapore. He studies self-presentation, social influence, and meaningfulness in digitally-mediated contexts.

Appendix

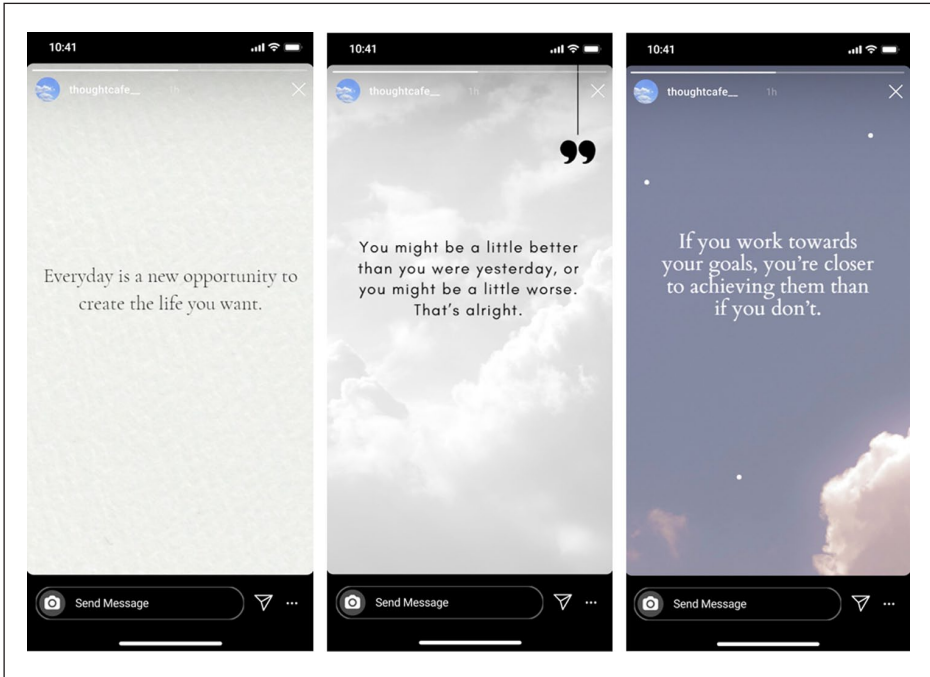
Stimuli used in Study 2



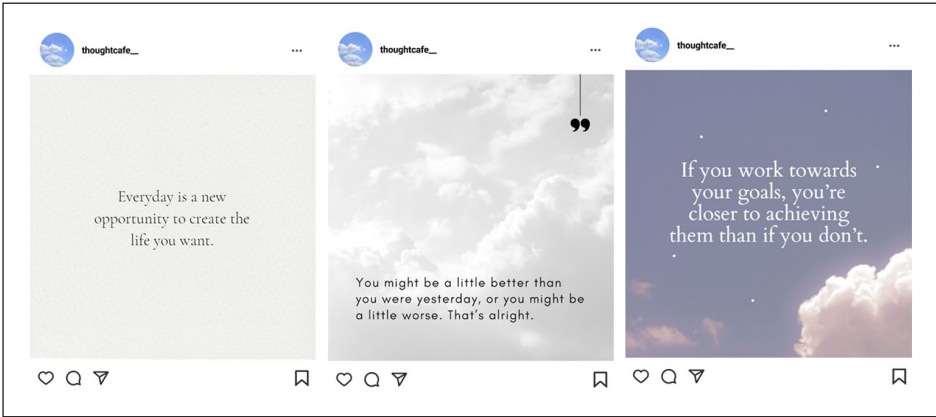
Condition 1: Ephemeral Posts with Toxic Positivity Message Characteristics (TPMC).



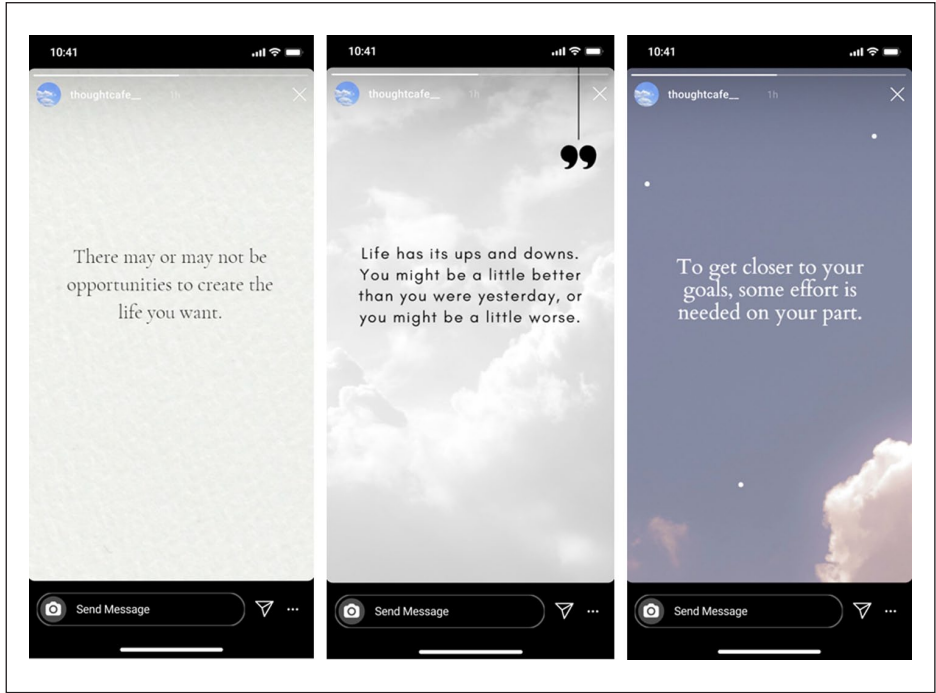
Condition 2: Permanent Posts with Toxic Positivity Message Characteristics (TPMC).



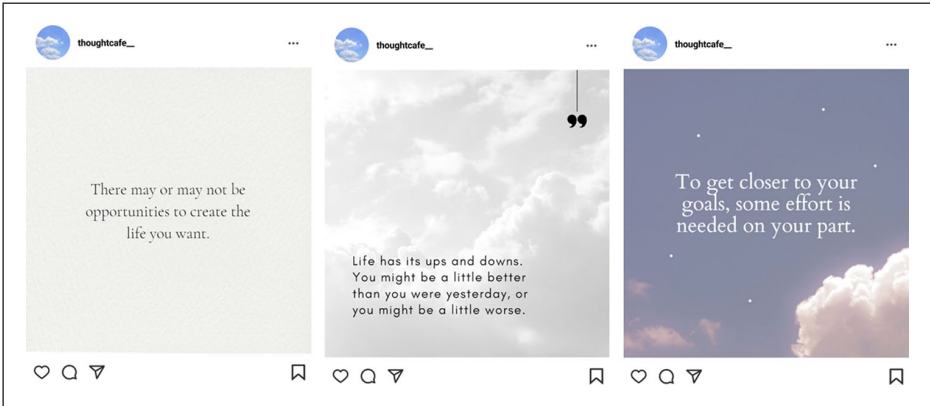
Condition 3: Ephemeral High Positivity Posts.



Condition 4: Permanent High Positivity Posts.



Condition 5: Ephemeral Low Positivity Posts.



Condition 6: Permanent Low Positivity Posts.