Article

new media & society
2025, Vol. 27(5) 2972–2995

DOI: 10.1177/14614448231213944
journals.sagepub.com/home/nms
Sage

Toxic positivity on social media: The drawbacks and benefits of sharing positive (but potentially platitudinous) messages online

Zijian Lew
Nanyang Technological University, Singapore

Andrew J Flanagin
University of California, Santa Barbara, USA

Abstract

Sharing positive messages on social media can produce positive outcomes for message senders due to *self-effects*—the effect of sending messages on message senders themselves. In this domain, one question is whether the performative display of positivity can engender positivity. By examining the sharing of personal experiences in a positive manner on social media, several boundary conditions to self-effects were found: displaying positivity is beneficial to message senders *only* if message senders have higher (vs. lower) self-esteem or if they experience less (vs. more) toxicity—defined as the suppression of the negative aspects of one's perceived reality due to engagement with or sending a positive message. Otherwise, displaying positivity can dampen enjoyment or make message senders reluctant to commit to their public self-presentations. However, after people receive feedback from friends, perceived social approval is a better predictor of enjoyment and commitment than displaying positivity.

Keywords

Self-effects, self-presentation, social media, toxic positivity

Corresponding author:
Zijian Lew, Wee Kim Wee School of Communication and Information, Nanyang Technological University, 31 Nanyang Link, 637718, Singapore.
Email: zlew@ntu.edu.sg

Although communication research has traditionally focused on how audiences are influenced by the messages they view, the capacity for ordinary individuals to send messages to others via social media has spawned a new line of inquiry that turned this research focus on its head. Instead of studying the effect of messages on audiences, *self-effects* research studies the effect of sending messages on the message senders themselves. The term *self-effects* was proposed by Valkenburg (2017) to encapsulate a commonality across seemingly-disparate research areas such as self-persuasion, self-concept change, expressive writing, and political deliberation: "how message creators/senders involuntarily influence their own cognitions, emotions, attitudes, or behavior" (Valkenburg, 2017: 478). For example, in the self-persuasion paradigm, people who publicly put forth a particular argument—which may be counter-attitudinal—subsequently aligned their attitudes or behaviors with what they said (e.g. Loman et al., 2018; see Aronson, 1999 for review). And in the self-concept change paradigm, research on identity shift theory found that people reported changes to their self-concepts after they publicly self-presented with certain personality traits, again in accordance with their self-presentations (see Carr et al., 2021 for review).

The general finding across these paradigms is that after people self-presented as someone they are not, their non-veridical self-presentations ironically induced changes to their attitudes and/or self-concepts. The very simplicity of the idea is appealing: mere utterance is often enough to change people's attitudes/self-concepts, when performed on public platforms such as social media.

This introduces the possibility that sharing positive posts or messages on social media can result in positive outcomes for message senders by dint of self-effects. This is particularly relevant as a positivity bias abounds online (Wahl-Jorgensen, 2018). For instance, "confidence culture," a phenomenon where people use positive messages to exhort themselves and others into self-assurance, has become normalized in contemporary social media (Orgad and Gill, 2021). Therefore, a central question is whether performative posturing results in positive outcomes for the message senders. Can the *display* of positivity actually *engender* positivity? Equally importantly, can the display of positivity have negative consequences?

To address these questions, this research examined how the sharing of a personal experience in a positive manner on social media can lead a message sender to a more positive memory of the experience (i.e. enjoyment). This is an especially pertinent question as social media have become very popular avenues through which personal memories are stored and shared (Van House, 2011). Yet, given how memories are dynamically reconstructed with each recollection, "it is not so much what is remembered that is critical, as how memories are expressed" (Fivush et al., 2017: 128). It is therefore plausible that if people share an experience on social media in a positive manner, they can look back upon the experience through rose-tinted lenses and rate the experience as more enjoyable. As such, one important outcome that the present research focused on was people's *enjoyment* of an experience they shared on social media, which served as an indicator of how positively message senders felt about the experiences they shared, after the sharing.

The second dependent variable that the present research examined was *commitment* to one's self-presentation. Research on online self-effects has consistently theorized that

when message senders are identifiable (vs anonymous) and when messages are made public (vs kept private), people become highly committed to their self-presentations due to their desire to appear consistent to others (Carr et al., 2021; Valkenburg, 2017). In other words, commitment is a mechanism for self-effects—but commitment is usually inferred, rather than measured, in existing studies. To close this empirical gap, the present research explicitly measured commitment and reframed the logic by conceptualizing commitment as a *proximal* outcome caused by sharing messages publicly on social media (thus implying that self-effects, such as sharing more positive messages on social media in future, are *distal* outcomes). In this way, commitment was used as a proxy to gauge the strength of self-effects—the greater one's commitment to a self-presentation, the stronger the self-effects.

In addition, several boundary conditions (namely: toxicity, self-esteem, and volition) were proposed that can result in a negative outcome for message senders despite the display of positivity. Toxicity, self-esteem, and volition are factors that may dampen the self-effects of sharing positive messages. It was hypothesized that enjoyment/commitment were inhibited if message senders felt that their positive messages were toxic (i.e. if messages suppressed the negative aspects of message senders' perceived reality), if they had low self-esteem, or if experimental instructions stripped them of the volition to share what they want. Finally, using a longitudinal study design, the influence of getting social approval from one's social media contacts on one's assessment of the shared experience is also assessed.

Self-effects and the display of positivity

Self-effects

The term *self-effects* was coined to broadly encompass several distinct lines of research, including identity shift and expressive writing (Valkenburg, 2017). Current theorizing in identity shift argues that self-effects arise when people feel beholden to their public self-presentations and therefore alter their privately-held self-concepts or attitudes to suit their public self-presentations (Schlenker, 1986). Gonzales and Hancock (2008) established the identity shift paradigm based on their research showing that participants who self-presented publicly on a blog rated themselves as more extroverted if they wrote about extroverted experiences, and as more introverted if they wrote about introverted experiences, whereas participants who self-presented privately in a text document did not show changes to their self-concepts. Subsequent studies built on identity shift to show that people generally internalize the opinions they publicly express on social media (e.g. Lane et al., 2019; Winter et al., 2022). As such, it may be possible that people will similarly internalize their own messages when they express positivity regarding an experience.

Research on expressive writing is also informative regarding self-effects. In expressive writing research (Pennebaker, 1997), participants were usually tasked to write, for 3–5 days, and for 15–30 minutes each day, either about emotionally-significant experiences (treatment group) or about superficial topics (control group). Studies generally showed that treatment group participants had better physical and mental health than

control group participants (see Pennebaker and Chung, 2011 for review). One theorized reason, among several, for the effectiveness of the treatment is that "verbally labeling an emotion may itself influence the emotional experience" (Pennebaker and Chung, 2011: 428). And although expressive writing is typically used to cope with negative emotional experiences (see Baikie and Wilhelm, 2005 for review), its core principles have been successfully applied to positive emotional experiences such as the maintenance of romantic relationships (Slatcher and Pennebaker, 2006). Within a self-effects context, this reasoning can be extended to ask whether the mere use of emotion words in itself can induce people to remember the experience as being more *enjoyable*, than if neutral, negative, or non-emotion words were used.

Separately, another line of research has brought to bear the influence of retelling personal experiences on self-effects: narrative meaning-making. Narrative meaning-making refers to the process through which "persons turn episodes in time into subjective, meaningful experiences [that] shape self-identity, guide future behavior, and connect individuals to others" (Graci and Fivush, 2017: 489). Research in the narrative meaning-making tradition shows that narrating a *personal* story can play a constitutive role in how people understand themselves (see for review Fivush et al., 2017). Among other things, narrative meaning-making allows message senders to create coherent narratives for themselves by linking their cognitions to their external environments (Fivush et al., 2017). For example, narrating a photographed experience using positive captions and posting it on social media assists people in explicitly connecting the experience to positive emotion. In addition, this process is self-reinforcing as people draw inferences from their narratives about themselves (Singer, 2004). Thus, through associating positive emotion words with an experience and through making inferences about oneself, positively describing an experience may elicit greater recalled enjoyment of the experience than neutrally describing the experience.

H1. Sharing experiences in a positive manner results in greater enjoyment than sharing in a neutral manner.

Besides enjoyment, another outcome that deserves attention is commitment (i.e. how likely people are to portray themselves consistently over time), which has been theorized as an important mechanism driving self-effects (Carr et al., 2021; Valkenburg, 2017). Due to the public nature of social media posts (as compared with, for example, writing in a diary), people may feel compelled to commit to their public self-presentations. As Schlenker (1986) argued, "public compared to private behavior is more committing, in that it is more difficult to revoke, implies that the actor will behave commensurately in the future, and implies that he or she has behaved similarly in the past" (p. 27). However, most studies within the self-effects paradigm have only inferred the activation of commitment—as opposed to having measured it as a manifest variable—when participants' post-experimental self-concepts/attitudes were commensurate with the messages they sent during the experiment. The present study measures commitment and tests it as an outcome variable (as opposed to testing it as a mechanism for a separate, distal self-effect outcome).

Assuming that people are intrinsically motivated to self-present positively by constructing desirable images of themselves to others (Leary and Kowalski, 1990), from a traditional self-effects perspective, message senders should be more committed to their positive self-presentations than to their neutral self-presentations. Therefore,

H2. Sharing experiences in a positive manner results in greater commitment than sharing in a neutral manner.

Thus, from a traditional self-effects perspective, it seems like the display of positivity can engender enjoyment or commitment. However, the expression of positivity can also be *toxic*, forming a boundary condition on self-effects.

Toxic positivity and toxicity

The term *toxic positivity* has been used in the popular psychology literature to describe the phenomenon where people portray a positive attitude, such as happiness or confidence, but in a way that appears platitudinous (Goodman, 2022; Lukin, 2019). Examples of toxic positivity include posting a smiling photo of oneself on social media with the caption "Grow through what you go through" when there is no realistic evidence for such unbridled optimism in a difficult time, or sharing some hackneyed motivational quote.

Accordingly, the present research conceptualizes *toxicity* as the extent to which a positive message encourages the suppression of the negative aspects of one's perceived reality. When people view the "live, laugh, love" poster on the wall or share the "tough times never last, but tough people do" quote on social media, if they feel psychologically pressured to be optimistic at the expense of level-headedly facing their reality, these messages can be said to be toxic. Within a self-effects perspective, toxicity involves—like the self-glorification that is common in selective self-presentation—a discrepancy between one's perceived actual self and one's presented self. But whereas selective self-presentation does not imply any suppression of one's own reality, toxicity does.

There may be several reasons why the sharing of positive messages on social media can result in negative outcomes for message senders due to toxicity. First, suppressing negative emotions is emotionally deleterious (see Gross, 2002 for review). In one study, suppressing negative emotions regarding an upsetting event led some participants to experience more negative mood compared with not receiving any instructions to control emotions regarding an upsetting event, especially if participants were predisposed toward negative affect (Dalgleish et al., 2009). Second, a person's overall well-being is not wholly determined by positive emotions—negative emotions also play a part. Specifically, suppressing negative emotions can constrain people's ability to feel gratitude. Research on gratitude suggests that the negative aspects of people's lives are important to overall well-being as they facilitate people to be more appreciative and thankful for what they have (see A. Wood et al., 2010 for review). Therefore, sending positive messages that suppress one's negative reality may encumber one's ability to be grateful, detrimentally affecting one's well-being. Finally, toxicity can engender feelings of cognitive dissonance. A study on emotional labor among customer service workers found that the more

those workers performed emotions that they did not feel (e.g. pretend to be in a good mood when interacting with customers), the more emotionally exhausted they were and the less job satisfaction they had (Pugh et al., 2011). In all, across several distinct lines of research, the suppression of one's perceived negative reality detrimentally affects people's cognitions and emotions.

Toxicity and inauthenticity have some notable overlaps, in that both concepts involve biased processing that denies what a message sender regards as true (see Kernis and Goldman, 2006 for review). Yet, there are also key differences between the two concepts. First, inauthentic self-presentations (a discrepancy between one's perceived actual self and one's self-presentation) do not necessarily engender negative affect in a message sender. For example, people can be satisfied that their inauthentic online dating profiles are attracting many potential partners, or that their inauthentic LinkedIn profiles are attracting the attention of job recruiters. By contrast, toxic self-presentations are theorized to be (superficially) positive but to elicit negative outcomes.

Yet, inauthenticity can engender negative outcomes as well, suggesting the importance of a second difference: inauthenticity may engender self-effects due to identity shift—but toxicity is unlikely to do so. According to identity shift theory, even inauthentic self-presentations can be internalized into message senders' self-concepts (see Carr et al., 2021 for review). Crucially, identity shift does not require message senders to suppress their reality even when their self-presentations are inauthentic. For example, Gonzales and Hancock (2008) instructed participants to draw on their past and present experiences to self-present as either introverted or extroverted. That is, even if introverts/extroverts were tasked to (inauthentically) self-present as extroverted/introverted, they were tapping into, rather than suppressing, some facet of their reality.

A third difference pertains to people's awareness of inauthenticity or toxicity. People should be at least primarily aware when their online self-presentations are inauthentic, assuming they were minimally strategic (i.e. they *intended* to be inauthentic). Where toxicity is concerned, however, people may well be unaware of the negative, unintended consequences of their self-presentations, whatever their intentions may be (e.g. conforming to norms on social media by being positive).

In all, sharing positive messages by acknowledging only the positive aspects of life and not the negative aspects potentially induces negative outcomes. Therefore, sharing experiences in a positive but toxic manner should result in less enjoyment than sharing in a neutral manner. Conversely, sharing experiences in a positive, non-toxic manner should result in greater enjoyment than sharing in a neutral manner. Therefore,

H3. The relationship between the valence (positive vs neutral) of a shared experience and people's enjoyment of the experience (as described in H1) is moderated by toxicity.

Considering these arguments, it follows that toxicity may also moderate the relationship between message valence and commitment. That is, people are likely to have greater commitment to their self-presentations if they share their experiences in a positive manner, than if they share their experiences in a neutral manner—but only if the messages are

deemed non- or less-toxic. Conversely, sharing experiences in a positive but toxic manner results in less commitment than sharing in a neutral manner. Therefore,

H4. The relationship between the valence of a shared experience and people's commitment to the portrayed self-image (described in H2) is moderated by the toxicity of what they shared.

Self-esteem

Another potential boundary condition of self-effects concerns not whether the emotional valence of a message matches perceived reality (as is the case for toxicity), but whether people's individual dispositions match the content of their messages. To be precise, self-esteem can limit the extent to which message senders reap the benefits of sharing their positive experiences.

According to Swann (2011), "people prefer others to see them as they see themselves, even if their self-views happen to be negative . . . Presumably, people seek self-verification because self-verifying evaluations make the world seem coherent and predictable" (p. 23). For example, among people with high self-esteem, making positive statements about themselves improves mood compared with not making any self-referential statements; but among people with low self-esteem, making positive statements about themselves actually decreases mood compared with not making any self-referential statements (J. Wood et al., 2009). Compared with people with high self-esteem, people with low self-esteem are also more likely to repress positive emotions or avoid thinking of positive events than remembering or celebrating the positive event (Goodall, 2015).

Therefore, sharing experiences in a positive manner should result in greater enjoyment than sharing in a neutral manner, but only if those who share have high self-esteem when they share. Conversely, sharing experiences in a positive manner results in less enjoyment than sharing in a neutral manner if those who share have low self-esteem when they share. As such,

H5. The relationship between the valence of a shared experience and people's enjoyment of the experience (described in H1) is moderated by people's self-esteem.

Of course, H5 would make sense only if the sharing of positive experiences online does not influence message senders' level of self-esteem, however momentarily. This premise seems reasonable, as existing research suggests it is high self-esteem that leads to positive affect, and not positive affect that leads to high self-esteem (Joshanloo, 2022; Leary and Baumeister, 2000). One online photography study on women's body image found that taking selfies resulted in lower self-esteem than taking photos of objects due to increased self-objectification (Fox et al., 2021), but no selfies were permitted in the present study (see Method section). Nevertheless, given the prior evidence on how sharing messages online can influence people's self-concepts (e.g. Gonzales and Hancock, 2008), the present study should check if the positivity of shared messages influences senders' self-esteem (see Results section).

Similarly, sharing experiences in a positive manner should result in greater commitment than sharing in a neutral manner, but only if those who share have high self-esteem when they share. In contrast, sharing experiences in a positive manner results in less commitment than sharing in a neutral manner if those who share have low self-esteem when they share. Therefore,

H6. The relationship between the valence of a shared experience and people's commitment to the portrayed self-image (described in H2) is moderated by people's self-esteem.

Volition

Self-effects can occur even when participants were instructed to self-present in a way that is different from how they typically see themselves, or when the position they were instructed to advocate is counter-attitudinal (see Valkenburg, 2017 for review). In more naturalistic environments, however, it would be more reasonable to expect that individuals have full control over the things they share on social media. As a consequence, self-effects in naturalistic environments should be the product of individuals' conscious *volition*, and not due to randomly-assigned instructions.

Indeed, Carr et al. (2021) argue that people's "desire and ability to change" (p. 210) may be important boundary conditions for self-effects. Similarly, Walther and Lew (2022) point out that in non-experimental settings, "volition—and by extension, choice and/or customization—is the starting point of any transformation of self" (p. 150).

Although it is difficult to give participants full volition to write whatever they like in an experiment, the degree of volition that participants may have in crafting their social media posts can still be incorporated into the experimental design. Concerns over volition highlight a potential point of tension in identity shift research. On one hand, experimentally instructing participants exactly what to write on social media (low volition) is good for experimental control. In the present research, this is advantageous for eliminating random variability in terms of each participant's own interpretation of what positive or neutral messages may be. On the other hand, with a greater degree of volition, participants should also be more motivated during the sending of messages (Deci and Ryan, 2012), which could strengthen the potency of self-effects.

Furthermore, across several studies on expressive writing, participants did not gain any health benefits if experimental instructions interfered with how they could make sense of and express themselves in their own way (see Pennebaker and Chung, 2011 for review). These (non-significant) findings are commensurate with the present argument: volition seems crucial for self-effects, and removing participants' freedom of expression may impede the formation of positive self-effects outcomes or even lead to negative outcomes.

If volition were indeed important to self-effects, then the effects of sharing messages on social media should be strongest when the messages are volitionally written in a positive manner. It is therefore possible that sharing experiences in a positive manner results in greater enjoyment than sharing in a neutral manner, and this relationship is stronger when people have a high degree of volition than when they have a low degree of volition.

H7. The relationship between the valence of a shared experience and people's enjoyment of the experience (described in H1) is moderated by volition.

Likewise, it is possible that sharing experiences in a positive manner results in greater commitment than sharing in a neutral manner, and this relationship is stronger when people share with a high degree of volition than when they share with a low degree of volition.

H8. The relationship between the valence of a shared experience and people's commitment to the portrayed self-image (described in H2) is moderated by volition.

Long(er)-term self-effects

An under-examined aspect of self-effects research concerns how long-lasting self-effects may be, due to the relative paucity of longitudinal studies compared with cross-sectional studies within the self-effects paradigm (see Carr et al., 2021; Walther and Lew, 2022 for review). Unlike in-person interactions, feedback is seldom obtained immediately after one makes a post on social media. As such, a longitudinal design can delineate the immediate self-effects of sending messages and the subsequent effects due to receiving feedback. In addition, the context of social media creates relatively unique conditions to study feedback. Whereas offline self-presentation is limited to a small number of in-person audiences, online self-presentation tends to have much larger audiences, whereby people can collectively leave feedback (such as likes, reactions, or comments) that reinforces self-effects (Walther et al., 2011). Such feedback gives a message sender some sense of *perceived social approval*, which may strengthen the sender's commitment to an existing online self-presentation.

Audience feedback and perceived social approval

A key factor in understanding how social media influence self-effects is audience feedback, sometimes termed paralinguistic digital affordances, or "cues in social media that facilitate communication and interaction without specific language associated with their messages" (Hayes et al., 2016: 172–173). Examples include likes or favorites (to use Facebook and X/Twitter terminology).[1] At a surface level, a social media post that has received more likes seems as if it has gotten more favorable feedback than a social media post that has received fewer likes. However, what may be more important is not the raw number of likes, but how likes are interpreted. For example, if one expects a post to receive 50 likes, but it got only 25, the feedback can potentially be interpreted as negative. But if one expects a post to get 5 likes, and it got 25, the feedback may be interpreted as positive. Indeed, what constitutes a "successful" number of likes by those who share social media posts is influenced by social comparison and reciprocity norms (Carr et al., 2018). It is therefore appropriate to consider the *perceived social approval* accrued due to a social media post instead of the raw number of likes.

In the present context, perceived social approval refers to the extent to which one believes that others display favorable attitudes toward one's shared social media content.

Insofar as text-based feedback (e.g. comments, personal direct messages) is concerned, the valence of the feedback can be assumed for the most part to be neutral to positive. As for feedback in the form of likes, it is reasonable to think that there is a positive correlation between the raw number of likes and the social approval people think that they have obtained, despite the more equivocal interpretation of likes than text (Hayes et al., 2016). Therefore,

H9. As people perceive greater social approval of the experiences they share, they feel greater enjoyment toward the experience.

In addition, the perception of social approval should also encourage message senders to commit to the self-image they portrayed on social media:

H10. As people perceive greater social approval of the experiences they share, they have greater commitment to their portrayed self-image.

There may also be an interaction between perceived social approval self-esteem. If people with low self-esteem prefer to verify their negative self-views (Swann, 2011), they may struggle to reap the benefits of sharing positive messages as they may not view others' feedback or social approval as rewarding. Yet, people with low self-esteem may nonetheless seek out positive feedback if they believe it is warranted (Hoplock et al., 2019). Therefore,

RQ1. Do perceived social approval and self-esteem produce an interaction effect on enjoyment?

RQ2. Do perceived social approval and self-esteem produce an interaction effect on the commitment to one's portrayed self-image?

Method

A 2 (valence: positive/neutral captions) × 2 (volition: volitional/non-volitional captions) between-subjects experiment was conducted to test the hypotheses. The study had two parts. Part 1 required participants (undergraduate students at the University of California, Santa Barbara) to photograph three different landmarks on the university campus, then share the photos to their actual social media accounts. Participants were given different instructions based on their randomly-assigned condition. They had to write either positive captions or neutral captions when they uploaded their photos to social media, and they either came up with a suitable positive/neutral caption volitionally, or were randomly given a predetermined positive/neutral caption to write non-volitionally. After participants uploaded their photos to their own social media accounts, they filled out a questionnaire, thus completing Part 1. Part 2 consisted of an online questionnaire that was emailed to participants 3 days after Part 1.

Sample

Participants were undergraduates who, as a qualifying criterion, posted at least three photo/video posts about their personal experiences (i.e. not memes or news stories) on any social media app in the 4 weeks preceding Part 1. Participants were rewarded with course credit.

Power analyses via G*Power based on a small-to-medium effect size ($f^2 = .085$), as is common in social science research, showed that 176 participants were required to have .80 power for the analysis needed for hypothesis testing. Participants who failed either of two attention check questions in the Part 1 questionnaire were removed from analyses ($n = 51$). Participants who uploaded their photos to VSCO ($n = 2$) were also removed because VSCO has no function for likes or comments, and does not display a follower count, making it unsuitable for testing commitment or perceived social approval. A few participants were eliminated ($n = 3$) due to a highly unusual weather event during the photo task, which may have affected their experience in an unknown fashion. Consequently, $N = 308$ participants were retained for analyses.

Participants were mostly female (64.3%; male = 20.5%; the remaining 15.3% did not indicate their sex). Their ages ranged from 18 to 27 years old ($M = 19.7$, $SD = 1.3$). White participants comprised 34.1% of the sample, Asian participants 20.1%, Hispanic or Latino participants 16.2%, Black or African American participants 1.3%, Native Hawaiian or Pacific Islander 0.6%; 12.4% indicated mixed ethnicity and 15.3% did not state their ethnicity.[2]

Procedure

Following the screening criteria mentioned earlier, eligible participants were randomly placed into experimental conditions and received further instructions in person. All participants were instructed to take photos of three different prominent campus locations. Participants then received condition-specific instructions telling them that they would eventually be asked to post the photos to the social media platform of their choice, alongside neutral or positive captions that were either assigned (non-volitional) or up to them to make up (volitional).

For participants in the non-volitional captions conditions, those assigned to write neutral captions were told to "describe the location of your photos, e.g., 'At the [university landmark]'"; those assigned to write positive captions were told to write one of the following lines, which were randomly selected: "It's a good day to be a [university nickname] / Finding joy in the ordinary / Always thriving :) / Do more of what makes you happy / Think happy, stay happy." For participants in the volitional captions conditions, those assigned to write neutral captions were told: "Add, to your post/photos, neutral captions of your choice. The key idea is to write captions that are neutral, and not come across as too strong in opinion or feelings"; and those assigned to write positive captions were told, "Add, to your post/photos, positive captions of your choice. The key idea is to write captions that are uplifting to audiences." See the Online Supplemental Materials for the complete instructions.

After taking the required photos, participants returned to the lab, where they were told to post the photos to their actual social media accounts. Condition-specific instructions were also repeated to guide participants. Participants then took screenshots of what they posted, and emailed the screenshots to a researcher for verification. They then completed the Part 1 questionnaire. Three days later, participants received a link to the online questionnaire for Part 2 of the study via their university email accounts, which concluded the study.

Measures

Confirmatory factor analyses of the continuous variables measured at Part 1 and Part 2, separately, were reported in the Online Supplemental Materials.

Part 1. Toxicity was measured using five original items. Participants were told, "Indicate how much you agree/disagree with the following statements." The question stem, "The social media post I made," was followed by five statements scored from 1 (*strongly disagree*) to 7 (*strongly agree*): "Fails to acknowledge the hardships in life," "Is unrealistically optimistic," "Pressures me to be positive," "Alienates me from my true feelings," and "Ignores the challenges I'm facing." Cronbach's $\alpha = .75$.

Self-esteem was measured using five items adapted from Heatherton and Polivy (1991). The question stem read, "How true are each of these statements for you right now?" Items were scored on a scale from 1 (*strongly disagree*) to 7 (*strongly agree*): "I feel satisfied with myself," "I feel that others respect me," "I feel good about myself," "I feel that others admire me," and "I feel confident," Cronbach's $\alpha = .87$.

Enjoyment was an original measure with the question stem, "How do you feel about the activity depicted in your post? It was: . . ." and four items scored on a 7-point semantic differential scale. The items were: *unpleasant–pleasant*, *unenjoyable–enjoyable*, *dissatisfying–satisfying*, and *uninteresting–interesting*, Cronbach's $\alpha = .89$.

Commitment was measured using three original items. The question stem read, "Compared to the way you portrayed yourself in your post," and it was followed with three items: "how committed are you to maintaining this image of yourself in the future? (1 = *very uncommitted*, 7 = *very committed*)," "how consistent with this portrayal do you think you will be in your future self-portrayals? (1 = *very inconsistent*, 7 = *very consistent*)," and "how differently would you portray yourself in the future? (1 = *very differently*, 7 = *very similarly*)." Cronbach's $\alpha = .86$.

Part 2. Perceived social approval was measured using six original items scored from 1 (*strongly disagree*) to 7 (*strongly agree*): "Other people approve of my post," "Other people have favorable opinions toward my post," "Other people have positive impressions of my post," "Other people react supportively toward my post," "Other people are not impressed by my post" (reverse coded), and "Other people don't care about my post" (reverse coded), Cronbach's $\alpha = .78$. Enjoyment (Cronbach's $\alpha = .90$) and commitment (Cronbach's $\alpha = .86$) were measured with the same items as before. Bivariate correlations of all measured variables are shown in Table 1.

Table 1. Means, standard deviations, and Pearson's correlation coefficients between measured variables.

	Mean	Standard deviation	Toxicity (T1)	Self-esteem (T1)	Perceived social approval (T2)	Enjoyment (T1)	Enjoyment (T2)	Commitment (T1)
Toxicity (T1)	4.14	1.18	–					
Self-esteem (T1)	4.91	1.14	−.218**	–				
Perceived social approval (T2)	4.33	0.87	−.169**	.185**	–			
Enjoyment (T1)	5.89	1.10	−.111	.181**	.314**	–		
Enjoyment (T2)	5.67	1.09	−.210**	.105	.360**	.713**	–	
Commitment (T1)	4.18	1.45	−.125*	.203**	.353**	.396**	.380**	–
Commitment (T2)	4.10	1.48	−.197**	.145*	.372**	.370**	.434**	.664**

*$p < .05$; **$p < .01$. T1 and T2 refer to the two times when data were collected.

Table 2. Linear regression model predicting enjoyment.

	b	*SE*	*t*	*p*
Intercept	5.883	0.062	94.388	< .001
Valence	−0.033	0.125	−0.268	.789
Volition	0.109	0.123	0.889	.375
Toxicity	−0.076	0.055	−1.379	.169
Self-esteem	0.157	0.056	2.809	**.005**
Valence × volition	−0.064	0.246	−0.260	.795
Valence × toxicity	−0.060	0.110	−0.548	.584
Valence × self-esteem	0.280	0.112	2.510	**.013**

Neutral valence = −0.5, positive valence = 0.5; non-volitional captions = −0.5, volitional captions = 0.5.
The bold faced values are the significance values.

Results

The photos that participants took during the photography tour in Part 1 were shared on Instagram by 60.0% of participants (as posts: 32.1%, as stories: 27.9%), on Snapchat by 23.1%, on Twitter by 7.8%, on Facebook by 5.2%, and on WeChat by 2.3%. The remaining 1.6% of participants shared on Weibo, TikTok, or a combination of several apps. Participants had complete freedom[3,4] to choose the app on which they shared their photos.

Linear regression models were used to test message expression effects, which utilized data from the Part 1 questionnaire. Mixed models were used to test post-expression effects, which utilized data from both Part 1 and Part 2 questionnaires. Categorical variables were coded as such: neutral valence = −0.5, positive valence = 0.5; non-volitional captions = −0.5, volitional captions = 0.5; time 1 (Part 1) = −0.5, time 2 (Part 2) = 0.5. All continuous predictors—toxicity, self-esteem, and perceived social approval—were mean-centered.

Time 1 analysis: self-effects only

Before any hypothesis testing, there was a need to find out whether participants' self-esteem was influenced by the experimental manipulations. A valence × volition factorial ANOVA showed that self-esteem was independent of valence, $F(1, 304) = 0.034$, $p = .854$, volition, $F(1, 304) = 0.002$, $p = .964$, and the interaction between valence and volition, $F(1, 304) = 264$, $p = .608$. Thus, self-esteem can be entered into the regression models, below, as an independent predictor and moderator.

To test H1, H3, H5, and H7, enjoyment was regressed on valence, volition, toxicity, self-esteem, and the following interaction terms: valence × volition, valence × toxicity, and valence × self-esteem. The overall model was significant, $F(7, 300) = 2.966$, $p = .005$, $R^2 = .065$, $R_{adj}^2 = .043$. Table 2 presents a breakdown of the result for each predictor.

The main effect of valence on enjoyment was non-significant, $b = -0.033$, $p = .789$. Self-esteem significantly predicted enjoyment, $b = 0.157$, $p = .005$, $\eta_p^2 = .03$, showing that the more self-esteem people had, the more they enjoyed the photography experience. The valence × self-esteem interaction was also significant, $b = 0.280$, $p = .013$, $\eta_p^2 = .02$.

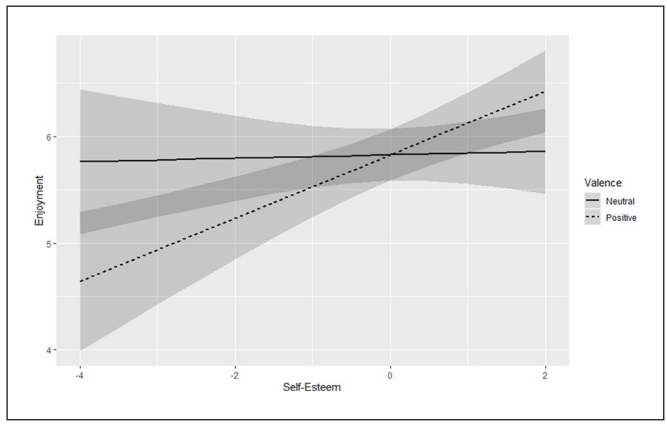

Figure 1. Effect of valence and self-esteem on enjoyment immediately after message expression.

Among people with lower self-esteem, those who shared photos alongside positive captions enjoyed the photography experience *less* than those who shared photos alongside neutral captions; but among people with higher self-esteem, those who shared photos alongside positive captions enjoyed the photography experience *more* than those who shared photos alongside neutral captions (see Figure 1). A simple effect analysis of the interaction (using the *emtrends* function in R) showed that the slope representing the relationship between self-esteem and enjoyment for neutral captions was not different from a horizontal line, $b = 0.017$, $p = .837$; but the slope for positive captions was significant, $b = 0.297$, $p < .001$. Other predictors were non-significant. Therefore, H5 was supported, but not H1, H3, or H7.

To test H2, H4, H6, and H8, commitment was regressed on the same set of predictors: valence, volition, toxicity, self-esteem, valence × volition, valence × toxicity, and valence × self-esteem. The overall model was significant, $F(7, 300) = 3.348$, $p = .002$, $R^2 = .072$, $R_{adj}^2 = .051$. See Table 3 for details.

The main effect of valence on commitment was non-significant, $b = 0.126$, $p = .441$. Self-esteem significantly predicted commitment, $b = 0.249$, $p = .001$, $\eta_p^2 = .03$. The more self-esteem people had, the more they wanted to commit to their self-presentations. The valence × toxicity interaction also predicted commitment, $b = -0.356$, $p = .014$, $\eta_p^2 = .02$. Among people who thought their post was more toxic, those who shared photos alongside positive captions had *less* commitment to self-present in a similar way in the future than those who shared photos alongside neutral captions; but among people who thought their post was less toxic, those who shared photos alongside positive captions had *more* commitment than those who shared photos alongside neutral captions (see Figure 2). A simple effect analysis of the interaction showed that the slope representing the relationship between self-esteem and toxicity for neutral captions was non-significant, $b = 0.107$, $p = .332$; but the slope for positive captions was significant, $b = -0.248$, $p = .008$. Other predictors were non-significant. Therefore, H4 was supported, but not H2, H6, or H8.

Table 3. Linear regression model predicting commitment.

	b	*SE*	*t*	*p*
Intercept	4.150	0.082	50.831	< .001
Valence	0.126	0.163	0.771	.441
Volition	0.104	0.161	0.647	.518
Toxicity	−0.071	0.072	−0.980	.328
Self-esteem	0.249	0.073	3.410	**.001**
Valence × volition	−0.093	0.323	−0.289	.773
Valence × toxicity	−0.356	0.144	−2.468	**.014**
Valence × self-esteem	0.031	0.146	0.212	.832

Neutral valence = −0.5, positive valence = 0.5; non-volitional captions = −0.5, volitional captions = 0.5.
The bold faced values are the significance values.

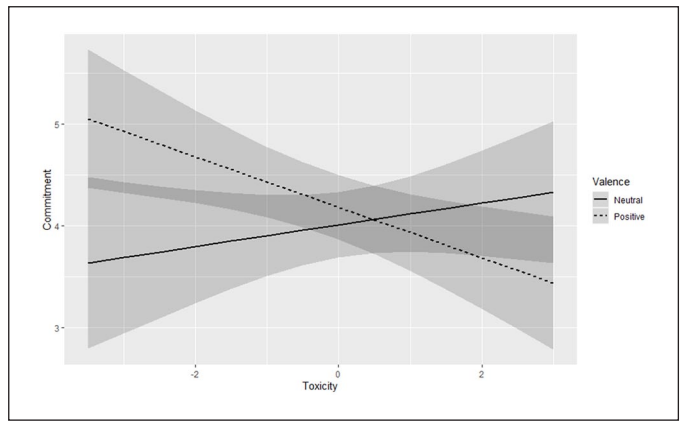

Figure 2. Effect of valence and toxicity on commitment immediately after message expression.

Time 2 analysis: self-effects after feedback

After excluding participants who failed to complete Part 2 of the study, the sample size for tests of post-expression effects was reduced to $n = 270$. To test H3, H5, H7, H9, and RQ1, a mixed model analysis was performed. To estimate fixed effects, enjoyment was regressed on valence, volition, toxicity, self-esteem, perceived social approval, time, and the following interaction terms: valence × volition, valence × toxicity, valence × self-esteem, and self-esteem × perceived social approval. Perceived social approval, time, and the self-esteem × perceived social approval interaction were new predictors, that is, they were not found in the earlier Time 1 analyses. Random intercepts were modeled for each participant to account for the repeated measurement of enjoyment within each participant. In this way, the analysis incorporated two different enjoyment scores for each participant (i.e. Time 1 and Time 2 scores), while the predictors were constant across time. Random slopes were not modeled. The *lme4* package in R was used to fit the mixed model using the restricted maximum likelihood (REML) method. Following that, the *lmerTest* package was used to obtain Satterthwaite approximations for degrees of

Table 4. Fixed effects for mixed model predicting enjoyment.

	b	*SE*	*t*	*p*
Intercept	5.745	0.059	97.047	< .001
Valence	−0.084	0.117	−0.723	.470
Volition	0.109	0.116	0.943	.347
Toxicity	−0.096	0.053	−1.821	.070
Self-esteem	0.081	0.054	1.496	.136
Perceived social approval	0.372	0.068	5.430	< **.001**
Time	−0.193	0.051	−3.786	< **.001**
Valence × volition	−0.108	0.231	−0.467	.641
Valence × toxicity	−0.106	0.105	−1.005	.316
Valence × self-esteem	0.174	0.106	1.634	.103
Self-esteem × perceived social approval	0.049	0.055	0.883	.378

Neutral valence = −0.5, positive valence = 0.5; non-volitional captions = −0.5, volitional captions = 0.5; time 1 = −0.5, time 2 = 0.5.
The bold faced values are the significance values.

freedom (see Luke, 2017), from which *p*-values were derived. The η_p^2 effect sizes reported for all mixed models were approximates obtained via the *effectsize* package.

The overall model predicting enjoyment had the following model fit statistics: pseudo-R^2 for fixed effects only = .15, pseudo-R^2 for the entire model = .72. Perceived social approval was significant, $b = 0.372$, $p < .001$, $\eta_p^2 = .10$, showing that as people perceived more social approval by their social media contacts, the more enjoyable they rated the photography experience to be. Time was also significant, $b = -0.193$, $p < .001$, $\eta_p^2 = .05$; people rated the photography experience as less enjoyable over time. Other predictors were non-significant. See Table 4 for the fixed effects results. Therefore, H9 was supported, but there was no evidence for H3, H5, H7, or RQ1.

To test H4, H6, H8, H10, and RQ2, a second mixed model analysis was performed on the outcome variable of commitment, using the same predictors and methods (i.e. random intercepts, REML estimation, Satterthwaite approximation) as before. The overall model had the following fit statistics: pseudo-R^2 for fixed effects only = .18, pseudo-R^2 for the entire model = .67. Perceived social approval was the only significant predictor, $b = 0.536$, $p < .001$, $\eta_p^2 = .12$; as people perceived more social approval by their social media contacts, they had more commitment toward how they portrayed themselves in their social media posts. Table 5 shows the fixed effects results. Therefore, H10 was supported, but there was no evidence for H4, H6, H8, or RQ2.

Discussion

Adopting a self-effects perspective, this study presents important evidence regarding the effect of valenced expressions of past experiences via social media on (a) people's recalled enjoyment of the experience and (b) their commitment to their self-presentations. Experimental findings show evidence for several boundary conditions of this domain of self-effects, in the form of the significant interaction between valence and

Table 5. Fixed effects for mixed model predicting commitment.

	b	*SE*	*t*	*p*
Intercept	4.094	0.076	53.812	< .001
Valence	−0.031	0.150	−0.205	.838
Volition	0.287	0.148	1.932	.054
Toxicity	−0.108	0.068	−1.581	.115
Self-esteem	0.143	0.069	2.057	.041
Perceived social approval	0.536	0.088	6.093	**< .001**
Time	−0.086	0.074	−1.175	.241
Valence × volition	−0.215	0.297	−0.722	.471
Valence × toxicity	−0.212	0.136	−1.561	.120
Valence × self-esteem	0.105	0.137	0.768	.443
Self-esteem × perceived social approval	0.119	0.071	1.672	.096

Neutral valence = −0.5, positive valence = 0.5; non-volitional captions = −0.5, volitional captions = 0.5; time 1 = −0.5, time 2 = 0.5.
The bold faced values are the significance values.

self-esteem on enjoyment and between valence and toxicity on commitment. Results demonstrated that people with lower self-esteem found writing positive captions less enjoyable than writing neutral captions, potentially because they did not want their social media contacts to view them with the manufactured halo of positive, upbeat messages (Swann, 2011). Conversely, people with higher self-esteem can reap the benefits of writing positive captions, which produced more enjoyment than writing neutral captions.

Another boundary condition of self-effects—insofar as valenced online expressions of past experiences were concerned—was supported when people who wrote positive captions found their positivity more toxic, and they wanted to commit to their self-presentations less than people who wrote neutral captions. But when people who wrote positive captions found their positivity less toxic, they professed greater commitment to their self-presentations than people who wrote neutral captions. Within the self-effects literature, identity shift theory (Carr et al., 2021) posits that commitment is a crucial reason why people's self-concepts and attitudes can change after a public self-presentation, but commitment was often inferred rather than measured. The present study reconceptualized commitment as a proximal outcome (that may lead to other distal self-effects outcomes beyond the scope of this study), and found some empirical evidence for commitment as an important psychological aspect of self-effects, with toxicity as a boundary condition.

In all, these results show that displaying positivity before experiencing positivity can be beneficial to message senders only if the display of positivity is aligned with people's self-views (i.e. high self-esteem) or if the content of the messages is deemed less toxic. This is consistent with research on the social sharing of emotions, which assumes that people would have experienced some emotionally valenced event before social sharing, and then experience heightened emotions after sharing (Rimé, 2009). Otherwise, displaying positivity can backfire on message senders, dampening enjoyment or creating a situation where people are reluctant to commit to, or even disavow, their public self-presentations.

The present study also revealed limits as to how much people can change their cognitions by dint of self-effects. Whereas past research has shown that people can shift their attitudes in favor of the policy opinions they publicly espoused on social media (Winter et al., 2022) or shift their self-concepts in line with their public online self-presentations (Gonzales and Hancock, 2008), the positive (vs neutral) expression of an experience did not engender cognitive changes among the participants of the present study. Valence, as a "main effect" on its own, did not influence enjoyment or commitment at both time points when measures were collected. There was no evidence that the mere sharing of positive messages will make one feel more positive or become more committed to being positive. Whether the mere sharing of other types of messages can influence message senders' cognitions remains to be seen (see Valkenburg, 2017 for review). Future research can continue studying self-effects beyond outcomes related to attitudes or self-concepts.

Unsurprisingly, self-effects do not always last. After people send messages on social media, social factors—such as perceived social approval—grow in relevance as people receive feedback (e.g. likes, comments, reactions) from their social media connections. After message senders had the opportunity to receive social media feedback, perceived social approval was the only significant predictor for commitment, and was—alongside time—one of two significant predictors for enjoyment. In the present context of toxic positivity, this should be a beneficial outcome for message senders: the negative aspects of displaying positivity despite low self-esteem or toxicity is temporary, and can be rendered irrelevant in the longer term if one perceives that one's social media friends show approval. Indeed, an exploratory post hoc t-test analysis revealed that writing positive captions ($M = 4.37$, $SD = 0.96$) did not result in greater perceived social approval than writing neutral captions ($M = 4.30$, $SD = 0.78$), $t(268) = 0.687$, $p = .493$. Unlike in previous studies (e.g. Carr and Hayes, 2019; Walther et al., 2011), feedback was not manipulated, and by extension, perceived social approval was engendered naturalistically. In the broader context of self-effects, the lack of a naturalistic relationship between valence and perceived social approval suggests that social factors do not merely reinforce or counteract self-effects. They can impact outcomes independently of self-effects, influencing people's cognitions above and beyond the act of sending messages publicly on social media.

There was no evidence for the interaction effect of self-esteem and perceived social approval on either enjoyment or commitment. In contrast to self-verification (Swann, 2011) and despite prior evidence that people with low self-esteem struggle to obtain rewards from the sharing of positive messages (J. Wood et al., 2009), participants with low self-esteem were not worse off than participants with high self-esteem after they had time to receive feedback from their social media contacts. This may be because people with low self-esteem do not universally reject positive feedback. For example, Kille et al. (2017) found that people with low self-esteem are less likely to benefit from a compliment if they think of the compliment in an abstract manner, but can benefit if they think, in concrete terms, how they will act in a manner worthy of praise. Considering this uncertainty, future research can study the conditions under which perceived social approval can reinforce or subvert self-effects.

Limitations

Several limitations due to the experimental design should be noted. For instance, the manipulation of volition could be stronger. Volition was not found to have an influence on either enjoyment or commitment. This was unexpected, considering that volition (autonomy) is considered an intrinsic human need that is essential for satisfaction (Deci and Ryan, 2012). The non-significant findings regarding volition may be due to a limitation with its manipulation: participants in the volitional captions condition had control over how to express positive or neutral captions, but they were given explicit instructions on the valence of their captions as well as the specific landmarks to photograph. Put differently, even participants in the volitional captions condition had to contend with experimental instructions that were enforced non-volitionally. As a result, the manipulation of volition may not have been strong enough. The lack of a manipulation check also created uncertainty as to whether actual volition—creating new captions or copying predetermined captions—was in agreement with perceived volition.

Relatedly, although the manipulation of volition gave participants some autonomy over their self-presentations, it did not go far enough to permit an examination of their motivations—which is important in environments with greater ecological validity. When people share their experiences in a highly positive manner on social media, are they motivated by the desire to manage others' impressions of themselves (see Leary and Kowalski, 1990 for review), obtain self-affirmation via social feedback (see Sherman, 2013 for review), explore different facets of their identities (see Walther and Lew, 2022 for review), persuade themselves that things will be better eventually (see Valkenburg, 2017 for review), or a combination of different reasons, including those not listed here? Future research can consider how people may recall greater enjoyment regarding their experiences or have greater commitment to their self-presentations if the needs driving their self-presentation motivations are fulfilled. Doing so should permit greater generalization of research findings to real-life scenarios.

Future research should carefully weigh the advantages and disadvantages of adopting a more naturalistic research method. The experimental design constrained participants to taking photographs of three university landmarks, but this may have affected the results in unknown ways. For example, toxicity may be magnified by the extent to which participants perceived that the selected landmarks were incongruent with the captions. Giving participants free rein over what to photograph can eliminate this plausible confound in future research. Yet, more naturalistic approaches have their weaknesses as well. For example, participants in the present study could share their photos on any social media platform of their choice. But would participants have chosen a platform on which they had fewer followers if they thought their post was toxic? This study did not have data on whether there was an inverse relationship between toxicity and the number of followers.

Another limitation is that certain key variables—such as toxicity, self-esteem, and perceived social approval—were measured rather than manipulated. Although there was no clear evidence for multicollinearity in the regression analyses, the correlation between self-esteem and perceived social approval ($r = .185$, $p = .002$; see Table 1) was undesirable. Future research should try to manipulate key moderators within an experimental

paradigm instead of measuring them survey-style, so as to ensure the statistical orthogonality of the moderators.

Finally, with a college sample, it is possible that some of our participants follow each other on social media. As such, they could be exposed to a number of very similar-looking posts when the study was running. Again, this may have influenced the results (and perhaps perceived social approval) in unknown ways—although random assignment should have evenly distributed the participants who followed or did not follow other participants across the various conditions.

Conclusion

The present study revealed that there are some limits regarding the extent to which publicly sending online messages can influence the message senders themselves. More precisely, it showed that there are some costs to keeping up with the positivity bias or the culture of confidence that are prevalent on social media (Orgad and Gill, 2021; Wahl-Jorgensen, 2018). Expressions of positivity can backfire on content creators if the content creators have low self-esteem or perceive toxicity in what they have created (i.e. deem that their expressions of positivity suppress some negative aspect of reality, such as their negative emotions).

At a wider academic and social level, there seems to be a rethinking of what it means to be positive or to self-present in a positive manner. Positive psychology researchers already recognize the benefits of supposedly negative emotions in fostering gratitude, thereby promoting overall well-being (Emmons and Stern, 2013). Popular psychology circles have started to discuss toxic positivity, or how (re)framing things as positive can be a poor way to emotionally cope (Lukin, 2019). Another parallel can be found in female body image research, where some evidence suggests that rather than striving for body positivity (loving one's body whatever its physical form may be), body neutrality (acknowledging that one's worth is not found in one's bodily appearance, that one should not have to *love* one's own body, and that *accepting* one's body is sufficient, etc.) is better for mental health (Cohen et al., 2021). The present study is one small step in synchronization with this broad research thrust, which should gradually become less disparate as future research produces more and clearer evidence.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

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

Andrew J Flanagin https://orcid.org/0000-0002-0928-5861

Supplemental material

Supplemental material for this article is available online.

Notes

1. Because the term "likes" has entered common parlance, it will be used as a generic term referring to paralinguistic digital affordances in general, as opposed to being a specific Facebook feature.
2. An analysis of variance (ANOVA) showed that participants' age did not differ by valence or volition. Chi-square tests also showed that participants' sex and ethnicity did not differ significantly by valence or volition. Insofar as these demographic variables are concerned, random assignment to condition was successful.
3. As participants were permitted to choose where they posted to, the persistence/ephemerality of their posts (as predictors of enjoyment/commitment) may have confounded results. To address this concern, the various posts were classified as either "persistent" or "ephemeral" and thereafter independent samples *t*-tests (for Part 1) and repeated-measures ANOVAs (for Part 2; with persistence/ephemerality as a between-subjects factor and time as a within-subjects factor) were performed to test the effect of persistence/ephemerality on enjoyment and comment. No significant effects were found for all analyses. Nine participants who shared on WeChat or Weibo—both of which cannot be neatly categorized into either persistent or ephemeral—were not analyzed for these tests (but were included in hypothesis testing).
4. As participants were permitted to choose where they posted to, the independent variables (valence and volition) could have influenced the persistence/ephemerality of their posts. The breakdown between valence and persistence/ephemerality was as follows: $n = 79$ for positive caption & ephemeral post, $n = 64$ for positive caption & persistent post, $n = 81$ for neutral caption & ephemeral post, $n = 75$ for neutral caption & persistent post. The breakdown between volition and persistence/ephemerality was as follows: $n = 87$ for volitional caption & ephemeral post, $n = 62$ for volitional caption & persistent post, $n = 73$ for non-volitional caption & ephemeral post, $n = 77$ for non-volitional caption & persistent post. A logistic regression showed that valence, volition, and the valence × volition interaction did not significantly influence persistence/ephemerality. Like in Note 3, the nine participants who shared on WeChat or Weibo were excluded from this test.

References

Aronson E (1999) The power of self-persuasion. *American Psychologist* 54(11): 875–884.

Baikie KA and Wilhelm K (2005) Emotional and physical health benefits of expressive writing. *Advances in Psychiatric Treatment* 11(5): 338–346.

Carr CT and Hayes RA (2019) Identity shift effects of self-presentation and confirmatory and disconfirmatory feedback on self-perceptions of brand identification. *Media Psychology* 22(3): 418–444.

Carr CT, Hayes RA and Sumner EM (2018) Predicting a threshold of perceived Facebook post success via likes and reactions: a test of explanatory mechanisms. *Communication Research Reports* 35(2): 141–151.

Carr CT, Kim Y, Valov JJ, et al. (2021) An explication of identity shift theory. *Journal of Media Psychology* 33(4): 202–214.

Cohen R, Newton-John T and Slater A (2021) The case for body positivity on social media: perspectives on current advances and future directions. *Journal of Health Psychology* 26(13): 2365–2373.

Dalgleish T, Yiend J, Schweizer S, et al. (2009) Ironic effects of emotion suppression when recounting distressing memories. *Emotion* 9(5): 744–749.

Deci EL and Ryan RM (2012) Self-determination theory. In: Van Lange PAM, Kruglanski AW and Higgins ET (eds) *Handbook of Theories of Social Psychology*. New York: Sage, pp. 416–436.

Emmons RA and Stern R (2013) Gratitude as a psychotherapeutic intervention. *Journal of Clinical Psychology* 69(8): 846–855.

Fivush R, Booker JA and Graci ME (2017) Ongoing narrative meaning-making within events and across the life span. *Imagination, Cognition and Personality* 37(2): 127–152.

Fox J, Vendemia MA, Smith MA, et al. (2021) Effects of taking selfies on women's self-objectification, mood, self-esteem, and social aggression toward female peers. *Body Image* 36: 193–200.

Gonzales AL and Hancock JT (2008) Identity shift in computer-mediated environments. *Media Psychology* 11(2): 167–185.

Goodall K (2015) Individual differences in the regulation of positive emotion: the role of attachment and self esteem. *Personality and Individual Differences* 74: 208–213.

Goodman W (2022) *Toxic Positivity: Keeping It Real in a World Obsessed with Being Happy*. New York: TarcherPerigee.

Graci ME and Fivush R (2017) Narrative meaning making, attachment, and psychological growth and stress. *Journal of Social and Personal Relationships* 34(4): 486–509.

Gross JJ (2002) Emotion regulation: affective, cognitive, and social consequences. *Psychophysiology* 39(3): 281–291.

Hayes RA, Carr CT and Wohn DY (2016) One click, many meanings: interpreting paralinguistic digital affordances in social media. *Journal of Broadcasting & Electronic Media* 60(1): 171–187.

Heatherton TF and Polivy J (1991) Development and validation of a scale for measuring state self-esteem. *Journal of Personality and Social Psychology* 60(6): 895–910.

Hoplock LB, Stinson DA, Marigold DC, et al. (2019) Self-esteem, epistemic needs, and responses to social feedback. *Self and Identity* 18(5): 467–493.

Joshanloo M (2022) Self-esteem predicts positive affect directly and self-efficacy indirectly: a 10-year longitudinal study. *Cognition and Emotion* 36(6): 1211–1217.

Kernis MH and Goldman BM (2006) A multicomponent conceptualization of authenticity: theory and research. *Advances in Experimental Social Psychology* 38: 283–357.

Kille DR, Eibach RP, Wood JV, et al. (2017) Who can't take a compliment? The role of construal level and self-esteem in accepting positive feedback from close others. *Journal of Experimental Social Psychology* 68: 40–49.

Lane DS, Lee SS, Liang F, et al. (2019) Social media expression and the political self. *Journal of Communication* 69(1): 49–72.

Leary MR and Baumeister RF (2000) The nature and function of self-esteem: sociometer theory. In: Zanna MP (ed.) *Advances in Experimental Social Psychology*, vol. 32, pp. 1–62. Cambridge, MA: Academic Press.

Leary MR and Kowalski RM (1990) Impression management: a literature review and two-component model. *Psychological Bulletin* 107(1): 34–47.

Loman JG, Müller BC, Oude Groote Beverborg A, et al. (2018) Self-persuasion on Facebook increases alcohol risk perception. *Cyberpsychology, Behavior, and Social Networking* 21(11): 672–678.

Luke SG (2017) Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods* 49(4): 1494–1502.

Lukin K (2019) Toxic positivity: don't always look on the bright side. *Psychology Today*, August 1. Available at: https://www.psychologytoday.com/us/blog/the-man-cave/201908/toxic-positivity-dont-always-look-the-bright-side

Orgad S and Gill R (2021) *Confidence Culture*. Duke University Press.
Pennebaker JW (1997) Writing about emotional experiences as a therapeutic process. *Psychological Science* 8(3): 162–166.
Pennebaker JW and Chung CK (2011) Expressive writing: connections to physical and mental health. In: Friedman HS (ed.) *Oxford Handbook of Health Psychology*. Oxford: Oxford University Press, pp. 417–437.
Pugh SD, Groth M and Hennig-Thurau T (2011) Willing and able to fake emotions: a closer examination of the link between emotional dissonance and employee well-being. *Journal of Applied Psychology* 96(2): 377–390.
Rimé B (2009) Emotion elicits the social sharing of emotion: theory and empirical review. *Emotion Review* 1(1): 60–85.
Schlenker BR (1986) Self-presentation: toward an integration of the private and public self. In: Baumeister RF (ed.) *Public Self and Private Self*. Cham: Springer-Verlag, pp. 21–62.
Sherman DK (2013) Self-affirmation: understanding the effects. *Social and Personality Psychology Compass* 7(11): 834–845.
Singer JA (2004) Narrative identity and meaning making across the adult lifespan: an introduction. *Journal of Personality* 72(3): 437–459.
Slatcher RB and Pennebaker JW (2006) How do I love thee? Let me count the words: the social effects of expressive writing. *Psychological Science* 17(8): 660–664.
Swann WB Jr (2011) Self-verification theory. In: Van Lange PAM, Kruglanski AW and Higgins ET (eds) *Handbook of Theories of Social Psychology*. New York: Sage, pp. 23–42.
Valkenburg PM (2017) Understanding self-effects in social media. *Human Communication Research* 43(4): 477–490.
Valkenburg PM and Peter J (2008) Adolescents' identity experiments on the Internet: consequences for social competence and self-concept unity. *Communication Research* 35(2): 208–231.
Van House NA (2011) Personal photography, digital technologies and the uses of the visual. *Visual Studies* 26(2): 125–134.
Wahl-Jorgensen K (2018) The emotional architecture of social media. In: Papacharissi Z (ed.) *A Networked Self and Platforms, Stories, Connections*. New York: Routledge, pp. 77–93.
Walther JB and Lew Z (2022) Self-transformation online through alternative presentations of self: a review, critique, and call for research. *Annals of the International Communication Association* 46(3): 135–158.
Walther JB, Liang YJ, DeAndrea DC, et al. (2011) The effect of feedback on identity shift in computer-mediated communication. *Media Psychology* 14(1): 1–26.
Winter S, Remmelswaal P and Vos A (2022) When posting is believing: adaptation and internalization of expressed opinions in social network sites. *Journal of Media Psychology* 34(3): 177–187.
Wood AM, Froh JJ and Geraghty AW (2010) Gratitude and well-being: a review and theoretical integration. *Clinical Psychology Review* 30(7): 890–905.
Wood JV, Elaine Perunovic WQ and Lee JW (2009) Positive self-statements: power for some, peril for others. *Psychological Science* 20(7): 860–866.

Author biographies

Zijian Lew is an Assistant Professor at the Wee Kim Wee School of Communication and Information, Nanyang Technological University, Singapore. He researches various topics in web-based contexts.

Andrew J Flanagin is a Professor in the Department of Communication at the University of California, Santa Barbara, where he is a former Director of the Center for Information Technology and Society. His work broadly considers processes of social influence in digitally-mediated environments.