Goodbye Likes, Hello Mental Health: How Hiding Like Counts Affects User Behavior & Self-Esteem

Roy Klaasse Bos*
Tilburg University, The Netherlands
r.j.klaassebos@tilburguniversity.edu

Likes are widely available on social network services and are known to influence people's self-image. An emerging literature has started to look at potential detrimental effects of social media use among teenagers. We study how Instagram users' posting frequency, variety, like behavior, and relative self-esteem are affected by an intervention in which like counts were hidden in selected treatment countries. Using a unique panel data set of individual users' Instagram posts across multiple years, we find evidence that users posted more frequently and more varied than in the months prior to the intervention. On the other hand, the number of likes decreases as people are no longer influenced by others' evaluations, especially among users with a small following. Further, in an experiment we show that the number of likes people see on others' posts affects their relative self-esteem, and that users are more likely to self-disclose once they rate themselves more positively. These results are critical to understanding the dynamics on visual-based social media in order to foster a healthy online environment.

Publication Date: August 2020
Keywords: Visual social media, self-disclosure, variety, likes, self-esteem
Supplemental Material: Data and the online appendix are available on https://git.io/JJ4fT

1. Introduction

Teenagers seem to associate the number of likes on social media with their social status. Some of them even go as far as deleting posts that did not receive enough likes (Contrera 2016). Pruning and revising profile content is an important part of their online identity management. This is in line with the finding that surveillance and knowledge of others are the two most influential reasons of social network services usage (Sheldon and Bryant 2016). Instagram, a mobile photo and video-sharing social network service used daily by 72% of all teenagers in the United States, plays an integral role in shaping their online identities. In recent years more and more media have raised concerns that social network services contribute to low self-esteem and feelings of inadequacy among young users. Teenagers express concerns about bullying, addiction, peer pressure, and mental health issues, among others (Anderson and Jiang 2018). They take steps to shape their online reputation and manipulate their profile in order to garner the maximum number of likes (Madden et al. 2013). At the same time, likes have lost part of their value as they are no longer about expressing interest in a brand or individual but rather a loose way of confirming a friendship or expressing support to other teenagers (Kesvani 2019). Likes are more akin to a read receipt than an endorsement and not liking can be read as a sign of rejection. Liking everything, however, could mess up the algorithm which is key to Instagram's business model. More specifically, the platform stands or falls by targeted advertisements for which it is imperative to

* The author (Tilburg School of Economics and Management, Marketing Department, Student ID: 361482) is supervised by Hannes Datta (TiSEM) and Niels van de Ven (TiSEM).

accurately derive users' personal interests. Once teenagers become less selective in their liking behavior, this could turn into an issue for the newsfeed algorithm. Last year Instagram ran several experiments which tap into this matter. In particular, users in selected countries could not see like counts anymore. Although they could still view how many likes they received on their own posts, their followers were no longer able to see like counts. The tech company announced that hiding likes is designed to shift the focus from the number of likes to the photos and videos users share. It should result in a less pressurized environment where i) people feel comfortable expressing themselves (Conger 2019) and ii) users have to decide for themselves if something is worth liking rather than following the crowd (Constine 2019). Since Instagram never disclosed the results of the experiment, one of the main interests of this study is in understanding the consequences of the intervention on user behavior and self-esteem. More specifically, in the observational study we determine whether users indeed post more often, share more varied content, and like fewer posts after hiding like counts. In the randomized experiment we then provide additional support and measure the effects of likes on users' levels of relative self-esteem.

We contribute to the literature in three ways. First, we respond to calls for more research on the visual modality on social media (Babić et al. 2016; King et al. 2014). More specifically, we compare the similarity of shared pictures on Instagram before and after the intervention as well as with other related users. This perspective aims to shed light on the notion that visual similarity of pictures being posted on social media has increased as users mimic each other's pictures (Hidrėlėy 2019; Zhang 2018). To this end, we operationalize variety in both a within and between-subjects temporal study design context. That is, the variety of topics someone discloses over time and the degree of differentiation of self-disclosure compared to others, respectively. Using this approach, we conduct a content analysis of a unique panel data set of Instagram pictures to capture and compare the objects presented in those pictures. This allows us to disentangle the effects of hiding like counts on variety more accurately. In fact, conventional approaches cannot be used to examine image similarity in a between-subjects research design.

Second, we investigate the effect of hiding like counts on user behavior and self-esteem ng. This differs from earlier work in which participants were exposed to either a below average or above average number of likes (Burrow and Rainone 2017; Sherman et al. 2016; Vogel et al. 2014). Although these studies contribute to our understanding of the effect of the number of likes on peer influence, it does not explain the differences between those who can or cannot see the number of likes. This study, however, considers whether levels of self-esteem differ between participants exposed to a post with (control) or without (treatment) the number of likes displayed. Moreover, we examine whether people with high self-esteem are also more likely to post and like pictures on Instagram.

Third, we shed light on the effect of a platform imposed privacy policy change which hides previously available information to all users and thus investigate how powerful self-regulation of digital platforms really is. Although the role of privacy concerns and privacy value on self-disclosure has already been extensively investigated (Chen 2013; Dienlin and Metzger 2016; Zlatolas et al. 2015) these studies measure whether participants worry about their privacy online and how many privacy enhancing measures they have taken themselves. Our study complements earlier work on opt-out privacy mechanisms (Johnson, Shriver, and Du 2020) and the privacy calculus (Krasnova et al. 2012) but differs in the sense that a privacy design change is externally imposed to users. That is, Instagram determines which users can and cannot see like counts in the experiment. Thus, we investigate the power of social media in shaping user behavior and self-esteem through their privacy designs.

2. Hypothesis Development

First, we explore the effects of hiding like counts on posting frequency, variety, and like behavior. Second, we investigate the effects of the intervention on users' levels of self-esteem.

2.1. Posting Frequency

Getting enough likes on content is a major reason for continued social network use (Scissors et al. 2016; Smith 2014). The number of likes a person receives may indicate popularity online and visual self-representation is aimed at gaining acceptance by the peer group (Mascheroni et al. 2015). Conformity to beauty standards and peer conventions is rewarded with peer validation expressed by the number of likes and followers. Yet pressure to post content that will be popular and gets a lot of likes and comments is also one of the main dislikes of Facebook users (Smith 2014). More and more young users delete or archive original posts that did not receive many likes (Contrera 2016).

This may be driven by a phenomenon Scissors et al. (2016) refer to as the *like paradox*: on average people's friends receive roughly twice as many likes per post as they do. This can be attributed to the fact that their friends have more friends to provide those likes (Feld 2016). It is reasonable to assume that individuals use the average number of likes their friends get as a proxy to determine whether they get enough likes. In that case, most users would draw the conclusion that their friends get more likes and may be considered more popular than they are. Hiding like counts removes this external reference point and thereby hinders the negative consequences of social comparison.

At the same time, hiding like counts may remove the peer pressure to always look "perfect" in pictures to garner the maximum number of likes (Mascheroni et al. 2015). Ma et al. (2016) conducted an online experiment in which they demonstrate that anonymity increases self-disclosure by lowering the risk of harming one's personal image. Under anonymity people feel less constrained by the expectations of others. Although hiding like counts does not directly affect users' anonymity, it does add an extra layer of privacy. Privacy concerns and values in turn are associated with self-disclosure (Chen 2013; Dienlin and Metzger 2016; Zlatolas et al. 2015). Interestingly, Chen also found that the more concerned people are regarding their privacy, the more they will engage in acts of self-withdrawal such as deleting posts from Facebook. Thus, the effect of privacy on self-disclosure is two-fold. Less privacy concerned individuals will i) post more frequently and ii) are less likely to delete it after publishing. Hence, we formulate the following hypothesis:

HYPOTHESIS 1: *Posting frequency increases after the intervention of hiding like counts.*

2.2. Variety

Tiggeman and Zaccardo (2018) conducted a content analysis of images on Instagram marked with the hashtag fitspiration. They manually coded 600 images and found that most images were posed shots and featured a specific body part. The range of body types was only limited: the great majority of men was of average build and had a high level of muscularity while most images of women featured a thin and toned body. Similarly, Hum et al. (2011) studied Facebook profile images and identified a consistent pattern: most profiles were inactive, posed, appropriate, and only included the subject. This is in line with the finding that conformity to beauty standards is rewarded with peer validation (Mascheroni et al. 2015). For instance, they found that girls posting provocative photos chose to conform to a sexualized stereotype as a means of being socially accepted by peers. Others have attempted to identify visual attributes that can predict virality (Deza and Parikh 2015; Tellis et al. 2019). It turns out that specific emotions, objects, and environments are more likely to drive virality. With this in mind, we expect that users may have capitalized on these trends by sharing like-seeking analogous images in order to garner the maximum number of likes. Hiding like counts, however, may encourage them to share more authentic content across a wider spectrum as getting enough likes becomes less important. Thus, we propose the following:

HYPOTHESIS 2: *Variety of posted visual content increases after the intervention of hiding like counts.*

2.3. Like Behavior

Herd behavior refers to the phenomenon of people following a crowd regardless of individual information suggestion something else (Banerjee 1992). Individuals who follow the crowd expect to make fewer errors and decision-making requires less mental effort (Lee et al 2015). For instance, Bikhchandandi et al. (1992) consider the submission of a paper to a journal. In case a reviewer learns that the paper was rejected at two previous journals the chance of acceptance decreases. This is based on the theory of informational cascade that says that individuals fear of loss of reputation when dissenting from the majority opinion of earlier responders. This theory has also been applied to marketing and management studies. Chen (2008) demonstrates that consumers use product choices of others as cues in making online purchase decisions. More specifically, sales volume and star ratings induce purchase intentions. Hanson and Putler (1996) tested an experimental design in which the download count of a software program was upwardly manipulated and concluded that it leads to higher online product popularity.

Sherman et al. (2016) conducted a study on the relationship between popularity and the way a photo was perceived. Specifically, adolescents were more likely to like a photo that had received many likes from peers than photos with few likes. A higher like count was interpreted as a signal of popularity which increased users' odds of liking a post as people tend to imitate peers when they do not know the proper behavior for a certain situation. At the same time, Lee et al. (2015) demonstrate that the need to adhere to the majority opinion is less salient when fewer opinions are expressed. Like Moe and Schweidel (2012), they find that consumers are more likely to post an opinion when the ratings already posted are more positive. Thus, individuals tend to rely on the crowd's opinion when making a choice or giving a rating, as this makes them less prone to errors or reputation damage and requires fewer mental resources.

Receiving many likes signals popularity and thus induces users to engage with the post. Since users can no longer assess how well a post is doing, the intervention may lead to less herd behavior and more differentiation behavior. They will not be influenced by others' evaluation and are, therefore, less likely to like a post which would have normally received many likes. Therefore, we hypothesize that users' tendency to like photos and videos without like counts declines:

HYPOTHESIS 3: *The number of likes decreases after the intervention of hiding like counts.*

2.4. Self-Esteem

The like count on social networks allows for easy impression management and provide high comparison standards (Appel et al. 2015). Social comparison plays an important role in explaining potential negative consequences that can arise. In fact, more frequent Instagram use is associated with greater depressive symptoms through the mechanism of social comparison, and moderated by the share of strangers one follows (Feinstein et al. 2013; Lup et al. 2015). Likewise, more frequent Facebook use is related to lower trait self-esteem which is mediated by a greater exposure to upward social comparison (Vogel et al. 2014). Specifically, participants saw a greater discrepancy between the target person and themselves in a high like condition compared to a profile with a low number of likes. Ratings of the self and target, however, did not differ in the low like count condition. That is to say, downward social comparison did not affect levels of self-esteem. Viewing a high number of likes and comments, on the other hand, deflates self-evaluations and state self-esteem.

Furthermore, there is supporting evidence that low self-esteem and depression are strongly related (Battle 1978; Battle et al. 1988; Orth et al. 2008; Ozmen et al. 2007). A more recent meta-analysis (Sowislo and Orth 2013) of 77 longitudinal studies providing information about the temporal order of the relationship between self-esteem and depression concluded that low self-esteem contributes to depression and not vice-versa. Accordingly, they recommend increasing self-esteem in order to reduce the risk of depression.

Although social comparison on social network services can make people feel depressed, online interactions can users also make feel more connected (Yang 2016) and improve their self-esteem (Burrow and Rainone 2017). Thus, it is not evident whether like counts contribute to mental issues or not. At the same time, it could well be the case that the potential positive impact of interactions cancels out because users feel that their social connectedness is not as satisfactory as others'. In that light, it is interesting to consider what can explain someone's tendency to make social comparisons. Kleemans et al. (2018) suggest that individuals prefer to make social comparisons to similar others. That is, social comparison may be stronger when perceived similarity is high (Suls et al. 2002). For instance, teenage girls are more likely to compare themselves with females of a comparable age. Along the same line of reasoning, Fardouly and Vartanian (2015) argue that the positive relationship between Facebook usage and body image concerns among female university students, is mediated by upward comparisons to celebrities and distant peers. Participants judged their own appearance to be less attractive because a lack of personal contact made it difficult for people to accurately gauge how realistic the appearance of celebrities and distant peers are on Facebook. Indeed, Lup et al. (2015) find evidence that the amount of strangers followed significantly moderates the association of Instagram use with social comparison and the indirect relationship between Instagram use with depressive symptoms through social comparison. At the lowest level of strangers followed only, more frequent Instagram use was associated with more positive social comparison, whereas at the highest levels of strangers followed higher usage had direct associations with greater depressive symptoms. Chou and Edge (2012) find similar results suggesting that users who included more people they did not personally know as their Facebook friends more strongly believed that others had a better life. Users were more likely to exhibit attribution error towards strangers. Personal interactions and knowing how people actually live avoid the correspondence bias. That is, the tendency to assume that others' actions and words reflect their personality or stable personal disposition rather than being affected by situational factors. In other words, Instagram users might conclude others are happy when they see their joyful pictures posted on social media without considering the circumstances that made them happy.

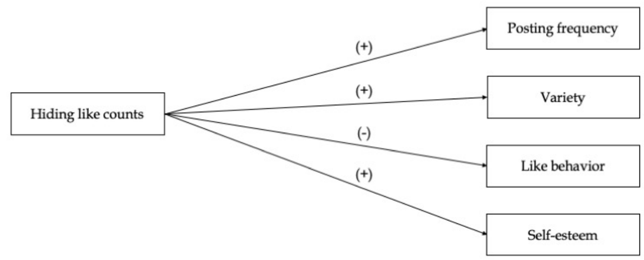
In sum, it would be expected that users who observe high like counts are more likely to make upward social comparisons which lowers their self-esteem. A direct consequence of hiding like counts is that users can no longer make social comparisons based on the number of likes. Therefore, a reduced tendency to engage in social comparison can explain an improvement of one's self-esteem:

HYPOTHESIS 4: *Hiding like counts is associated with higher levels of self-esteem.*

2.5. Conceptual Model

Thus far, we have formulated several hypotheses about how hiding like counts affects posting frequency, variety, like behavior, and self-esteem. Figure 1 presents the conceptual model that underpins this study. Next, we will explain how we measure these constructs and investigate corresponding hypotheses.

Figure 1. Conceptual model



Notes. The "+" and "-" symbols indicate positive and negative hypothesized relationships respectively. Self-esteem refers to the relative self-evaluation compared to a target person.

To measure posting frequency on Facebook Hollenbaugh and Ferris (2014) used a modified version of Revised Self-Disclosure Scale (RSDS). Variety of posted content was measured through a five-item scale (e.g., "My Facebook posts range over a wide variety of topics"). However, this questionnaire-based approach is not ideal to measure self-disclosure in a temporal research design since we want to investigate posting frequency and variety in the months prior to the intervention of hiding like counts as well as the period thereafter. Moreover, the fact that posts cover a wide variety of topics before and after the intervention can mean one of two things: users shared different types of images and nothing changed, or they adopted a different photography style that involves many yet different topics. Another challenge lies in the operationalization: variety can be defined within an individual but also across groups of individuals. That is, the variety of topics someone discloses over time and the degree of differentiation of self-disclosure compared to others, respectively. In order to explore differences between subjects it does not suffice to ask about variety of posted content, we need to capture and compare the objects presented in the images of both individuals. Thus, to accurately disentangle the effects on variety we need to conduct a content analysis of the posted images. Therefore, we have collected observational data from Instagram across multiple years to examine the effects of hiding like counts on variety of posted images. Although observational data provides a novel perspective and demonstrates external validity, it has two major limitations. First, we cannot measure nor make any causal claims about the effects on users' self-esteem as we cannot randomly assign Instagram users into test conditions. Second, it is unknown whether individuals in the control group were exposed to downward or upward social comparison stimuli once they saw the post. An experimental design in which participants are randomly assigned to a photo with below average, above average, or no likes overcomes these problems and allows us to measure self-esteem in a controlled environment. Although we asked participants in the experiment to imagine the user on the photo was their neighbor, it remains a hypothetical question whether they would engage with the post or not. Therefore, we also exploit our observational data to measure like behavior. More specifically, we expect that the number of likes decreases as users can no longer draw on others' evaluations. In summary, we collect post-level data from Instagram to measure changes in posting frequency, variety, and user engagement in response to hiding like counts. In addition, we conduct a randomized experiment to provide additional support for the relationships we found in our observational study and overcome aforementioned limitations.

Table 1. Mapping of hypotheses

	Construct	Observational study	Experiment
H1	Posting frequency	X	X
H2	Variety	X	
H3	Like behavior	X	X
H4	Self-esteem		X

3. Observational Study

We now describe the data we have collected to shed light on how the intervention of hiding like counts changes posting frequency (H1), variety (H2), like behavior (H3), and self-esteem (H4) on Instagram. Thereafter, we discuss the method and results related to these hypotheses.

3.1. Institutional Background

Late April 2019 Instagram announced that it would run an experiment among Canadian users in which the like counts were hidden (Constine 2019). Three months later, around mid-July, they expanded the treatment to users in various other countries including Australia, Canada, and Italy. Users located in these countries could not see the number of likes on media posted by others, whereas users living anywhere else could still view like counts (Loren 2020). Thus, treatment groups enter the treated pool of persons sequentially, and assignment to the treatment or control condition was dependent on users' geography. To construct our sample, we therefore identified users' country of origin which we explain next.

3.2. Identification Strategy

Since user geography is only available on an aggregated level by influencer, we applied the following sampling approach to derive individuals' country of origin (detailed procedure in Web Appendix B). Starting from a seed of 13,451 top influencers from 21 countries (HypeAuditor, 2020), we obtained the percentage of influencers' followers by country. From this sample we selected influencers of which most followers were concentrated in the same country of origin. If those followers were settled in a country part of the treatment group, they would not be able to observe like counts on others' posts after the introduction of the intervention. Likewise, we identified influencers whose following was predominantly composed of users in one of the countries in the control group. For each selected influencer, we drew a random sample of followers and validated their country of origin. Although we acknowledge that this limits our sample to users who follow influencer accounts, we believe this includes the great majority of Instagram users as Lup et al. (2015) found that almost half of the Instagram accounts 18-29 year-olds follow are celebrities. This can be attributed to the platform design in which following someone only goes in one direction. Still, consumers who follow many influencers are more likely to be sampled. To counteract the risk of oversampling users from specific niches, we choose our top influencers across a wide variety of categories. Our sample of influencers operate in 18 diverse categories ranging from how-to-style to cars:

Table 2. Overview of Instagram categories

Beauty	Entertainment	Pets
Beer & Wine	Fitness	Photography
Cars	Home & Garden	Restaurants
Children	How-To-Style	Sports
Clothes	Movies & TV	Travel
Education	Music	Videogames

Notes. We obtained the full list of categories in February 2020. Influencers are assigned to one or more categories based on their bio and posts.

In addition, we validated whether sampled users matched our sampling criteria. First, users should have a public profile so that we could access their Instagram posts. We argue that if private accounts would differ from public accounts, they are likely more privacy concerned than users whose profile is open to the public. As such, potential effects of hiding like counts on posting frequency (H1) and variety (H2) in our sample of public accounts are probably more pronounced among privacy concerned individuals as they have deliberately chosen to take privacy enhancing measures themselves already. Second, users should have posted at least 50 pictures of which at least 5 prior to and after the intervention. This ensures individuals are active Instagram users, and we can compare their user behavior before and after hiding like counts. Third, users were followed by up to 5000 other users since this study centers around the effect on everyday consumers rather than influencers. In fact, influencers may be more involved in how hiding like counts affects them than consumers as social media is part of their work. Fourth, accounts are not used for business purposes for the same reason as above. Fifth, we validated users' country of origin through location tags and language usage in post captions. We repeated this process until we gathered 40 Instagram profiles in each of the following countries: Australia, Canada, France, Germany, Italy, Netherlands, Spain, and the United Kingdom.

In sum, we have collected data from 320 consumer accounts spread across 3 treatment and 5 control countries. For each account we obtain user and post information (Table 3), and we store a link to the photo or video associated with each post. Hereafter we explain how these characteristics relate to the hypotheses.

Table 3. Instagram Users and Posts

	Mean	Median	SD	Min.	Max.
User level					
#*Posts*	347	199	494	50	5,169
#*Followers*	848	580	777	62	4,239
#*Follows*	892	682	833	29	7,197
Post level					
#*Likes*	38.38	21.00	86.79	0.00	22,187
#*Comments*	1.99	1.00	5.08	0.00	978

Notes. The full sample consists of 320 users and 111,011 posts obtained in July 2020.

3.3. Measurement of Constructs

Below we describe how we operationalize the constructs relevant for the observational study.

3.3.1. Posting Frequency

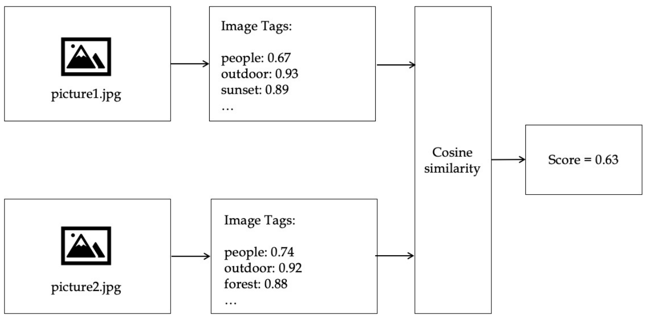
We measure the number of photo and video posts each user shares in a month on Instagram. Since users enter the treatment pool sequentially, we have defined the posting frequency relative to the introduction of the intervention. That is, the number of months since likes have been hidden for a given user. Along the same line of reasoning, we set the time frame to include 12 months of data before and after the intervention. For instance, for Canadians this period runs from April 2018 to April 2020.

3.3.2. Variety

As we explained earlier, we conduct a content analysis on collected Instagram images. This allows us to study variety within and between subjects. That is, the variety of topics someone discloses over time and the degree of differentiation compared to others. Given the sheer amount of posts (Table 3), we are not able to code all images manually. Instead, we use Azure Cognitive Services Computer Vision Application Programming Interface (API). Although computer vision methods are rarely used in the field of marketing and an in-depth assessment of the accuracy of the API is beyond the scope of this article, some evidence of the validity can be inferred. Rietveld et al. (2020) underline the reliability of computer vision APIs after detecting brand logos, objects, and faces in Instagram images. Next, we provide a high-level description of the data returned by the API and how we convert the raw data into a measure for variety. Details and examples are available in Web Appendix E.

For every image, the API returns a vector of detected objects (i.e., image tags) which we use to compute image similarity within and between-subjects[1]. More specifically, we calculate the cosine similarity between each pair of image vectors to judge similarity (Figure 2). The higher the cosine similarity, the more similar two pictures are. Since the type of image content is often temporally dependent (e.g., Christmas), we again set the time frame to include a year of images prior to and after likes have been hidden. This way we control for time-varying factors that could distort the comparison.

Figure 2. Cosine Similarity



Notes. The computer vision API returns a vector of image tags and confidence scores for each image. For instance, the computer vision algorithm is 67% confident that people are portrayed in picture 1. Given these image tags and confidence scores, we compute the cosine similarity for each pair of images. This is a widely used metric to measure how similar two documents are irrespective of their size. That is, the number of image tags can be different for picture 1 and 2. Also, the image tags themselves can vary from one picture to another (e.g., "sunset" only appears in picture 1). Confidence scores are captured through the cosine similarity operation which measures the angle between two vectors and determines whether two vectors are pointing in the same direction. More precisely, we replace the confidence scores of missing image tags with zeros (e.g., "forest" in picture 1) and multiply the confidence scores of pictures 1 and 2 for each image tag (e.g., for people: 0.67 x 0.74) and divide by the multiplication of the length of both vectors. Mathematically, this can be denoted as: $sim(r,c) = (r \cdot c)/(||r|| \cdot ||c||)$ where *r* and *c* are the image vectors for picture 1 and 2 respectively, and $||r||$ is defined as $\sqrt{r_1^2 + r_2^2 + \ldots + r_n^2}$ (Knox and Datta 2020). A larger confidence score thus has a larger weight and more overlapping image tags gives a higher cosine similarity score.

First, we make within-subject comparisons to address whether the variety of posts changes after the introduction of the intervention. Second, we make between-subject comparisons to determine whether treated users share more unique content relative to others.

Within-subject

For each user i we compute the cosine similarity between pictures taken by the same user i. We distinguish between pictures taken before ($1_{before} \ldots n_{before}$) and after ($1_{after} \ldots n_{after}$) the intervention. This yields a similarity matrix in which each picture *before* [*after*] the intervention is compared with all pictures *before* [*after*] hiding like counts (i.e., white squares in Figure 3). Each row (r) and column name (c) presents a picture from user i in k, where k can take on the value *before* or *after.* Given these two separate subsets k, we calculate how similar each picture is to all other pictures in the same subset. That is, for each row we take the row average excluding the diagonal values. Finally, we aggregate the results across all rows in k:

$$\varpi_{ik} = \frac{1}{n_{ik}(n_{ik}-1)} \sum_{r=1}^{n} \sum_{c=1 \mid c \neq r}^{n} sim(r_{ik}, c_{ik})$$

Thus, for each user i we obtain the within-subject similarity before ($\varpi_{i,before}$) and after ($\varpi_{i,after}$) hiding like counts. In the follow-up analysis we determine whether the difference between those two values differs for the treatment and control group. Specifically, a drop in the within-subject similarity after the intervention would indicate that users shared more varied content[2].

Between-subjects

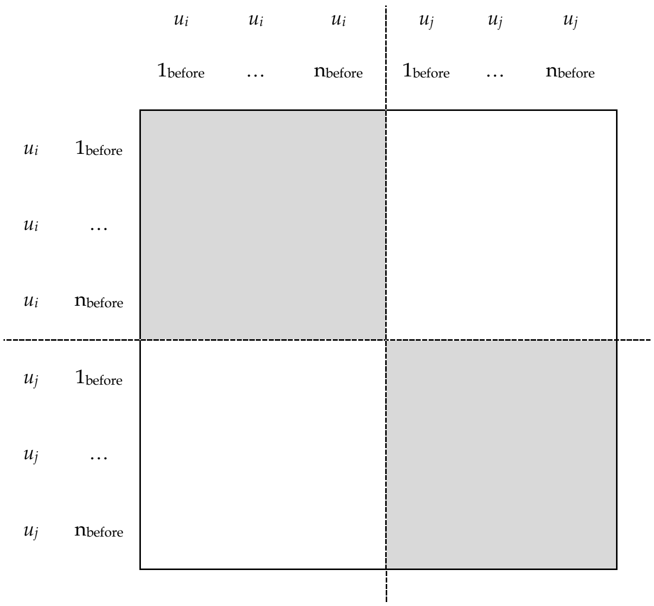
To assess between-subjects similarity ($\mathfrak{B}$) we distinguish between cohorts of users in the treatment and control group. We choose for these comparisons for two reasons. First, Instagram users may especially stay on top of the trends in their local market and therefore their postings might have already been more like other treatment units prior to the intervention. Second, by defining cohorts we establish more homogeneous clusters of users. Within these two cohorts, we determine the cosine similarity of each user pair (u_i, u_j) in k = {*before*, *after*} (i.e., white squares in Figure 4). That is, how similar pictures from user i are to pictures from another user j on average, where u_i and u_j belong to the same cohort.

Figure 3. Within-Subject Similarity Matrix

	1_{before}	…	n_{before}	1_{after}	…	n_{after}
1_{before}	1.00					
…		1.00				
n_{before}			1.00			
1_{after}				1.00		
…					1.00	
n_{after}						1.00

Notes. The row and column names represent pictures *before* and *after* the intervention for user i (u_i). Values in the matrix denote the cosine similarity for each picture pair (only the diagonal of 1s have been reported). For the purpose of this analysis we restrict ourselves to the white squares in the top left ($1_{before} \ldots n_{before}$) and bottom right ($1_{after} \ldots n_{after}$) quadrant of the figure. Within these areas we compute row means (excluding the diagonal values) to determine how similar a given picture is to all other pictures in the same subset on average. Thereafter, we derive the *before* and *after* within-subject similarity by taking the average of the row means in the top left and bottom right squares, respectively. Note: calculating column means, rather than row means, yields identical outcomes.

Figure 4. Between-Subjects Similarity Matrix



Notes. The row and column names represent pictures *before* the intervention for user *i* (u_i) and user *j* (u_j) in the same cohort. Values in the matrix denote the cosine similarity for each picture pair (note: values are left out for simplicity). For the purpose of this analysis we restrict ourselves to the top right or the bottom left white square. Within this area we compute row means to determine how similar a given picture from u_i [u_j] is to all pictures from u_j [u_i] on average. Thereafter, we sum up the row means and divide by the number of rows (n_{ik} [n_{jk}]) to derive the *before* between-subjects similarity ($\mathfrak{B}_{i,j,before}$). In a similar fashion, the *after* between-subjects similarity ($\mathfrak{B}_{i,j,after}$) can be determined. Note that only one of both white squares should be used to avoid duplicates.

Mathematically, this can be denoted as follows:

$$\mathfrak{B}_{ijk} = \frac{1}{n_{ik}\, n_{jk}} \sum_{r=1}^{n} \sum_{c=1}^{n} sim(r_{ijk}, c_{ijk})$$

Thereafter, we determine whether the change in cosine similarity of the cohort of treatment unit pairs differs from the cohort of control unit pairs.

3.3.3. Like Behavior

As a result of hiding like counts, we expect herd behavior to decrease which follows from a decline in users' tendency to like a post. The Instagram content lifespan is 48 hours on average which implies that posts do not receive much engagement after two days. Therefore, we can use like counts recorded in July 2020 for the analysis of like behavior across all historical posts. We measure the log mean number of likes per post each user gets in a month. Furthermore, we control for the number of followers and opt for a time frame of 1 year before and after the intervention to control for temporally dependent factors in the image content.

3.4. Method

3.4.1. Comparison of Treated and Control Groups

The data generation process in the observational study lacks a randomized assignment of Instagram users into treatment and control conditions. More specifically, Australians, Canadians, and Italians were part of the treatment group, whereas users in other countries were not. It could therefore be the case that treatment units vary from control units in ways we have not accounted for, which may bias our estimates. For instance, differences in when treatment and control users joined Instagram (i.e., adoption speed) might cause changes in how both groups use the platform. As such, we use a quasi-experimental matching procedure to match treated units with similar control units such that the distribution of observed baseline covariates will be similar between treated and control units (Austin 2011). In addition, posting frequency, variety, and like behavior could already differ between treatment and control countries prior to the intervention of hiding like counts. The difference-in-differences (DiD) method can be applied in these settings in which the treatment and control units differ because of time-invariant unobserved characteristics (Deschenes et al. 2018). That is, DiD compares average pre- and post-treatment differences in the outcome of a treatment and control group. We now explain how we use the propensity score method (PSM) to construct our matched sample for the DiD approach.

3.4.2. Propensity Score Estimation and Matching

The PSM allows researchers to construct counterfactuals using observational data (Li 2013). This reduces biases due to a lack of distribution overlap and different density weighting. Contrary to regression analysis the PSM adjusts the distribution of the groups and is less susceptible to violate model assumptions such as the functional form of covariates. To that end, we rebalance our data through matching non-treated users to treated ones on similar covariate values. First, we estimate a probit model of receiving treatment on the number of followers, how many accounts a user follows (i.e., followings), the adoption speed of Instagram use, and the percentage of image posts relative to all types of media posts (e.g., videos). We pick these variables for the following reasons. Users with a larger audience may experience more pressure in self-disclosure, a higher following count suggests that a user follows more other accounts and thus is more likely to engage in social comparison, early adopters have more experience using Instagram and therefore may be less likely to break out of recurring patterns, and video posts differ from image posts in the sense that the number of views as opposed to the number of likes is hidden which could affect the impact of the intervention on like behavior[3]. Second, we compute the Mahalanobis distance for each treated and control user pair and select unique matches sequentially, in order of closeness of their Mahalanobis distances. We matched without replacement such that control units are only allowed to be used as a match once. This approach is recommended provided that there are enough good matches (Li 2013). Each treatment unit was matched with a single control unit as a higher number of matches deteriorated matching quality. That is, after matching with multiple units more covariates significantly differed between the treatment and control group. Third, we conduct an imbalance check before and after matching (Table 4). From the matching results it follows that treatment units follow more users and adopted Instagram earlier than control units. However, after matching we find that the mean covariate values in the control group are comparable to the matched treatment group. That is, none of the covariate values differ significantly between the treatment and control group after matching. To control for differences in any other factors we did not capture with PSM, we added user fixed effects in the DiD model, which we explain next.

Table 4. Propensity Score Matching Results

	Treatment	Control
Before matching		
Followers	854.80	832.67
Followings	914.06*	864.42*
Adoption speed	2134.50***	1920.50***
Image share	0.87	0.87
After matching		
Followers	854.80	879.38
Followings	914.06	915.89
Adoption speed	2134.50	2110.10
Image share	0.87	0.85

Notes. Mean covariate values in the treatment and control group before and after matching. The final sample consists of 238 Instagram users of which 119 treatment and 119 control units. We computed the Mahalanobis distance for each treated and control pair for the follower count, following count, adoption speed (i.e., number of days since first post), and the percentage of image posts. Then, we matched each treatment unit with a control unit sequentially, without replacement, in order of closeness of their Mahalanobis distances (Web Appendix G). Matching quality was assessed by the statistical significance of the difference between the mean covariate scores in the treatment and control group. We run 1000 bootstrap samples to compute these p-values where *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

3.4.3. Difference in Differences

We use a DiD approach to estimate the effect of hiding like counts on posting frequency and like behavior. We compare the outcome measures of Instagram users in the treatment countries with those in control countries. As Canadians enter the treatment group prior to Australians and Italians, we estimate a DiD model in which the time in months (t) is relative the introduction of the intervention. For instance, t = 0 corresponds with April 2019 (Canadians) and July 2019 (Australians and Italians). To account for long-term trends in user behavior we add a trend variable to the equation. Taken together, we propose the following model:

$$Y_{it} = \alpha_i + \mathrm{I}_t + \mathrm{I}_t \cdot \tau_i + \gamma_t + \varepsilon_{it}$$

where Y_{it} is the dependent variable for user i at time t, α_i is either a user-level fixed or random effect to control for time-invariant user characteristics[4], I_t is 1 if the intervention was implemented at time t and 0 otherwise, τ_i is 1 of if user i was assigned to the treatment group and 0 otherwise, γ_t a trend variable, and ε_{it} is the error term. Given above equation, we are especially interested in the coefficient estimate and significance of the interaction between the intervention (I_i) and treatment group (τ_i) as this indicates whether treatment units respond significantly different to the intervention than control units. To test these relationships we use robust standard errors clustered at the user level to account for any serial correlation (Bertrand et al. 2004).

3.5. Results

The first question we address is whether hiding like counts leads to a higher posting frequency (H1). We next investigate the effect of the intervention on the variety of image content (H2). Thereafter, we study whether users whose like counts are hidden get fewer likes than before the intervention (H3).

3.5.1. Posting Frequency

We estimate a fixed and random effects model for the posting frequency. The dependent variable is the number of Instagram posts by a given user in a month, I_t the main effect of the intervention, $\mathrm{I}_t \cdot \tau_i$ the interaction effect of hiding like counts on treatment units. The Hausman test indicates that a fixed effects model is more efficient than a random effects model, $\chi^2(3, 5016) = 163.02$, $p < 0.001$. Therefore, in Table 5 we present the coefficient estimates of various fixed effects models.

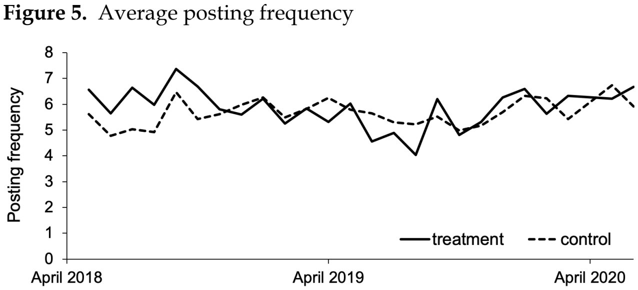
Table 5. Posting frequency higher after hiding like counts

	Fixed effects	Fixed effects + interaction	Fixed effects + double interaction
I_t	4.84***	4.65***	4.65***
	(0.13)	(0.16)	(0.15)
$\mathrm{I}_t \cdot \tau_i$	0.06	0.44*	-0.80***
	(0.14)	(0.22)	(0.22)
γ_t	-0.32***	-0.31***	-0.30***
	(0.008)	(0.01)	(0.01)
$\gamma_t \cdot \tau_i$		-0.04*	-0.15***
		(0.02)	(0.02)
$\gamma_t \cdot \tau_i \cdot \mathrm{I}_t$			0.47***
			(0.03)
R-Squared	0.30	0.30	0.35
F	686.80	516.78	511.38
p-value	0.000	0.000	0.000

Notes. The table shows a fixed and random effects model with the standard robust errors between parenthesis. Estimates are calculated on a matched sample of 238 Instagram users observed one year before and after the intervention (5016 observations). The dependent variable is the log number of posts by user in a month. The independent variables are treatment group dummies (τ_i), intervention step dummies (I_t), a trend variable (γ_t), their interactions, and user-level fixed effects.
* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

The coefficient estimates for the effect of the intervention (I_t) indicate that the posting frequency has increased. This implies that after the intervention Instagram use increased across the spectrum. Furthermore, the trend variable (γ_t) is negative and significant which also follows from Figure 5; especially treatment units posted less and less often in the months prior to the intervention. To control for these time-varying factors of both groups, we augment the fixed effects models with additional interaction terms (middle and right column). The R-Squared is significantly higher in the fixed effects model with double interaction terms, $F = 17.80$, $p < .001$. From this model it follows that the intervention first causes a negative shock ($\mathrm{I}_t \cdot \tau_i$) after which the slope of the trend becomes more positive. Specifically, the overall negative trend among treatment units ($\gamma_t \cdot \tau_i$) turns into a positive trend after the intervention ($\gamma_t \cdot \tau_i \cdot \mathrm{I}_t$). Thus, the post-treatment trend effect is positive which means that the posting frequency of treatment units has gone up which is in line with our hypothesis[5].

Figure 5. Average posting frequency



3.5.2. Variety

As we explained earlier, we conduct a content analysis of the posted images. To that end, we use a computer vision API that returns a vector of image tags and corresponding confidence scores (examples in the Web Appendix D). Tags are objects, living beings, scenery or actions found in the image. As follows from Table 6, the output of the API is comparable for the treatment and control group. Most images are assigned a multitude of tags, and the confidence scores indicate that most image elements are classified with high accuracy.

Table 6. Image Output Summary Statistics

	Mean	Median	SD	Min.	Max.
Treatment					
#Tags	11.24	10	5.66	1	74
Confidence	0.778	0.840	0.217	0.005	1.000
Control					
#Tags	11.16	10	6.26	1	142
Confidence	0.775	0.832	0.216	0.005	1.000

Notes. The table shows summary statistics for the number of tags and confidence scores for images from users in the treatment and control group. Estimates are calculated on a sample of 25,836 images.

We use these vectors of image tags to compute the cosine similarity along two dimensions. First, we study the image contents on a within-subject level. That is, individual's image variety before and after the intervention. Second, we examine whether the change in between-subjects image variety after hiding like counts differs between treatment and control unit cohorts.

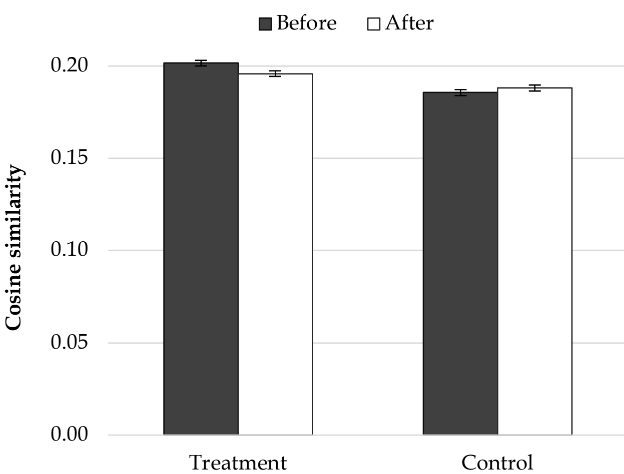
Within-subject

We compute the cosine similarity between the pictures a user shared in the year before and after the intervention. For each user we thus obtain two similarity scores which are dependent on one another. Hence, we conducted a repeated measures ANOVA to compare the effect of hiding like counts on variety. There was neither a main effect of the intervention on within-subject image similarity, $F(1,227) = .42$, $p = .52$, $\eta_p^2 = .0002$, nor an interaction between the treatment group and hiding like counts, $F(1,227) = .32$, $p = .57$, $\eta_p^2 = .0002$. That is to say, hiding like counts did not affect within-subject variety at all. Although our results suggest that variety did not change on an individual level, users may respond to one another. Hence, we conduct a between-subjects similarity analysis to examine whether variety did change between cohorts of users.

Between-subjects

Figure 6 shows the mean between-subjects cosine similarity for treatment and control units before and after hiding like counts. The intervention had a significant effect on image similarity, $F(1, 13808) = 8.90$, $p < .001$, $\eta_p^2 = .00007$. The cosine similarity was higher before (M = .193, SD = .09) than after the intervention (M = .192, SD = .09) which implies that variety increased. Moreover, the interaction between the treatment group and the intervention was significant, $F(1, 13808) = 69.89$, $p < .001$, $\eta_p^2 = .0005$. Paired t-tests were used for the post hoc analysis of the interaction term. Hiding like counts decreased between-subjects similarity among pairs of treatment units with .006, $p < .001$, $d_{avg} = .07$. However, in the control group the cosine similarity increased by .002, $p = .0013$, $d_{avg} = .03$.

Figure 6. Between-subjects variety increases in the treatment group



Notes. The bars present the mean cosine similarity among the cohort of treatment and control units before and after hiding likes. Lower cosine similarity reflects higher image variety. The cosine similarity before and after the intervention did significantly change for both cohorts of users.

In summary, we find that variety measured at an individual level did not change as a result of hiding likes. However, at an aggregated level the treatment group shared pictures that were significantly less alike, which confirms our hypothesis. This result could be attributed to a larger power of the between-subjects test and the dynamics of social media in which users respond to one another. Users may adapt their own posts based on the accounts they follow. As such, adopting a more varied image style may encourage others to reciprocate this behavior. Another explanation could be that users are less likely to jump on Instagram trends as getting enough likes becomes less important to them.

3.5.3. Like Behavior

As we hypothesized earlier, we expect users' tendency to like to decrease after hiding like counts. Table 7 presents the results of a fixed and random effects model for like behavior. The dependent variable is the log mean number of likes by user in a month. The Hausman test indicates that a random effects model is more efficient than a fixed effects model, $\chi^2(3, 4436) = 4.75$, $p = .19$. We find that the number of likes dropped after the intervention (I_t), especially among users whose like counts were hidden ($I_t \cdot \tau_i$). Treatment and control units experienced a fall in the number of likes of 11.2% and 6.7%, respectively. The latter could be attributed to spillover effects and heterogeneity in treatment conditions of users' follower base. On the other hand, we find an upwards trend of the number of likes over time (γ). A rise of popularity of Instagram and user's follower growth could account for this finding. Hence, we also estimate a random effects model in which the number of followers and followings are added as control variables. Indeed, a higher number of followers is associated with more likes, $b = .85$, $t(4436) = 13.63$, $p < .001$. Intuitively, this makes sense as the audience size grows as the follower count goes up. A larger audience in turn increases the potential number of likes.

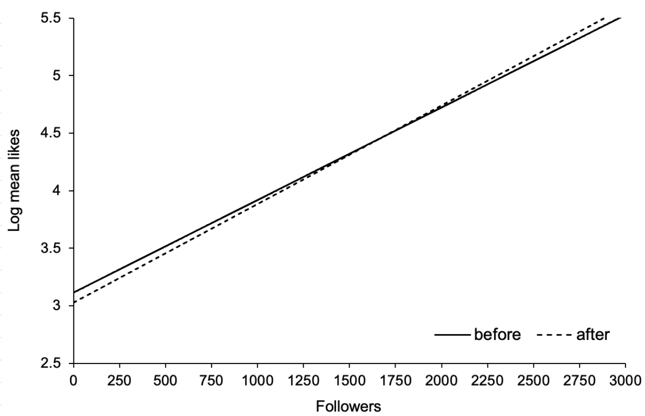
Table 7. The number of likes drops after the intervention

	Fixed effects	Random effects	Random effects + controls
I_t	-67.00**	-66.98**	-66.74**
	(22.06)	(22.06)	(22.06)
$I_t \cdot \tau_i$	-45.44*	-44.87*	-44.97*
	(20.37)	(20.36)	(20.37)
γ_t	6.22***	6.16***	6.15***
	(1.44)	(1.44)	(1.44)
Followers			0.85***
			(0.062)
Followings			-0.081
			(0.059)
R-Squared	0.006	0.02	0.06
F	8.56	75.17	306.15
p-value	0.000	0.000	0.000

Notes. The table shows a fixed (FE) and random effects (RE) model with the standard robust errors between parenthesis. Estimates are calculated on a matched sample of 238 Instagram users observed one year before and after the intervention (4436 observations). The dependent variable is the log mean number of likes a user receives on a post in a given month. The independent variables are treatment group dummies (τ_i), intervention step dummies (I_t), a trend variable (γ_t), an interaction term ($I_t \cdot \tau_i$), and the number of followers and followings. Regression coefficients and standard errors have been multiplied by 1000 to improve readability. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Additionally, we explore whether the number of followers moderates the effect of the intervention on the number of likes. That is, whether users with few and many followers are affected equally by hiding like counts. After all, the effect of hiding a low and high count on users' like behavior might vary. We run another random effects model that incorporates the interaction between the followers count and the intervention, and obtain a significant positive coefficient, $b = .05$, $t(4436) = 3.99$, $p < .001$ (Web Appendix H). Thus, users with many followers are less affected by the intervention than users with only a small audience. Figure 7 illustrates how the effect of the intervention flips at 1663 followers. After this point the dashed line exceeds the solid line and therefore the positive effect of the number of followers outweighs the negative effect of the intervention on like behavior.

Figure 7. Hiding like counts only hurts users with small audiences



Notes. The solid and dashed lines present users before and after hiding like counts, respectively. Estimates are calculated on a sample of 119 treatment units observed one year before and after the intervention (2199 observations). The vertical axis is the log mean number of likes a user receives on a post. The horizontal axis is a user's number of followers. The log mean likes before and after the intervention differ significantly for users with less than 1507 or more than 1819 followers.

These results show that Instagram users follow the crowd with their like behavior. Herd behavior can explain why posts get more engagement once the number of likes is visible. However, this does not hold for users with an above average following. Their posts may be generally accepted and, therefore, liking the post without observing the like count is less likely to cause reputation damage, which we will further test in the experiment. Thus, our results suggest that the positive effect of social validation on others' like behavior may be nonlinear. That is, those without a strong following benefit the most from like counts in terms of the number of likes they get on their posts.

4. Experiment

To overcome the limitations of observational data and determine the effects of hiding like counts on self-esteem, we run an experiment in which participants are randomly shown a photo with a low, high, or hidden like count. Thereafter, we measure how this manipulation affects their posting frequency (H1), like behavior (H3), and self-esteem (H4).

4.1. Method

4.1.1. Participants

Our participants were recruited via Prolific. We required workers to be aged between 18 to 30 years and active Instagram users to ensure they are familiar with the platform dynamics. We received 600 responses (51% female) of which participants' median age was 24 (M = 24.26, SD = 3.64). The sample was 62% White, 12% Black, 9% Hispanic, 15% Asian, and 2% unknown race(s). To ensure randomization was done correctly, we performed a check using ANOVAs and verified there were no significant differences between participants in each like condition in terms of demographics.

4.1.2. Manipulation

After participants confirmed the study requirements, they viewed an Instagram photo of a target person of which the like count was low, high, or hidden. The target persons were chosen to be same-gender strangers so that it was difficult to gauge how realistic their appearance is (Fardouly and Vartanian 2015) and individuals were likely to make social comparisons (Kleemans et al. 2018). To control for composition and framing, we made sure the scenery was identical for both gender conditions, and their body pose and facial expression were highly similar. The low and high like count figures were chosen based on the 20th and 80th percentile of the average like count among a representative sample of Instagram users in our observational study. The number of days since publishing (3 days ago) was included in the picture to control for a scenario where a low like count could be attributed to the recency of the post. Furthermore, we copied the Instagram mobile design, but left out the description and comments in all conditions. Like the real design, the username of one of the people who liked the post was incorporated in the like count. We choose a name that did not reveal the gender and used it across all conditions (Web Appendix I).

Participants were forced to view the photo for at least 10 seconds before they could proceed to the next screen. They were asked to imagine that the person portrayed on the photo was their neighbor. Then, we asked participants to make specific evaluations of themselves and the target person in terms of likeability, popularity, and attractiveness (Vogel et al. 2014). Thereafter, we measured user's intention to like, comment, or share the picture after viewing the post. Finally, the questionnaire concluded by collecting the frequency of Instagram use and demographic information of the participants such as age and ethnicity.

4.1.3. Measures

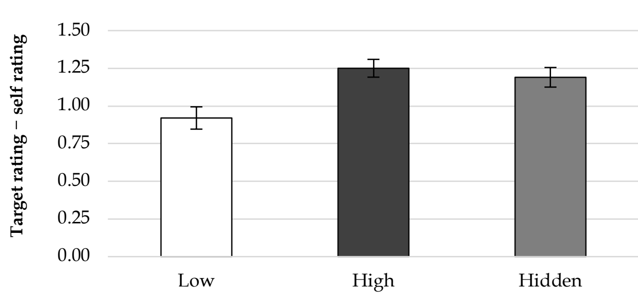
Target and self-rating. Participants indicated on a 7-point scale (1=strongly disagree; 7=strongly agree) to what extent they agreed that the target person and themselves were likeable, popular, and attractive (Vogel et al. 2014).

Instagram use. Participants reported how much time they spend on Instagram daily, with choices including 10 minutes or less, 11–30 minutes, 31–60 minutes, 1–2 hours, 2–3 hours, and more than 3 hours (adapted from Lup et al. 2015).

4.1.4. Results

To examine the impact of the manipulation on self-esteem, ratings on the set of items were collapsed separately for the self (α = .78) and the target (α = .67). First, we subtracted the aggregated ratings for the self-rating from the target-rating to derive the individual level differences (i.e., relative self-esteem). Second, we conducted a one-way ANCOVA[6] to determine a statistically significant difference between the like count conditions on self-esteem controlling for gender, age, ethnicity, and Instagram usage. Our core hypothesis was that the discrepancy between the target person and self would be largest in the high like count condition, while hiding the like count would eliminate popularity cues and thereby attenuate the difference between self and target rating. Indeed, there is a significant effect of the like count condition on relative self-esteem, $F(2, 582) = 6.09$, $p = .037$, $\eta_p^2 = .01$. In the high like count condition the target rating was highest (M = 5.60, SD = 0.85) and the self-rating was lowest (M = 4.35, SD = 1.12). Although in the hidden like count condition the target rating was lower (M = 5.57, SD = 0.88) and the self-rating higher (M = 4.38, SD = 1.14), it did not differ significantly from the high like count condition. On the other hand, in the low like count condition the discrepancy was smallest (Figure 8): the target rating was lowest (M = 5.35, SD = 1.05) while the self-rating was highest (M = 4.43, SD = 1.20). Moreover, post hoc comparisons using the Tukey HSD test indicated that relative self-esteem was significantly higher among participants exposed to a low as opposed to a high like count. Thus, it appears that the number of likes has an impact on how people judge themselves relative to a target person. In particular, relative self-esteem was lower when the target person picture contained a high like count. Hiding like counts marginally attenuates the discrepancy between target and self-ratings, yet relative self-esteem is higher in a low like count condition. In other words, the effect of hiding like counts depends on the number of likes the target person used to get prior to the intervention. Next, we examine how i) like counts and ii) target and self-ratings affect posting frequency and like behavior.

Figure 8. Difference in ratings largest for high like count condition



The effect of like counts on user behavior

After participants had filled out the question related to relative self-esteem, we asked them to indicate their likelihood to engage in various activities on Instagram. In Table 8 the summary statistics for a low, high, and hidden like count are shown. We conduct a MANOVA to test whether the manipulation would affect any of the follow-up user behaviors but did not find any effect, Pillais' Trace = .006, $F(6, 1162) = .56$, $p = .76$. This implies that participants' responses to these questions were not significantly different from one another.

Table 8. Follow-up user behavior as a function of the like count

	Low		High		Hidden	
	M	SD	M	SD	M	SD
Post	2.05	1.70	1.78	1.71	1.87	1.66
Like	3.37	1.94	3.37	1.87	3.36	1.93

Notes. Participants were asked to indicate how likely they were to like the photo, comment on the photo, and share a photo on Instagram themselves on a 7-point scale (1 = *extremely unlikely*; 7 = *extremely likely*).

It may be the case that our experimental manipulation in which participants were asked to imagine the person on the photo was their neighbor, was inappropriate to directly measure user behavior on Instagram. In reality users who cannot observe like counts may draw on other factors beyond those available to our study participants to determine whether a photo is worth liking. For instance, the observational study demonstrated that the effect of the intervention on like behavior differs between users with a low and high follower count. Hence, we extend our empirical investigation to explore whether target and self-ratings can explain differences in posting frequency and like behavior.

The effect of target and self-ratings on user behavior

First, we test whether our measurements for user behavior are independent of the experimental condition. The Chi-Square Test of Independence suggests observations can be considered independent, $\chi^2(2) = 33.63$, $p = 1.00$. Also, the Intraclass Correlation Coefficients (ICC) for liking (.01) and posting (.03) suggests that there is no tendency for values from the same test condition to be similar. As user behaviors are submitted on a seven-point scale, we model posting frequency and like behavior using an ordered logit and probit model (Moe and Schweidel 2012; Ying et al. 2006). Table 9 present the regression estimates for participants' likelihood to post or like which we will discuss next.

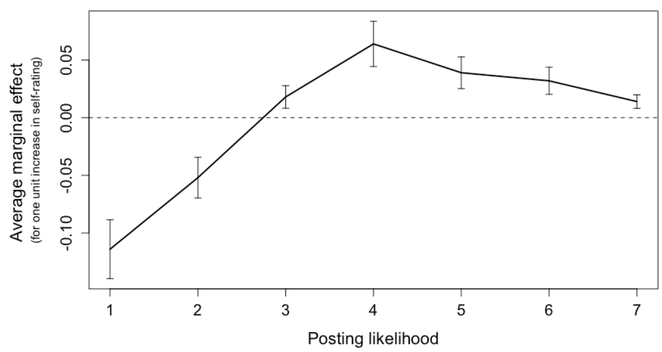
First, the McFadden R^2 and AIC model evaluation criteria favor the logit model for posting likelihood. Figure 9 shows the corresponding average marginal effect (AME) plot. That is, the percentage change in posting likelihood for each level in the scale. For instance, a one unit increase in self-rating is associated with a 3.2% higher probability to *likely* post (posting likelihood = 6).

Table 9. Target and self-ratings affect user behavior

	Post (logit)	Post (probit)	Like (logit)	Like (probit)
Target-rating	-0.052	-0.012	0.77***	0.44***
	(0.087)	(0.049)	(0.090)	(0.050)
Self-rating	0.67***	0.39***	0.53***	0.30***
	(0.073)	(0.042)	(0.069)	(0.04)
McFadden R^2	0.064	0.063	0.10	0.10
Log likelihood	983	984	1005	1006
AIC	2000	2002	2043	2012

Notes. The table shows an ordinal logit and probit model with the standard errors between parentheses. The dependent variables are the likelihood to post on Instagram and participants' tendency to like a picture shown to them in the experiment. The independent and control variables are the target and self-ratings, Instagram use, age, gender, ethnicity, and treatment condition[7]. A higher McFadden R^2 and lower Akaike Information Criterion (AIC) indicate better model-fit, respectively. *** $p < 0.001$.

Figure 9. Posting likelihood higher once self-ratings increase



Notes. The average marginal effect (AME) plot visualizes the percentage change in users' posting tendency after a one unit increase in self-rating for each of the seven posting likelihood levels (1=*extremely unlikely*, 7=*extremely likely*). The error bars present 95% confidence intervals for each likelihood level. Values above and below the dotted line indicate an increased and decreased probability of a given likelihood level, respectively.

The AME also goes up for *somewhat likely* and *extremely likely*, 95% CIs: [.03; .05] and [.008; .02] respectively. At the same time, the probability to not post decreases. The results, therefore, suggest that self-ratings are positively related to posting likelihood. In turn it could be argued that a higher posting likelihood goes hand in hand with a higher posting frequency.

Second, the probit model is preferred for liking behaviors because of a lower AIC value. As could be expected, the coefficient for the target rating is significant and positive, $b = .44$, $t(581) = 8.86$, $p < .001$. High ratings for the target person's attractiveness, likeability, and popularity are associated with a higher tendency to like the picture. In that sense, participants' first impression about the stranger shown to them seems to affect their like behavior. The same holds for individuals' self-rating, $b = .29$, $t(581) = 7.30$, $p < .001$. Higher self-ratings not only increase posting frequency but also liking tendency. This finding complements earlier work from Forest and Wood (2012) in which they demonstrate that people with high self-esteem elicit high positive disclosures on social media. In sum, this experiment adds to a growing corpus of research showing the influence of likes on social media (Vogel et al., 2014). Our data suggests that like counts affect how users evaluate themselves compared to others. As a consequence, the effects on target and self-ratings are related to their like behavior and posting frequency on Instagram, respectively. Next, we discuss the broader implications of these results for consumers, marketeers, platforms, and policymakers and propose directions for future research.

5. Implications

We document the effects of hiding like counts on user behavior and self-esteem: after the intervention users posted more frequently and more varied than in the months before. We also provide evidence that self-image is a key driver for user behavior on social media. The number of likes influences how we evaluate others as well as ourselves. Users are more likely to self-disclose once they rate themselves more positively. On the other hand, our results point to a drop in the number of likes after hiding like counts. These findings impact stakeholders on various levels, which we describe next.

Consumers
First, consumers may be less inclined to chase likes as their followers no longer see their like counts. Instead, they may feel freer to authentically express themselves on social media. This transition could counteract a culture of unhealthy comparison, superficiality, and insecurity among young users (Pei Ting 2019).

Marketeers
Second, influencers may start promoting other forms of engagement, such as comments and direct messages, in an attempt to counterbalance fewer likes (Ribeiro 2019). Their value to advertisers may be measured differently as the number of likes becomes a less important factor in determining whether a campaign is effective or not. At the same, hiding like counts could create a window of opportunity for advertisers to more accurately target individuals based on their authentic preferences.

Platforms
Third, online platforms could benefit from this change for two reasons. An increase in posts will contribute to users spending more time on the platform which implies that more ads can be served. In addition, business users may increase advertising spending in order to compensate for lower engagement on their posts and decreased transparency of the influencer market.

Policy-makers
Fourth, while governments more and more regulate internet services' policies around privacy, copyright, and the use of personal data (Ribeiro 2019), there is still an ongoing debate about whether social media platforms should be publicly regulated. Although there have been widespread calls for social media firms to take more responsibility for their content (Dubicka and Theodosiou 2020), this intervention illustrates that platform design changes can fundamentally shape user behavior as well. Therefore, it raises the question whether policy-makers should expand regulation that forces social media firms to implement health and privacy by design.

Thus, hiding like counts can affect the lives of many people. As more and more people adopt social network services, the responsibility of online platforms will keep growing and growing. On the other hand, researchers and media put social media platforms under heightened scrutiny. As such, creating a healthy online environment seems to be the right step forward to secure long term sustainable growth.

6. Conclusions

Hiding like counts was supposed to create a less pressurized environment in which users feel comfortable expressing themselves. Accordingly, the focus should be on the photos and videos users share as opposed to the number of likes. In the observational study we demonstrate that despite a downwards trend treatment users have posted more frequently on Instagram. Moreover, their pictures have become more varied compared to others after the intervention. This may be considered a promising aspect as it not only responds to calls for more original user-generated content, but it also demonstrates the effectiveness of the intervention. Further, we find that the number of likes on posts has gone down, yet the degree to which depends on users' following size. Target and self-ratings attribute to the underlying mechanism that can explain these findings. People with high relative self-esteem are more likely to share pictures on Instagram.

In our analysis we concentrated on the main effects of the intervention on user behavior and self-esteem. As our data indicate that the effect of hiding like counts depends on users' characteristics, future research should further develop and confirm these initial findings. Examining a broader range of moderating variables will extend and expand our findings. For instance, the number of likes users get in comparison to the accounts they follow may affect their self-ratings. Recent adopters of Instagram may behave differently than users who have seen like counts previously as they lack prior knowledge about how many likes their friends would normally get. More privacy-concerned individuals perhaps react more strongly toward the intervention, while Instagram users who only occasionally spend time on the platform may hardly respond at all.

We built on insights from observational and experimental data to reap the benefits from both perspectives. A limitation of the current study, however, is the experimental design in which we measured like behavior in a hypothetical scenario. In contrast, Sherman et al. (2016) asked participants to submit their own photos to determine the influence of low and high virtual peer endorsement on herd behavior on Instagram. Also, the researchers curated a series of images themselves. Their results demonstrate that the hypothesized effect was significantly stronger for participants' own images. Therefore, future experimental research could be directed to further explore the effect of hiding like counts on participants' like behavior in more realistic settings.

It is a question of future research to investigate the generalizability of our findings. While our results on user behavior pertain to visual-based social media, we expect similar effects to hold on text-based social media. Although our analysis on the effects of hiding like counts was mostly concerned with the frequency and variety of pictures users shared, we postulate that users share more intimate post captions after the intervention. Hiding likes may create a safer environment to disclose intimate feelings without having to fear rejection (Ma et al. 2016). Thus, studying how the text modality changes after hiding likes could be another fruitful area for future research.

Endnotes

[1] The API also assigns one or more categories to each image out of a list of 86 taxonomies. However, tags provide a more valid measure for variety than categories because images with distinct categories (e.g., "People Many" & "People Group") can still be highly equivalent despite a cosine similarity of zero.

[2] We acknowledge that a multitude of exogeneous variables (e.g., childbirth) can cause changes in image characteristics between consecutive years. However, we argue that there is no reason to believe these unexplainable factors occur systematically more often between July 2019 and July 2020 than in the year prior.

[3] The effect of videos on our analysis is neglectable for two reasons. First, 87% of all posts in our sample consists of pictures. Thus, Instagram users will be primarily exposed to these kinds of posts. Second, the intervention also hid impression counts (i.e., views). Both the number of likes and impressions can be considered a social cue that indicate popularity and induce herd behavior. As such, the effect of the intervention on pictures and videos may be comparable.

[4] We cannot disentangle α_i and τ_i, so in reality we estimate $\widehat{\alpha_i}$ which is a function of both terms.

[5] We replicated our analysis with a fixed effects model with an additional interaction term. ($\gamma_t \cdot \tau_i \cdot I_{before}$). We find a significant negative coefficient, $b = -.47$, $t(4773) = 18.50$, $p < .001$, which implies that the pre-treatment trend slope is negative. Therefore, the intervention can be regarded as a breaking point in the posting frequency trend. In fact, in Table 5 an upward shift of the posting frequency intercept after treatment (middle column: $\tau_i \cdot I_t$) masks an upward change in the trend slope after treatment (right column: $\gamma_t \cdot \tau_i \cdot I_t$). Further, the shock at the start of the intervention ($\tau_i \cdot I_t$) suggests that treatment units may initially await and watch how others are responding to the intervention.

[6] We additionally replicated the 2 (source: self or target) X 3 (likes: low, high, hidden) mixed-model ANOVA (Vogel et al., 2014). This approach gives comparable results for the interaction between source and likes, $F(2, 1225) = 2.91$, $p = .055$, $\eta_p^2 = .004$. Notably, our results suggest that seeing a low number of likes on others' posts increases relative self-esteem, whereas Vogel et al. argue that high like counts decrease self-esteem. By incorporating the hidden like count condition, we can rule out the latter. In fact, target and self-ratings are comparable for the high and hidden like count condition. As such, we suggest that the effect of a low and high number of likes should be interpreted compared to a control condition.

[7] To control for any remaining variance that originates from the manipulation (low, high, hidden likes), we add the treatment condition as a covariate. The coefficient estimate is insignificant in all four models.

References

Anderson, Monica, and Jingjing Jiang. 2018. "Teens, Social Media & Technology 2018." *Pew Research Centre*. http://publicservicesalliance.org/wp-content/uploads/2018/06/Teens-Social-Media-Technology-2018-PEW.pdf.

Appel, Helmut, Jan Crusius, and Alexander L. Gerlach. 2015. "Social Comparison, Envy, and Depression on Facebook: A Study Looking at the Effects of High Comparison Standards on Depressed Individuals." *Journal of Social and Clinical Psychology* 34(4): 277–89.

Austin, Peter C. 2011. "An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies." *Multivariate Behavioral Research* 46(3): 399–424.

Babić, Rosario Ana, Francesca Sotgiu, Kristine De Valck, and Tammo H.A. Bijmolt. 2016. "The Effect of Electronic Word of Mouth on Sales: A Meta-Analytic Review of Platform, Product, and Metric Factors." *Journal of Marketing Research* 53(3): 297–318.

Banerjee, A. V. 1992. "A Simple Model of Herd Behavior." *The Quarterly Journal of Economics* 107(3): 797–817.

Battle, James. 1978. "Relationship Between Self-Esteem and Depression." *Psychological Reports*: 745–46.

Battle, James, Lucy Jarratt, Sinha Smit, and D A N Precht. 1988. "Relations Among Self-Esteem, Depressoin, and Anxiety of Children." *Psychological Reports*: 999–1005.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How Much Should We Trust Differences-in-Differences Estimates?" *The Quarterly Journal of Economics*.

Bikhchandandi, Sushil, Ivo Welch, and David Hirshleifer. 1992. "A Theory of Fads , Fashion , Custom , and Cultural Change as Informational Cascades." *Journal of Political Economy* 100(5): 992–1026.

Burrow, Anthony L., and Nicolette Rainone. 2017. "How Many Likes Did I Get?: Purpose Moderates Links between Positive Social Media Feedback and Self-Esteem." *Journal of Experimental Social Psychology* 69: 232–36. http://dx.doi.org/10.1016/j.jesp.2016.09.005.

Chen, R. 2013. "Living a Private Life in Public Social Networks: An Exploration of Member Self-Disclosure." *Decision Support Systems* 55(3): 661–68. http://dx.doi.org/10.1016/j.dss.2012.12.003.

Chen, Yi Fen. 2008. "Herd Behavior in Purchasing Books Online." *Computers in Human Behavior* 24(5): 1977–92.

Chou, Hui Tzu Grace, and Nicholas Edge. 2012. "'They Are Happier and Having Better Lives than I Am': The Impact of Using Facebook on Perceptions of Others' Lives." *Cyberpsychology, Behavior, and Social Networking* 15(2): 117–21.

Conger, Kate. 2019. "Facebook Tests Hiding 'Likes' on Social Media Posts." *The New York Times*. https://www.nytimes.com/2019/09/26/technology/facebook-hidden-likes.html.

Constine, Josh. 2019. "Instagram to Test Hiding Like Counts in US, Which Could Hurt Influencers." *TechCrunch*. https://techcrunch.com/2019/11/08/instagram-hide-likes-us/.

Contrera, Jessica. 2016. "This Is What It's like to Grow up in the Age of Likes, Lols and Longing." *Washington Post*. https://www.washingtonpost.com/sf/style/2016/05/25/13-right-now-this-is-what-its-like-to-grow-up-in-the-age-of-likes-lols-and-longing.

Deschenes, Olivier, and Kyle C. Meng. 2018. 4 Handbook of Environmental Economics *Quasi-Experimental Methods in Environmental Economics: Opportunities and Challenges* ✶. Elsevier B.V.

Deza, Arturo, and Devi Parikh. 2015. "Understanding Image Virality." *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 07-12-June: 1818–26.

Dienlin, Tobias, and Miriam J. Metzger. 2016. "An Extended Privacy Calculus Model for SNSs: Analyzing Self-Disclosure and Self-Withdrawal in a Representative U.S. Sample." *Journal of Computer-Mediated Communication* 21(5): 368–83.

Dubicka, Bernadka, and Louise Theodosiou. 2020. *Technology Use and the Mental Health of Children and Young People*. https://kingsfundmail.org.uk/21A8-6O4NP-UB0FP6-3WVQV0-1/c.aspx.

Fardouly, Jasmine, and Lenny R. Vartanian. 2015. "Negative Comparisons about One's Appearance Mediate the Relationship between Facebook Usage and Body Image Concerns." *Body Image* 12(1): 82–88. http://dx.doi.org/10.1016/j.bodyim.2014.10.004.

Feinstein, Brian A. et al. 2013. "Negative Social Comparison on Facebook and Depressive Symptoms: Rumination as a Mechanism." *Psychology of Popular Media Culture* 2(3): 161–70.

Feld, S. L. 2016. "Why Your Friends Have More Friends Than You Do Author (s): Scott L . Feld Source : American Journal of Sociology , Vol . 96 , No . 6 (May , 1991), Pp . 1464-1477 Published by : The University of Chicago Press Stable URL : Http://Www.Jstor.Org/Stable/2." *Chicago Press* 96(6): 1464–77.

Hanson, Ward A., and Daniel S. Putler. 1996. "Hits and Misses: Herd Behavior and Online Product Popularity." *Marketing Letters* 7(4): 297–305.

Hidrèlèy. 2019. "Somebody Is Showing How Instagram Photos Are All Starting To Look The Same And It's Pretty Freaky." https://www.boredpanda.com/social-media-instagram-identical-photos/.

Hollenbaugh, Erin E., and Amber L. Ferris. 2014. "Facebook Self-Disclosure: Examining the Role of Traits, Social Cohesion, and Motives." *Computers in Human Behavior* 30: 50–58. http://dx.doi.org/10.1016/j.chb.2013.07.055.

Hum, Noelle J. et al. 2011. "A Picture Is Worth a Thousand Words: A Content Analysis of Facebook Profile Photographs." *Computers in Human Behavior* 27(5): 1828–33. http://dx.doi.org/10.1016/j.chb.2011.04.003.

Johnson, Garrett A., Scott K. Shriver, and Shaoyin Du. 2020. "Consumer Privacy Choice in Online Advertising: Who Opts out and at What Cost to Industry?" *Marketing Science* 39(1): 33–51.

Kesvani, Hussein. 2019. "Teens 'Like' Everything on Social Media Now." https://melmagazine.com/en-us/story/teens-like-everything-on-social-media-now.

King, Robert Allen, Pradeep Racherla, and Victoria D. Bush. 2014. "What We Know and Don't Know about Online Word-of-Mouth: A Review and Synthesis of the Literature." *Journal of Interactive Marketing* 28(3): 167–83. http://dx.doi.org/10.1016/j.intmar.2014.02.001.

Kleemans, Mariska, Serena Daalmans, Ilana Carbaat, and Doeschka Anschütz. 2018. "Picture Perfect: The Direct Effect of Manipulated Instagram Photos on Body Image in Adolescent Girls." *Media Psychology* 21(1): 93–110. https://doi.org/10.1080/15213269.2016.1257392.

Knox, George, and Hannes Datta. 2020. *Streaming Services and the Homogenization of Music Consumption*.

Krasnova, Hanna, Natasha F. Veltri, and Oliver Günther. 2012. "Self-Disclosure and Privacy Calculus on Social Networking Sites: The Role of Culture Intercultural Dynamics of Privacy Calculus." *Business and Information Systems Engineering* 4(3): 127–35.

Lee, Young Jin, Kartik Hosanagar, and Yong Tan. 2015. "Do i Follow My Friends or the Crowd? Information Cascades in Online Movie Ratings." *Management Science* 61(9): 2241–58.

Li, Mingxiang. 2013. "Using the Propensity Score Method to Estimate Causal Effects: A Review and Practical Guide." *Organizational Research Methods* 16(2): 188–226.

Loren, Taylor. 2020. "Instagram Is Now Hiding Likes in the US: Here's Everything You Need to Know." https://later.com/blog/hidden-likes-instagram/.

Lup, Katerina, Leora Trub, and Lisa Rosenthal. 2015. "Instagram #Instasad?: Exploring Associations Among Instagram Use, Depressive Symptoms, Negative Social Comparison, and Strangers Followed." *Cyberpsychology, Behavior, and Social Networking* 18(5): 247–52.

Ma, Xiao, Jeff Hancock, and Mor Naaman. 2016. "Anonymity, Intimacy and Self-Disclosure in Social Media." *Conference on Human Factors in Computing Systems - Proceedings*: 3857–69.

Madden, M., Lenhart, A., Cortesi, S., Gasser, U., Duggan, M., Smith, A., &, and M Beaton. 2013. "Teens, Social Media, and Privacy | Pew Research Center." http://www.pewinternet.org/2013/05/21/teens-social-media-and-privacy/.

Mascheroni, Giovanna, Jane Vincent, and Estefanía Jimenez. 2015. "'Girls Are Addicted to Likes so They Post Semi-Naked Selfies': Peer Mediation, Normativity and the Construction of Identity Online." *Cyberpsychology* 9(1).

Moe, Wendy W., and David A. Schweidel. 2012. "Online Product Opinions: Incidence, Evaluation, and Evolution." *Marketing Science* 31(3): 372–86.

Orth, Ulrich, Richard W. Robins, and Brent W. Roberts. 2008. "Low Self-Esteem Prospectively Predicts Depression in Adolescence and Young Adulthood." *Journal of Personality and Social Psychology* 95(3): 695–708.

Ozmen, Dilek et al. 2007. "The Association of Self-Esteem, Depression and Body Satisfaction with Obesity among Turkish Adolescents." *BMC Public Health* 7: 1–7.

Pei Ting, Wong. 2019. "The Big Read: Dangers Lurk in Youth's Chase for Social Media 'Likes.'" *Channel News Asia*. https://www.channelnewsasia.com/news/singapore/instagram-tiktok-social-media-danger-gen-z-youth-the-big-read-11760046.

Ribeiro, Celina. 2019. "How Turning 'Likes' Invisible Is Changing Instagram." *BBC Worklife*. https://www.bbc.com/worklife/article/20190822-how-turning-likes-invisible-is-changing-instagram.

Rietveld, Robert, Willemijn van Dolen, Masoud Mazloom, and Marcel Worring. 2020. "What You Feel, Is What You Like Influence of Message Appeals on Customer Engagement on Instagram." *Journal of Interactive Marketing* 49: 20–53. https://doi.org/10.1016/j.intmar.2019.06.003.

Scissors, Lauren, Moira Burke, and Steven Wengrovitz. 2016. "What's in a Like Attitudes and Behaviors around Receiving Likes on Facebook." *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW* 27: 1501–10.

Sheldon, Pavica, and Katherine Bryant. 2016. "Instagram: Motives for Its Use and Relationship to Narcissism and Contextual Age." *Computers in Human Behavior* 58: 89–97.

Sherman, Lauren E. et al. 2016. "The Power of the Like in Adolescence: Effects of Peer Influence on Neural and Behavioral Responses to Social Media." *Psychological Science* 27(7): 1027–35.

Smith, Aaron. 2014. "What People like and Dislike about Facebook." https://www.pewresearch.org/fact-tank/2014/02/03/what-people-like-dislike-about-facebook/.

Sowislo, Julia Friederike, and Ulrich Orth. 2013. "Does Low Self-Esteem Predict Depression and Anxiety? A Meta-Analysis of Longitudinal Studies." *Psychological Bulletin* 139(1): 213–40.

Suls, Jerry, René Martin, and Ladd Wheeler. 2002. "Social Comparison: Why, with Whom, and with What Effect?" *Current Directions in Psychological Science* 11(5): 159–63.

Tellis, Gerard J., Deborah J. MacInnis, Seshadri Tirunillai, and Yanwei Zhang. 2019. "What Drives Virality (Sharing) of Online Digital Content? The Critical Role of Information, Emotion, and Brand Prominence." *Journal of Marketing* 83(4): 1–20.

Tiggemann, Marika, and Mia Zaccardo. 2018. "'Strong Is the New Skinny': A Content Analysis of #fitspiration Images on Instagram." *Journal of Health Psychology* 23(8): 1003–11.

Vogel, Erin A, Jason P Rose, Lindsay R Roberts, and Katheryn Eckles. 2014. "Social Comparison, Social Media, and Self-Esteem." *Psychology of Popular Media Culture* 3(4): 206–22.

Yang, Chia Chen. 2016. "Instagram Use, Loneliness, and Social Comparison Orientation: Interact and Browse on Social Media, but Don't Compare." *Cyberpsychology, Behavior and Social Networking* 19(12).

Ying, Yuanping, Fred Feinberg, and Michel Wedel. 2006. "Leveraging Missing Ratings to Improve Online Recommendation Systems." *Journal of Marketing Research* 43(3): 355–65.

Zhang, Michael. 2018. "This Instagram Account Shows How Instagram Photos Look the Same." https://petapixel.com/2018/07/26/this-instagram-account-shows-how-instagram-photos-look-the-same/.

Zlatolas, Lili Nemec, Tatjana Welzer, Marjan Heričko, and Marko Hölbl. 2015. "Privacy Antecedents for SNS Self-Disclosure: The Case of Facebook." *Computers in Human Behavior* 45: 158–67.