UNIVERSITY OF CALIFORNIA

Santa Barbara

Adding Insult to Algorithm: How Unfavorable Behavioral Advertising Impacts Self-Concept

A Thesis submitted in partial satisfaction of the requirements

for the Master of Arts in Communication

by

Gwen Petro

Committee in charge:

Professor Amy Gonzales, Chair

Professor Joseph Walther

Professor Miriam Metzger

March 2021

The thesis of Gwen Petro is approved.

_____

Joseph Walther

_____

Miriam Metzger

_____

Amy Gonzales, Committee Chair

December 2020

ACKNOWLEDGEMENTS

First and foremost, I could not have completed this master's thesis without the continual guidance and support of my advisor Amy Gonzales. Thank you for taking the time to read through the many iterations of this and shaping my scattered, amorphous ideas into a cohesive product. I also owe much gratitude to my committee members, Joe Walther and Miriam Metzger, whose insight proved invaluable in transforming this project immensely from start to finish. Finally, I would like to thank my fellow graduate students in the communication department at UCSB, as well as my partner Kyle, who have supported me emotionally throughout this process and pushed me to persevere.

ABSTRACT

Adding Insult to Algorithm: How Unfavorable Behavioral Advertising Impacts Self-Concept

by

Gwen Petro

Online behavioral advertising (OBA) often highlights personal one's shortcomings (e.g., weight loss products, wrinkle creams) yet little research has explored the effects of OBA on negative self-concept or the mechanism for self-concept change following exposure to OBA of any kind. To address these research gaps, and explore possible moderators of this phenomenon, a 2 (*ad personalization:* personalized or non-personalized advertising) x 2 (*advertising favorability:* favorable or unfavorable) x 1 (no advertising control) between-subjects experiment ($N = 120$) was conducted to determine the effects of these factors on perceptions of publicness, self-concept, and privacy concern. Results found some support for the hypothesis that unfavorable OBA can temporarily shift self-concept even when it does not reference a specific instance of self-presentation. However, perceived publicness did not appear to mediate this effect, nor was there evidence for trait self-esteem or causal attributions as moderators. Finally, OBA favorability did not affect online privacy concern. Theoretical implications for future research are discussed, as well as practical implications for digital marketing, platform user experience, and consumer behavior.

Adding Insult to Algorithm: How Unfavorable Behavioral Advertising Impacts Self-Concept

Advertisers have long understood the value of tailoring messages to a target audience, but with advancements in interactive digital technology and artificial intelligence (AI), advertisers are now able to target an audience of one. The process of online behavioral advertising, or OBA, involves "the practice of tracking an individual's online activities in order to deliver advertising tailored to the individual's interests" (Federal Trade Commission, 2009, p. i). Although this may sound straightforward, the inner workings of digital advertising systems are highly complex and are often not fully understood by the people who design them (Ananny & Crawford, 2018), let alone by the average consumer (Yao et al., 2017). Furthermore, when one views a personalized advertisement online and seeks an explanation for why a particular ad was seen, the company's rationale behind that selection often remains ambiguous (Andreou et al., 2018; Eslami et al., 2018). Thus, although OBA is commonplace in the contemporary digital marketplace, it is a phenomenon still shrouded in mystery and one that consumers are often confused by and suspicious of (Eslami et al., 2018; Phelan et al., 2016; Yao et al., 2017).

Given this, it is not surprising that a growing body of scholarship explores how exposure to OBA may impact one's self-view. Summers et al. (2016) found that exposure to OBA that implied a social label altered people's self-perception, whereas exposure to non-personalized advertising or advertising targeted at people's demographic groups did not. Furthermore, French (2018) found that personalized career recommendations based upon the entirety of women's Facebook data influenced how they perceived themselves. These studies build upon earlier research in social psychology on the effects of self-focused attention

(Scheier & Carver, 1980), magnified scanning (Tice, 1992), and public commitment (Kelly & Rodriguez, 2006; Schlenker, 1994) on self-perception, as well as later research on identity shift in the realm of computer-mediated communication (Carr & Hayes, 2018; Gonzales & Hancock, 2008; Walther et al., 2011). All of this work points to the ways that the self-concept can be revised *post-hoc* to align aspects of identity with previous behavior and/or feedback associated with that behavior. Although the studies by Summers et al. (2016) and French (2018) tested the effects of algorithmically-personalized messages based on people's online behavior on self-perception, they only tested messages conferring more positive or neutral social labels. The effect of messages conferring more negative or unfavorable social labels, which to my knowledge has not been, serves as the focus of the present research.

Unfavorable online behavioral advertising, that which has a high potential to make salient one's undesirable traits and cause negative self-conscious emotions, is less understood in computer-mediated communication (CMC) research. When people see traditional broadcast advertisements that reflect less-enviable aspects of their identity (e.g., ads for weight loss programs, wrinkle removers, teeth whitening, treatments for embarrassing health conditions), they may respond negatively towards the advertisements for "airing their dirty laundry," especially when they are viewing the advertising in the presence of others (Putoni et al., 2015; Thomas et al., 2015). Similar reactions may occur when people view personalized advertisements generated by AI that make unfavorable or offensive assumptions about them, even when viewing these advertisements alone on a personal device. Indeed, there is evidence that people are sometimes unpleasantly surprised and caught off-guard by the assumptions AI makes about them based upon their online behavior, whether or not these assumptions have any basis in reality (Eslami et al., 2018; Hautea et al., 2020; Rainie, 2019).

In this study, I propose that one of the reasons unfavorable behavioral advertising may be especially upsetting to some is because OBA may create an awareness of having an audience for one's online behavior. That is, to the degree that people become aware of how someone else might view their public or private online behavior, it is expected that they will temporarily shift their self-concept in a manner that is consistent with the characteristics reflected in the advertising—perhaps even when these characteristics are unfavorable, upsetting, and unsubstantiated. To test this, I elaborate on the growing body of evidence demonstrating that AI-generated messages can alter perceptions of the self (e.g., French, 2018; Summers, 2016) in a few key ways. First, previous research in this domain has not clearly identified a mechanism for these effects on self-perception, but rather has often implied it. Therefore, based on previous CMC studies (Gonzales et al., 2008; Walther et al., 2011), I will measure perceptions of publicness as a mechanism for AI-induced self-concept change. I will also begin to disentangle exactly who or what constitutes the audience that people imagine when they encounter AI-generated messages, which previous research has not clearly answered (e.g. French, 2018; Summers, 2016; Walther et al., 2011). Given the opportunity for exposure to unfavorable OBA, I also consider the impact of this form of negative feedback to determine whether it can shape self-concept. To my knowledge, research on the self-concept effects of AI-generated feedback has exclusively looked at the effects of neutral or positive feedback on the self. I will consider whether people internalize a negative self-concept just as they internalize a neutral or positive self-concept. Moreover, there is a history of research on individual difference moderators of negative feedback delivered via interpersonal communication on self-concept (e.g., Baumeister, 1982; Fedor et al., 2001; Kim & Gonzales, 2018; Park & Crocker, 2008; Swann et al., 1989; Swann, 1992),

which I will utilize to evaluate potential moderators of the effects of negative OBA. For example, whereas most people are inclined to internalize positive feedback, the impact of negative feedback may depend upon individual differences such as trait self-esteem (Swann et al., 1989). Similarly, I will test whether people's attributions about why they are seeing certain ads also moderates the effect on self-concept change. And finally, this thesis attempts to broaden the public commitment literature to determine whether "AI as audience" has implications for people's subsequent privacy preferences. By doing all of this, this study aims to expand both the generalizability and theoretical underpinnings of research that explains the effects of OBA on perceptions of publicness, self-concept, and subsequent attitudes about online privacy.

Online Advertising & Self-Concept Change

The notion that we see ourselves as we believe others see us dates back to the concept of the looking-glass self in symbolic interactionism (Cooley, 1964; Goffman, 1959; Mead, 1934). In this view, the "self" is a fluid and contextually-dependent entity. Identity and self-concept theorists later distinguished between the more stable *core self-concept* and the more malleable *working self-concept* (Markus & Kunda, 1986). According to Markus and Wurf (1987), "the working self-concept, or the self-concept of the moment, is best viewed as a continually active, shifting array of accessible self-knowledge" (p. 306). In other words, people can see themselves in a variety of ways across time and in different situations. However, each situation will activate a particular set of self-concepts temporarily in working memory depending on how one believes they might appear to others in that moment. Thus, studies of change in self-views typically involve the working self-concept, as will be the case in the present study.

It is important to note that self-concept change does not necessarily involve the presence of actual observers—in other words, one need not physically be in public for self-concept change to occur. Decades of scholarship in psychology has demonstrated that the mere possibility of having an audience magnifies the extent of self-concept change by raising self-focused attention (e.g., Kelly & Rodriguez, 2006; Scheier & Carver, 1980; Schlenker et al., 1994; Tice, 1992). In some cases, one needs only to *imagine* how one's behavior might appear to other people or presume that a given self-presentation might have an audience at some point. For example, manipulations of "audience" in this body of literature sometimes involve participants privately observing their own behavior in a mirror (e.g., Scheier & Carver, 1980; Tice, 1992), imagining an audience during a writing exercise (Schlenker et al., 1994), or being video recorded for future audiences (e.g., Kelly & Rodriguez, 2006). It is therefore possible that OBA also activates an imagined audience by reminding internet users that all of their online activity was being recorded and analyzed by outside parties.

Scholars have proposed different mechanisms to explain why perceived publicness enhances self-concept change. As one key example, Tice (1992) explained her findings using *biased scanning*: "a self-perception process in which behavior calls the individual's attention to certain aspects or potentialities of the self, which are then highly accessible and therefore exert a powerful influence on subsequent self-assessment" (Tice, 1992, p. 435). Although biased scanning can occur in private, Tice coins the term *magnified scanning*, arguing that people may feel as though the aspects of the self that they present publicly will become the essence of how others will perceive them (Tice, 1992). In other words, biased scanning carries more weight when it occurs in the presence of an observer.

As an elaboration of the biased scanning approach, Schlenker et al. (1994) found that self-presentation affected people's self-views most intensely when they anticipated being observed by others. They used the notion of *public commitment*, or the idea that presenting oneself in a certain way publicly will commit one to behaving consistently with that self-presentation in the future. Specifically, public commitment argues that, "commitment is a force that ties the individual to some psychological entity. More precisely, it is a pledging or binding of self (a) to an action or set of actions, (b) to a person, group, or organization, or (c) to an idea, often a set of moral principles for conduct. It represents the establishment and recognition of a unit relationship between self and something else" (Schlenker et al., 1994, p.21). In one study, Schlenker and colleagues manipulated public commitment by asking participants to complete a writing task to prepare for an interview that either would or would not actually occur at some point in the future. They found that people altered their self-views to match their self-presentation only when they believed that the interview would occur, in which case their self-presentation would be public. Similarly, in this study, I ask whether people alter their self-views only when they believe that they are seeing online ads advertising personalized for them, in which case their *past* behavior online would have been public to some extent.

Research on public commitment has found that the observation of one's behavior by an audience is a powerful force that can even outweigh private recollection of contradictory past experiences to influence self-views (Kelly & Rodriguez, 2006; Schlenker et al., 1994). According to Kelly and Rodriguez (2006), the strength of public commitment lies in the extent to which one perceives that they will be associated with a certain trait or behavior: "When an actor has an audience who can identify him or her, the actor feels accountable to

the audience for his or her actions" (p. 186). As such, studies that have found evidence of audience effects often include measures of the degree to which participants feel *publicly identifiable* (Kelly & Rodriguez, 2006; Tice, 1992). I propose that because OBA systems operate through the recording and analysis of online behaviors previously enacted on digital devices, OBA may activate some degree of being publicly identifiable. As a result, people may adapt their self-concept to reflect the OBA content. If so, this study will broaden the boundary conditions of previous research looking at perceived publicness as a mechanism for self-concept change, which has traditionally focused on self-concept change in response to a specific and recent self-presentation. Unlike previous studies which manipulate one's self-presentation, I will manipulate the self-concept referenced in OBA, which is more implicitly and ambiguously linked to one's past self-presentation.

In support of this proposition, research has demonstrated that the effects of public self-presentation on self-concept are not limited to offline situations but also occur in online contexts. A study by Gonzales and Hancock (2008) on self-presentation on social media (e.g., creating a blog post) tested whether the interactive component of the *hyperpersonal model* was necessary to catalyze a change in one's self-concept (Walther, 1996). The hyperpersonal model proposes that a feedback loop exists in response to selective self-presentations online wherein message senders may alter their self-views in response to feedback from receivers (Walther, 1996). Gonzales and Hancock (2008) replicated the Tice (1992) study in an online blog setting, demonstrating that, even without interpersonal interaction, self-concept change occurs as a function of online self-presentation. They used the notion of public commitment to explain their findings (Gonzales & Hancock, 2008). That is, when people present themselves in an identifiable manner online, they are "tied to" that

public behavior and thus undergo a temporary shift in self-concept: an *identity shift*. A number of studies have since replicated and expanded upon the notion of identity shift in CMC by, for instance, testing whether feedback moderates this effect (e.g. Carr & Hayes, 2019; Kim & Gonzales, 2018; Walther et al., 2011). I elaborate on this research below.

OBA as Algorithmic Feedback

According to symbolic interactionism, not only is the self socially constructed via beliefs about how others might see us, but also through our interactions with others (Cooley, 1964; Goffman, 1959; Mead, 1934). In that vein, after demonstrating that self-concept change can occur as a function of public online self-presentation (Gonzales & Hancock, 2008), Walther and colleagues further elaborated the findings to demonstrate that feedback provides an additive effect on one's self-concept change (Walther et al., 2011). When people receive feedback confirming their public self-presentation, whether that feedback links people with a certain personality trait or a particular brand image, they are more likely to internalize that self-presentation after actually receiving confirmatory feedback compared to when the feedback is merely imagined (Carr & Hayes, 2019; Walther et al., 2011). This work is based on psychological research conducted offline which finds that people shift their self-presentation to align with the feedback they receive from others and ultimately confirm the other party's beliefs about them (Snyder & Swann, 1978; Snyder, Tanke, & Berscheid, 1977). Furthermore, Walther et al. (2011) and Carr and Hayes (2019) both found some degree of self-concept change occurred regardless of whether or not the explicit feedback participants received was generated by another person or by a computer, suggesting that feedback delivered by an OBA system might have a similar effect. These findings

demonstrate that reinforcing feedback, including computer-generated feedback, further magnifies the effects of audience on self-concept change.

In this study, I explore the effects of feedback in the form of OBA on self-concept when the feedback is more distally linked to one's past behavior than in prior studies. In the studies by Walther et al. (2011) and Carr and Hayes (2019), participants were asked to present themselves as though they exhibited a certain characteristic (e.g., extraversion) and then received explicit feedback about their recent self-presentation. Another study by Summers et al. (2016) demonstrated that OBA can also confer an *implied social label*, or "a characterization of the self that is implied to be held by an external agent" (p. 158) and thereby alter self-concept. Participants in the Summers et al. study engaged in an online shopping task which, like the studies mentioned above, generated a specific instance of self-presentation which would then be directly referenced in the feedback they received. In other words, in these previous studies, computer-generated feedback has been linked to a specific and recent act of self-presentation. In this study, I ask: when OBA is not so clearly linked to an act of self-presentation, and people have to determine for themselves how or why a particular message was personalized for them, will it still affect people's self-perception?

In reality, OBA does not always clearly reference self-presentations that are recent or specific, and it is rarely accompanied by a precise description of how it was personalized as it was in the study by Summers and colleagues. After all, OBA operates via sophisticated algorithms that might generate unexpected insights about people based upon their seemingly random data points. But there is reason to believe that in more ambiguous scenarios, people might still view OBA as a form of feedback about their online self-presentation. For example, French (2018) found that algorithmically-generated feedback in the form of a

career recommendation based upon women's Facebook data history—not an individualized act of online self-presentation—influenced participants' self-concept. Despite the fact that French's study did not look at advertising messages, it did show that feedback about the *entirety* of one's online self-presentation on a platform such as Facebook, as opposed to feedback about a *specific* instance of someone's behavior online, can induce self-concept change. French (2018) asserts that, "algorithmic personalization is a growing form of feedback about the self" (p. 9). Based on this framing, and in an attempt to optimize the external validity of research on public commitment and identity shift in a contemporary context, I also explore feedback that is linked to a more holistic history of online self-presentation, rather than in individual act, and the effects of that feedback on the self.

A couple of recent qualitative studies also seem to support the idea that inferences about people made by ad-personalization algorithms can serve as a form of feedback, even when they are not clearly linked to people's behavior. One such study by Hautea et al. (2020) found that when people viewed the ways in which Google or Facebook categorized them for advertising purposes and discovered surprising, seemingly irrelevant inferences made about them by algorithms, they "often believed that inferences were a reflection of actions they took online (searching, browsing, clicking) even when they could not recall taking any specific actions that would have led to the inference being associated with them" (p. 4). A study by Eslami et al. (2018) noted a similar pattern, stating that, "When participants started viewing these incorrect interests, rather than stating that the algorithm was wrong, many…tried to find reasons to justify those mistakes" (p. 8). Based on these studies, it appears that people will sometimes attempt to explain why an ad personalization system made certain inferences about them, thereby linking these algorithmic outputs to their past

behavior and allowing them to operate as a form of feedback. I want to test whether this process of accepting and internalizing OBA, even negative OBA, will subsequently shift the self-concept to be more aligned with that unfavorable feedback.

OBA: Hitting Where It Hurts

Self-concept favorability is essential to consider when studying the impact of personalized messages on self-concept, and so far, it has been absent from the research on digital communication and self-concept change. For instance, while Summers et al. (2016) found that OBA could confer social labels and alter self-concept, these effects were only tested using socially desirable personality traits (e.g. sophisticated, environmentally conscious). Yet people regularly search for, and thus make "public" stigmatized personal information (e.g., health conditions, relationship troubles, financial concerns) on Google rather than, or before sharing such information, with other people (Joinson & Banyard, 2002). As a result, OBA may not always reflect the positively-enhanced versions of ourselves that we are motivated to present in many interpersonal online contexts (e.g. Toma et al., 2008; Walther, 1996). Instead, the messages in certain OBA might point out one's negative traits or insecurities, leaving viewers feeling upset and offended. It is therefore ecologically valid to examine the effect of less flattering or unfavorable OBA on self-views which, to my knowledge, has not been tested in previous research.

It is unclear whether unfavorable OBA will have the same effect on self-concept as the more favorable OBA tested by Summers et al. (2016). Scholars and philosophers have long recognized that humans are driven by a desire to maximize pleasure and minimize pain. In support of this idea, research has found that people are generally inclined to interpret self-relevant information in ways that limit its negative and maximize its positive implications for

the self (e.g., Baumeister, 1982; Kunda, 1987; Sedikides, 1993; Shrauger & Lund, 1975). This process, known as *self-enhancement*, refers to the mechanism by which people evaluate information that pertains to the self (Alicke & Sedikides, 2009). As Alicke and Sedikides (2009) put it, people utilize "construal mechanisms such as reinterpreting the meaning of social or task feedback, misremembering or reconstructing events in a self-serving way, and making excuses for poor behaviour [sic] or performance" (p. 6). As such, people are generally more willing to accept praise, even when the feedback is randomly generated, and more reluctant to accept criticism (Baumeister, 1982). For instance, qualitative research by Eslami et al. (2018) found that, "when an algorithmically-inferred interest described an attribute a participant did not have but would be proud to possess (the ideal self), the participant was still satisfied" (pp. 432). Meanwhile, research shows that when people encounter negative feedback, they often distort or deny the experience (Crary, 1966) or are more critical of the information and its source (Ditto & Boardman, 1995; Eslami et al., 2018; Kunda, 1987). Therefore, people might be more inclined to internalize favorable traits referenced in personalized advertising than unfavorable ones, which may help explain why most research on the effects of audience on self-concept have considered neutral or positive self-concepts.

To summarize my argument thus far, there is reason to believe that personalized computer-generated messages such as OBA might serve as a form of feedback, even when they are not clearly linked to a specific self-presentation. As Walther et al. (2011) found, computer-generated feedback strengthens the effect of computer-mediated self-presentation on self-concept. This can explain the finding by Summers et al. (2016) that personalized advertisements can impact self-concept, whereas non-personalized advertisements, and

advertisements targeting one's demographic group rather than an individual, do not have the same effect—OBA constitutes feedback about one's self-presentation and thereby shifts one's self-concept, whereas other forms of online advertising do not. Furthermore, French (2018) found that algorithmically-personalized messages can affect self-concept even when they do not reference a specific, recent act of self-presentation. In this study, I compare the effect of non-personalized advertising to personalized advertising (OBA) on self-concept, but the stimuli used in this study will not reference a particular instance of self-presentation, in an attempt to enlarge the boundary conditions of these phenomena. Additionally, I build on previous studies by also manipulating whether an advertisement contains a favorable or unfavorable message in order to determine if there is an effect of ad personalization on self-concept in both cases. As a result, I pose the following hypothesis:

> **H1:** Personalized advertisements, both negative and positive, will lead to greater self-concept change than non-personalized advertisements.

Algorithm as Audience

The question still remains as to *why* it is the case that OBA or other forms of algorithmic feedback can shift one's self-concept, and whether this is due to heightened perceptions of publicness. Summers et al. suggest that, "It is only when an implied social label introduces information about how an external agent views one's level of [some trait] that self-perceptions change" (2016, p. 164). Indeed, Walther et al. (2011) measured perceptions of publicness in the case of algorithmic feedback to determine whether this accounts for its effects on one's identity. Although Walther et al. did *not* find evidence that people experienced greater perceptions of publicness when they received feedback from an algorithm compared to when they received no feedback, the authors suggest that their

measure of perceptions of publicness might not have been sensitive enough to capture perceptions of the algorithm as audience. In that study, participants were asked about the extent to which they felt publicly identifiable to other people and to the researchers in particular. In the present study, I amend that measure by asking whether participants who receive personalized versus non-personalized advertisements feel more identifiable not only to the researchers or other people, but *to the system itself*. In this way, I measure perceived publicness not in the conventional sense involving human audiences, but in a way that is more relevant to the algorithms that constantly monitor and record our online behavior. Thus, I pose this second hypothesis:

> **H2:** The extent to which one feels publicly identifiable will mediate the effect of personalized advertising on self-concept.

Finally, another contribution of this study is that I seek to clarify the ambiguity about who exactly constitutes the audience that people might perceive when they view OBA. Is it the algorithm, as French (2018) suggests? Is it marketers or advertisers, as Summers et al. (2016) propose? Or, is the people who might have access to one's data, such as the employees at Instagram or the researchers conducting the experiment, as Walther et al. (2011) considered? I pose the following research questions to gain additional insights into the public commitment effects associated with OBA:

> **RQ1:** Which actors are people most likely to perceive as an audience when they are exposed to OBA?

It is important to understand the audience that people imagine when exposed to OBA given previous work showing that the source of feedback matters. For instance, organizational research has shown that employees' performance improves more following feedback from a

more powerful superior (Fedor et al., 2001). Research in computer-mediated communication has begun to explore this topic as well: French and Hancock (2016) found that when people perceive that an online audience values a particular trait that they have exhibited, they are more likely to internalize that trait. Further, Carr and Foreman (2019) found that those who received feedback about a recent online self-presentation exhibited greater self-concept change when the feedback came from someone close to them compared to similar feedback from a stranger. Therefore, identifying the types of audiences that people imagine when they receive OBA may help explain variation in the strength of self-concept change following exposure to OBA which confers a particular trait.

Trait Self-Esteem as Moderator

Much of the work on the effects of negative feedback on self-views has focused on self-esteem as a moderating factor (Baumeister, 1982; Shrauger & Lund, 1975; Fedor et al., 2001; Kim & Gonzales, 2018; Park & Crocker, 2008; Swann et al., 1989; Swann, 1992). Trait self-esteem might therefore moderate the effect of offensive OBA on subsequent self-views. Leary and Baumeister (2000) define global trait self-esteem as "a person's appraisal of his or her value…related more strongly to perceptions of others' evaluations of oneself than to seemingly objective indicators of one's ability or goodness" (pp. 2). Although people are generally motivated to see themselves in a positive light due to the tendency towards self-enhancement, *self-verification theory* predicts that people are also inclined to maintain consistent self-views (Swann, 1983). For instance, individuals with low self-esteem are more likely to persist in relationships and in jobs with people that also hold them in low regard than with people that hold with higher regard (Swann et al., 1992; Wiesenfeld et al., 2007). Another study found that individuals with lower self-worth were less likely to discount

negative feedback received via email when it was public than when it was private, whereas individuals with high self-worth discounted negative feedback regardless of publicness (Kim & Gonzales, 2018). In other words, people may internalize negative feedback to the extent that it is consistent with their pre-existing self-views, especially when there is the perception of an audience. Given this research, I will test the following hypothesis to examine the moderating effects of trait self-esteem on self-concept change in response to negative or unfavorable OBA feedback:

H3: Self-esteem will moderate the effect of advertising message personalization on self-concept change.

Attributions as Moderator

People's causal attributions about algorithmically-generated feedback have also been shown to moderate the effect of such feedback on self-concept. According to French (2018), the strength of self-concept change as a result of exposure to a personalized recommendation is moderated by the locus of causality of users' attributions. That is, self-concept change is greatest when one attributes the cause of a personalized recommendation to something internal within oneself (e.g., one's actual identity, interests, and values) rather than to something external within one's environment (e.g., inferences made by the system or by advertisers, the behavior of one's social network or other users, chance). French (2018) argues that people who form more internal attributions to explain why they received a certain personalized recommendation feel more self-aware and conscious of how others might perceive their behavior than those who form more external attributions. As a result, perhaps due to heightened perceptions of publicness, people who make internal attributions for personalized recommendations are more likely to experience self-concept change. Similarly,

as Summers et al. (2016) put it, “implied social labels appear to be able to lower [or raise] participants’ self-perceptions on a trait as long as the behavioral targeting is at least moderately accurate” (p. 169).

Research in other domains further supports the idea that causal attributions impact self-concept, in both negative and positive directions. For instance, one study done in offline contexts found that internal attributions for academic success predict elementary school children’s intellectual self-concept such that high-achieving people who made more internal attributions for success were more likely to see themselves as intellectual (Kanoy et al., 1980). Similarly, Cadinu et al. (2006) found that internal attributions also moderate the effect of stereotype threat wherein females who were primed to think about a negative stereotype about women’s mathematical skills performed worse on a subsequent exam, particularly when they had a more internal locus of causality. Thus, people’s attributions for favorable or unfavorable events appear to predict whether their self-concept changes and whether that self-concept is also favorable or unfavorable.

Similarly, these same types of findings might occur following exposure to personalized advertisements, such that those who make more internal attributions about an ad will be more likely to experience self-concept change. For instance, someone who sees herself as overweight might experience a shift in self-concept when confronted with a personalized ad for a weight-loss remedy. That is, the person in this case would likely make a more internal attribution for why she was being presented with that advertisement (e.g., *I am seeing this ad because I often browse content related to weight-loss*), whereas someone who does not see themselves in this way might attribute the ad to external factors (e.g., *I am seeing this ad because some people on this website match the advertiser’s target*

demographic). In turn, a more internal attribution would contribute to greater self-concept change as a result of exposure to OBA designating a certain label. Accordingly, I thus pose the following hypothesis:

> **H4:** The more *internal* one's attributions about why they received a particular ad, the more likely one is to experience self-concept change.

Perceived Privacy Risk

A final theoretical contribution of this study involves the integration of public commitment and identity shift research with outcomes related to informational privacy. Informational privacy describes the right to control one's personal information and decide how it is used (Westin, 1967). Scholars have argued that OBA violates people's right to informational privacy because people lack sufficient control over how their personal information is collected and utilized (Turow, Hennessy, & Draper, 2015). Indeed, self-efficacy is generally low in this domain, meaning that people do not feel as though they are able to protect their privacy online (Boerman, Kruikemeier, & Borgesius, 2018). On Instagram, for example, users can "hide" a given ad from their feed or report it for various reasons, but they cannot easily prevent future targeted advertising based upon a certain inferred characteristic.

I argue that individuals who are exposed to unfavorable OBA should be more likely to desire greater privacy control. Negative emotions such as shame and embarrassment are associated with a greater desire to hide or disappear (Lewis, 1995), and public online behavior has been known to cause shame or embarrassment. For example, Oeldorf-Hirsh et al. (2017) found that people were more embarrassed by unflattering Facebook posts when they perceived a large, unknown audience. Furthermore, Agarwal et al. (2013) found that

people were concerned about receiving OBA with suggestive or embarrassing content and wanted a way to selectively filter that advertising content. However, to my knowledge no one has experimentally tested whether personalized advertising that confers a negative trait specifically activates a desire for greater privacy than personalized advertising conferring a positive trait. In particular, I propose that when publicness is triggered by algorithmically-derived OBA, and the advertising confers a negative trait rather than a positive one, people will be more concerned about their online privacy and will be motivated to engage in behavior to enhance online privacy if possible. I therefore propose a final hypothesis:

> **H5:** Those who are exposed to unfavorable OBA will be more concerned about their online privacy and more likely to desire greater privacy control than those who see more favorable OBA.

Methods

A 2 (*ad personalization:* personalized or non-personalized advertising) x 2 (*advertising favorability:* favorable or unfavorable) between-subjects experimental design was used to examine the effect of advertisement personalization and favorability on perceptions of publicness, self-concept, and privacy concern. The design of the stimuli used in this study was primarily based on Summers et al. (2016), which measured effects of OBA on self-concept (e.g. seeing oneself as sophisticated or outdoorsy), although that study neither manipulated ad favorability nor assessed public commitment as a mechanism of self-concept change. Also unlike in the Summers et al. study, participants in the present study provided the researchers with their personal data history from Instagram (a procedure adapted from French, 2018). This data presumably contains a multitude of different

presentations of the self and a range of self-concepts which targeted advertising might make salient.

Instagram users, who presently represent a considerable portion of the undergraduate student population, are often exposed to personalized advertising on this platform—as one scrolls though the platform, they can expect to see a personalized ad between every few posts that relates to their interests or needs. Several advertisements were pilot tested to establish their believability, their personal relevance, and how positively or negatively they would be perceived by viewers. Based on these findings, I selected a single self-concept manipulation using an advertisement for a fitness service. As a point of comparison, I also measured this self-concept among a separate group of participants who supplied their Instagram data but were not shown any advertisements.

Pilot Testing

Prior to conducting the experiment, I performed a pilot test to ensure that the cover story would be believable, the experimental stimuli were convincing, and that I would be able to manipulate advertising favorability as intended. To ensure that the cover story would be believable, I first asked a sample of 104 participants to rate on a seven-point scale (1 = *extremely unlikely*, 7 = *extremely likely*) the likelihood that an algorithm could infer various characteristics about them based on their Instagram behavior. These characteristics included their shopping habits ($M = 5.54$, $SD = 1.65$), the size of their social circle ($M = 4.94$, $SD = 1.67$), their eating habits ($M = 4.57$, $SD = 1.72$), their personality ($M = 4.54$, $SD = 1.53$), their personal grooming habits ($M = 4.54$, $SD = 1.67$), their level of physical fitness[1] ($M = 4.52$,

[1] The relatively low rating for fitness here might partially explain the lack of an effect in the study, despite an ultimately successful pilot test.

SD = 1.66), their level of mental health (*M* = 3.93, *SD* = 1.76), and their intelligence level (*M* = 3.66, *SD* = 1.65). The majority of these ratings were above the scale midpoint, suggesting that participants would find it plausible that advertisements in these domains might be personalized for them.

Next, specific advertisements corresponding to some of the categories listed above were pilot tested to corroborate that they would be perceived as personalized by participants. Pilot test participants each saw several advertisements for different businesses presented in a random order. Among these were some advertisements for clearly labeled corporate products and services including Chipotle (restaurant chain), Everlane (denim), Nike (sneakers), and Postmates (food delivery app), which were identical for all participants. In addition, participants were randomly shown one of two versions of advertisements for products and services that could be seen as favorable or unfavorable. Testing different versions of advertisement manipulations helped ensure that I could successfully manipulate advertising message favorability and that participants would believe that the advertisements appeared authentic and personalized for them. Among these test stimuli that I created were advertisements for ClassPass (an app to sign up for fitness classes), Curology (a skin-care product), a brand of protein powder, an online peer tutoring/mentorship service, and a public-service advertisement encouraging behavior to stop the spread of COVID-19. Each of these advertisements were accompanied by one of two captions intended to activate either a negative or positive self-concept. For example, the ad for the fitness app ClassPass featured one of the following captions: “Don’t let the quarantine get in the way of your workout! Join today to access thousands of virtual fitness classes” (favorable) or, “Being at home all day is no excuse to be lazy! Put the chips down and get off the couch with a virtual workout class”

(unfavorable). The former ad presumably implies that the viewer is someone who works out regularly, whereas the latter implies that the viewer is lazy and out of shape.

Pilot test participants were asked to review each ad and read its caption, then think about the types of advertisements they usually see on Instagram and answer questions about how these ads compared. First, they rated how favorably they perceived each advertisement using a seven-point semantic differential scale (1 = *very unfavorable*, 7 = *very favorable*) from Holbrook and Batra (1987). The four scale items include: I dislike the ad—I like the ad; My reaction to the ad is favorable—My reaction to the ad is unfavorable (reversed); I feel positively toward the ad—I feel negatively toward the ad (reversed); The ad is bad—The ad is good ($\alpha = 0.86$, $M = 4.99$, $SD = 1.19$). For the ads with captions that were manipulated, participants also rated the extent to which they felt three negative self-conscious emotions—guilt, shame, and embarrassment—on a seven-point scale from 0 (not at all) to 6 (very much) ($\alpha = 0.87$, $M = 1.88$, $SD = 1.17$). Next, they rated how relevant they found each advertisement on a four-item scale adapted from Zhu & Chang (2016): "I think this ad fits my interests," "I think this ad fits my preferences," "I think this ad fits my tastes," and "Overall, I think this ad fits me," ($\alpha = 0.98$, $M = 4.94$, $SD = 1.58$). Participants rated each statement on a seven-point scale (1 = *strongly disagree*, 7 = *strongly agree*). Finally, participants rated the degree to which they feel they would expect to see each advertisement in their Instagram feed. Participants indicated how much they agreed with the following statements on a seven-point scale (1 = *strongly disagree*, 7 = *strongly agree*): "This looks like an ad I would see in my own Instagram feed," "I would never see an ad like this on my Instagram feed (reversed)," and, "I would expect to see an ad like this when browsing Instagram," ($\alpha = 0.89$, $M = 5.39$, $SD = 1.27$).

Among the advertisements I tested to serve as the experimental manipulation, I opted for one that pilot test participants rated high in relevance and believability. However, I was also careful to select an ad which received different favorability ratings depending on its caption. After conducting a series of one-way MANOVAs, the ad that best met these criteria was the one for ClassPass. Overall, pilot test participants found it somewhat relevant (M = 4.46, SD = 1.68) and fairly believable (M = 5.18, SD = 1.52). Between those who saw the favorable or unfavorable caption, there were no significant differences in perceptions of ad relevance, $F(60) = .34$, $p = .56$, or believability, $F(60) = .48$, $p = .49$. However, those who saw the favorable caption rated the ad significantly higher in favorability (M = 4.85, SD = 0.86) than those who saw the unfavorable caption (M = 4.15, SD = 1.33), $F(60) = 5.99$, $p < .05$. Additionally, those who saw the favorable caption experienced lower levels of negative self-conscious emotions (M = 2.32, SD = 1.54) compared to those who saw the unfavorable caption (M = 3.27, SD =1.98), $F(60) = 4.41$, $p < .05$. This is the only ad that yielded statistically different favorability ratings in each condition, but did not yield overall differences in relevance and believability.[2]

The secondary purpose of the pilot test was to identify two appropriate filler ads. For the sake of variety, I decided to select one of the two food-related and one of the two fashion-related ads that I tested. I conducted a one-way MANOVA to compare each of the filler ads with one another and with both versions of the fitness ad to ensure that there would be no differences in terms of perceived favorability, relevance, and believability between any of the ads. I chose the Postmates advertisement over the one for Chipotle because the Chipotle ad

[2] See Table 1 in the appendix for how these ratings compare to those of the other advertisements tested to use as the manipulation.

was rated as significantly higher in relevance than the unfavorable version of the ClassPass ad, $t = -3.53$, $p < .05$, whereas there was no difference in relevance between the Postmates and ClassPass ads. Between the Nike and Everlane advertisements, I selected the one for Everlane because it was rated marginally statistically higher in believability ($t = 1.87$, $p = .07$), although it was not significantly higher in believability than the ClassPass ad. In sum, there were no significant differences between the four selected ads (Postmates, Everlane, ClassPass favorable, and ClassPass unfavorable) in terms of relevance ($F = 2.24$, $p = .09$) or believability ($F = 2.11$, $p = .10$). The final set of advertisements used in the experiment can be found in Appendix A.

Sample

298 undergraduate students from UC Santa Barbara participated in the first part of this study in exchange for course credit. Participants were pre-screened to ensure that they used Instagram regularly and were then randomly assigned to one experimental condition. Of these participants, 145 supplied us with their Instagram data. Several of the individuals who participated in the first part of the study but not the second communicated to us that they had second thoughts about sharing their data with us, but it is unknown exactly why many of them chose not to participate further.[3] 127 participants completed the second part of the study in exchange for additional course credit. A few responses were duplicated or were flagged as multivariate outliers (as evidenced by an outlying Mahalanobis distance) and were removed, leaving a total of 120 complete responses.

[3] I will consider the effect that this low retention rate may have had on my results in the discussion section.

An a-priori power analysis to determine the appropriate sample size had revealed that in order to conduct a two-way ANOVA with two levels to each variable at 80% power and achieve an effect size of 0.3 (the medium effect size in communication research according to Weber and Popova, 2012), $p < 0.05$, I would need 23 participants per condition. Therefore, the sample size of N=120 was appropriate. The final sample skewed female (70.8%, vs. 28.3% male and 0.8% gender non-conforming). Participants ranged in age from 18-25 (M = 19.37, SD = 2.19). The sample was fairly diverse: 50% self-identified as Asian, 37.5% White/Caucasian, 17.5% Hispanic/Latino, 5.8% Black/African American, and 4.2% some other race or ethnicity (14.2% identified as more than one race or ethnicity).

Experimental Procedure

First, participants completed a pre-test questionnaire. At the start of the questionnaire, they were told that the researchers were user-testing a new advertising system integrated with Instagram. Participants were informed that the study would involve granting the researchers access to their data from Instagram. They were assured that their data would only be used to train the new system to deliver ads personalized specifically for them, and that the researchers would not view participants' data or store it after the process was complete. This procedure was similar to that used by French (2018), only I asked for participants' Instagram data rather than their Facebook data. After providing informed consent, participants proceeded to a series of moderating or control survey items to measure their time spent using Instagram, trait self-esteem, and demographics. At the end of the questionnaire, participants were given instructions for downloading their Instagram data (see Appendix B) and sending it to the experimenter. They were asked to refrain from looking through their downloaded

Instagram data before the experiment was complete. They shared their data with the researchers via email who, unbeknownst to them, immediately deleted it.

The day after their data was received, participants were emailed a personalized link to complete the second part of the study. They began by reading the following introduction:

> *We will now show you a series of ads delivered from a new advertising system still in an early stage of development. The advertising system uses an ad-matching algorithm designed by a team of communication scholars and computer scientists at UCSB. We designed this algorithm in order to study how people feel about personalized advertising like that which they see on Instagram. Much like Instagram's ad-matching algorithm, our algorithm analyzes data about your behavior on Instagram and selects ads that are the best match for you. Please note that the system is only programmed with a limited subset of several hundred ads from which to choose.*

Next, participants were randomly assigned to one of the four experimental conditions in which they saw three similar advertisements (see Appendix A), or they were assigned to a control condition in which they did not see any advertisements. The participants in the personalized advertising conditions were told: "*Based upon your Instagram data, the ad system has matched you with the following three advertisements."* The participants in the non-personalized advertising conditions were shown the same ads, but these participants were told: "*Unfortunately, our system encountered an error and did not properly read your Instagram data. Please bear with us as the system is still in the user-testing phase, and we are currently working to prevent this issue from reoccurring. However, we would still like your feedback on some ads which the system will generate randomly*." The final group or participants assigned to the control condition were told, "*Unfortunately, our system*

encountered an error and did not properly read your Instagram data. Please bear with us as the system is still in the user-testing phase, and we are currently working to prevent this issue from reoccurring. However, we would still like you to answer some questions about yourself so that we can better understand the user base."

Participants in one of the four advertising conditions then viewed the advertisements, which were each presented in a mock-up of the Instagram layout. The two filler advertisements for Postmates and Everlane appeared first with the same captions in each experimental condition. These were followed by the advertisement for ClassPass, which was accompanied by one of two different captions, intended to manipulate either favorable or unfavorable feedback. Participants were instructed to carefully examine each ad as well as its caption and were required to view each ad individually for at least ten seconds before proceeding. Afterwards, participants were told to focus on the final advertisement they saw and answer a series of questions while it reappeared on the screen. The questionnaire first included a manipulation check to ensure that participants who were told that the advertisements were personalized for them perceived them as more personalized than those who were told that the advertisements were not personalized. Next, I assessed participants' perceptions of publicness and their emotional responses to the advertisement. Participants then completed a scale measuring self-concept and another scale measuring their attributions about why they saw that particular ad. Additionally, participants answered questions measuring privacy concern and desire for greater informational privacy. Finally, the researcher debriefed participants, reassuring them that their Instagram data was immediately deleted and never used, and thanked them for their time.

Measures

Manipulation Checks

Perceived Personalization of Ads. Participants rated the extent to which they agree with the following statements about the manipulated ClassPass advertisement: "The advertisement was related to my recent behavior on Instagram," "The ad seemed to be designed specifically for me," and "The ad targeted me as a unique individual." The first statement was created for the purposes of this study, and the second two statements were taken from Li (2016). Participants rated each statement on a seven-point scale (1 = *strongly disagree*, 7 = *strongly agree*), and an average was taken to form a composite score for each participant ($M = 3.86$, $SD = 1.36$, $\alpha = .81$).

Advertising Relevance. Participants rated how relevant they found the ClassPass advertisement on a four-item scale adapted from Zhu & Chang (2016): "I think this ad fits my interests," "I think this ad fits my preferences," "I think this ad fits my tastes," and "Overall, I think this ad fits me." Participants rated each statement on a seven-point scale (1 = *strongly disagree*, 7 = *strongly agree*), and an average was taken to form a composite score for each participant ($M = 4.38$, $SD = 1.45$, $\alpha = .91$).

Advertising Favorability. Perceptions of advertising favorability for the ClassPass were measured using a seven-point semantic-differential scale from Holbrook and Batra (1987). The four scale items include: I dislike the ad—I like the ad; My reaction to the ad is favorable—My reaction to the ad is unfavorable (reversed); I feel positively toward the ad—I feel negatively toward the ad (reversed); The ad is bad—The ad is good. The average of these statements was taken to form a composite score for each participant ($M = 4.52$, $SD = 1.11$, $\alpha = .85$).

Negative Self-Conscious Emotions. Participants rated the extent to which they experienced guilt, shame, and embarrassment in response to viewing the ClassPass advertisement. Participants indicated the extent to which they felt each emotion on a seven-point scale from 0 (not at all) to 6 (very much). A composite measure of negative self-conscious emotions was formed by taking the average of the amount of guilt, shame, and embarrassment ($M = 2.37$, $SD = 1.41$, $\alpha = .90$).

Mediating Variables

Perceived publicness. Perceived publicness was measured with an adapted version of the statement used by Tice (1992) as well as Gonzales and Hancock (2008). The adapted statement read, “Please indicate the extent to which you feel that the following groups would associate you with the last advertisement you saw." The groups listed were: the ad personalization system, Instagram employees, advertisers, the researchers, and your Instagram followers. Participants responded to these statements on a 7-point scale with endpoints labeled not at all (1) and very much (7). The average of these statements was taken to form a composite perceived publicness score for each participant ($M = 3.88$, $SD = 1.32$, $\alpha = .89$).

Moderating Variables

Trait Self-Esteem. In the initial questionnaire which participants completed prior to viewing any advertisements, self-esteem was measured using the 10-item Rosenberg self-esteem scale (Rosenberg, 1965). Example statements include, “I feel that I have a number of good qualities,” and “I wish I could have more respect for myself.” Participants rated each statement on a four-point scale (1 = *strongly disagree*, 2 = *disagree,* 3 = *agree,* and 4 = *strongly agree*), and an overall score was calculated by taking the average of these statements

for each individual (M = 2.85, SD = 0.46, α = .88). See Appendix C for a full list of statements and scoring information.

Causal Attributions. The extent to which people's attributions of personalized Instagram ads were internal or external was measured using a subset of questions from the Revised Causal Dimension Scale (CDSII) developed by McAuley, Duncan, and Russell (1992). Participants were instructed to think about the reason(s) why the system showed them the ClassPass advertisement, describe the reason(s) in an open-ended question, and rate the reason(s) on a seven-point semantic differential scale. The scale included three items which asked participants whether the cause was something that: reflects an aspect of the ad system—reflects an aspect of yourself; that does not have to do with you—that has to do with you; something about the ad system—something about you. These three items were averaged to compute a composite score for each individual, where higher values indicate a more internal locus of causality and lower values indicate a more external locus of causality (M = 4.73, SD = 1.66, α = .88).

Dependent Variables.

Appearance Self-Concept. Participants completed several subscales from the Self-Perception Profile for College Students (Neemann & Harter, 2012). Each subscale included four semantic differential statements along a four-point scale. In my analysis, I focused on the appearance subscale as it was the most related to the ClassPass advertisement for fitness classes, although participants completed four additional subscales in order to reduce demand characteristics. In this subscale, people were shown statements such as, "Some people are not happy with the way they look BUT Other people are happy with the way they look," and, "Some people wish their body was different BUT Other people like their body the way it is,"

and had to decide which side of the scale best reflected them (see Appendix D for the full instructions and list of statements). The four appearance subscale items were averaged to compute a composite score for each individual, where higher scores indicated more favorable beliefs about one's appearance (M = 2.33, SD = 0.78, α = . 82).

Privacy Concern. To measure the extent to which participants believed the OBA system posed a risk to their privacy, I averaged participants' responses to a series of three statements adapted from Boerman et al.'s (2018) measure of online privacy risk severity: "Having Instagram collect my online behavior is a problem for me," "Having Instagram use my online behavior to show me advertisements is a problem for me," and "Having Instagram share my online behavior with other companies is a problem for me." Participants rated these statements on a seven-point scale (1 = *strongly disagree*, 7 = *strongly agree*), and an average was taken to form a composite score for each participant (M = 4.58, SD = 1.06, α = .75).

Desire for Greater Privacy Control of OBA. Finally, I included an original measure of participants' desire for greater informational privacy control relevant to OBA using the statement, "I want to have greater control over how Instagram uses my personal data to show me advertising," (M = 5.48, SD = 1.17). I also included two statements used in recent research by Pew: "There should be more government regulation of what companies like Instagram can do with their customers' personal information" (Pew Research Center, 2019; M = 5.53, SD = 1.23) and "I believe that people should have the right to have certain information about themselves permanently deleted by the people or organizations who have that information" (Pew Research Center, 2020; M = 6.11, SD = .94). Participants rated these statements on a seven-point scale (1 = *strongly disagree*, 7 = *strongly agree*). Inter-item agreement was not high enough to combine these items into a composite measure, α = .67.

Results

Manipulation Checks

Advertising Favorability

First, it was expected that those who saw the unfavorable version of the ClassPass advertisement should have had a less favorable attitude towards it and experienced more negative emotion in response to it than those who saw the favorable version. A 2 (*advertising favorability:* favorable or unfavorable) x 2 (*personalization:* non-personalized or personalized advertising) ANOVA was used to test for differences in how favorably people perceived the ad in each experimental condition. Based on a marginally significant effect, those who saw the unfavorable version perceived it more unfavorably ($M = 3.68$, $SD = 1.18$) than those who saw the favorable version ($M = 3.28$, $SD = 1.18$), $F(94) = 3.57$, $p = .06$. This analysis was repeated using negative self-conscious emotion as the dependent variable. Those who saw the unfavorable advertisement experienced significantly more negative self-conscious emotion ($M = 2.67$, $SD = 1.54$) than those who saw the favorable version ($M = 2.06$, $SD = 1.21$), $F(94) = 4.45$, $p < .05$. These results suggest that the favorability of the ad message was manipulated successfully in this experiment.

Advertising Personalization

Second, it was also expected that those who were told that the advertisements they saw in the experiment were personalized should have perceived the advertisements as more personalized and higher in relevance than those who were told that the ads were not personalized. A 2 (*advertising favorability:* favorable or unfavorable) x 2 (*personalization:* non-personalized or personalized advertising) ANOVA was used to test for differences in perceived personalization by experimental condition. Unfortunately, the manipulation was

not successful: there was no statistical difference in perceptions of personalization for those who were told that the advertisements were personalized ($M = 4.05$, $SD = 1.47$) compared to those who were told that the advertisements were not personalized ($M = 3.70$, $SD = 1.25$), $F(94) = 1.70$, $p = .20$.

As a second step, I considered whether the manipulation of personalization worked differently when people were shown a favorable or an unfavorable ad. There was no main effect of ad favorability on perceived personalization, $F(94) = .04$, $p = .85$, but there was a significant interaction effect, $F(94) = 6.48$, $p = .01$. That is, among the participants who saw the more favorable ad, those who were told that it was personalized perceived significantly greater ad personalization ($M = 4.38$, $SD = 1.60$) than those who were told that it was not personalized ($M = 3.33$, $SD = 1.18$). However, this trend was reversed among those who saw the unfavorable ad.[4] Those who were told that the unfavorable ad was personalized actually perceived less ad personalization ($M = 3.74$, $SD = 1.31$) than those who were told that it was not personalized ($M = 4.08$, $SD = 1.23$). This unexpected finding is explored in greater detail in the discussion section. Group means for the other variables measured in this experiment can be found in Table 2.

This analysis was repeated using perceived ad relevance rather than perceived ad personalization. In contrast to the test for differences in perceived personalization, there was a main effect of personalization on ad relevance: those who were told that the advertisements

[4] Post-hoc pairwise comparisons showed that perceived personalization was significantly higher among those in the personalized-favorable condition ($M = 4.38$, $SD = 1.60$) than those in the not personalized-favorable condition ($M = 3.33$, $SD = 1.18$), $p < .05$, but perceived personalization did not differ between those in the personalized-unfavorable condition ($M = 3.74$, $SD = 1.31$) and those in the not personalized-unfavorable condition ($M = 4.08$, $SD = 1.23$), $p = .36$.

were personalized for them rated the ClassPass ad as significantly more relevant (M = 4.68, SD = 1.47) than those who were told that they were not personalized (M = 4.12, SD = 1.40), $F(94) = 3.75$, $p = .05$. There was no main effect of ad favorability on relevance, $F(94) = .31$, $p = .58$. However, there was a marginally significant interaction effect of the manipulated variables on ad relevance, $F(94) = 3.53$, $p = .06$, such that among those who saw the more favorable version of the ad, those who were told it was personalized rated it higher in relevance (M = 5.05, SD = 1.47) those who were told it was not personalized, (M = 3.93, SD = 1.34), whereas, for those that saw the unfavorable ad, there was no difference in perceptions of ad relevance between those who were told it was personalized (M = 4.34, SD = 1.42) and those who were told it was not personalized, (M = 4.32, SD = 1.46), $F(46) = .00$, $p = .97$.

In short, the personalization manipulation was more effective among those who saw the more favorable ad compared to those who saw the unfavorable version of the ad. Given these findings, all subsequent analyses include a measure of perceived personalization as a covariate.

Advertising Effects on Appearance Self-Concept

Hypothesis 1 predicted that personalized advertisements would lead to greater self-concept change than non-personalized advertisements. Specifically, I looked for self-concept differences related to perceptions of one's own appearance which might have been made salient by seeing advertising for ClassPass, a fitness app. This shift in self-concept was expected to be either positive or negative depending on the favorability of the OBA. I expected that those who saw the negative ad would rate their appearance self-concept significantly *lower* when they were told that the ad was personalized compared to when they

were told it was not personalized, and those who saw the positive ad would rate their appearance self-concept significantly *higher* when they were told that the ad was personalized compared to when they were told it was not personalized.

A 2 (*advertising favorability:* favorable or unfavorable) x 2 (*advertising personalization:* non-personalized or personalized) ANCOVA was used to test for the effects of the manipulation of personalization on appearance self-concept as it varied by the favorability of the advertising that participants viewed, while controlling for within-group differences in perceived personalization. There was a significant interaction effect[5] of advertising personalization and message on people's perceptions of their appearance, $F(94) = 4.81$, $p < .05$, $\eta_p^2 = .05$, which indicated initial support for H1. To further understand this interaction effect, I split the sample into those saw the favorable version and those who saw the unfavorable version of the ad. Next, I used separate univariate ANCOVAs to determine whether there was an effect of personalization on appearance self-concept in each group while controlling for perceived personalization. Among the participants who saw the more favorable advertisement, there was surprisingly little difference in perceptions of appearance self-concept for those that were told the advertisement was personalized ($M = 2.37$, $SD = .63$) compared to those who were told that it was not personalized ($M = 2.30$, $SD = .91$), $F(46) = .00$, $p = .96$. However, among the participants who saw the unfavorable advertisement, those who were told it was personalized rated their appearance self-concept lower ($M = 2.01$, $SD = .72$) than those who were told that the advertisement was not personalized, who rated their appearance higher than those in the other conditions, on average ($M = 2.62$, $SD = .72$), $F(47)$

[5] This interaction effect was significant with or without including perceived personalization as a covariate, $p < .05$.

= 9.70, $p < .01$, $\eta_p^2 = .18$. A plot depicting these results can be found in Figure 1. There results partially supported Hypothesis 1.

Comparison to the Control Group

An additional test of H1 was conducted to compare the mean appearance self-concept ratings of each of the four experimental conditions to the control condition ($M = 2.35$, $SD = .81$). To accomplish this comparison, a five-level categorical variable denoting experimental condition (personalized-favorable, non-personalized-favorable, personalized-unfavorable, non-personalized-unfavorable, or control) was computed and entered as the independent variable in a one-way ANOVA with contrasts. Each of the four groups who were exposed to advertising were assigned a contrast coefficient of 1, and the control group a coefficient of -4. This analytic technique for a dangling group design is outlined by Tabachnick and Fidell (2007, pp. 180-181). A Tukey HSD post-hoc test revealed that none of the mean appearance self-concept ratings for any of the experimental conditions significantly differed from that of the control group who did not see any advertising ($M = 2.35$, $SD = .81$). Thus, although appearance self-concept was significantly lower among those who saw unfavorable advertisements that were personalized compared to those who saw unfavorable ads that were not personalized, appearance self-concept among those who saw unfavorable advertisements that were personalized was *not* significantly lower than those who saw no advertising at all.

Perceived Publicness as a Mediator

Hypothesis 2 predicted that the extent to which one feels publicly identifiable will mediate the effect of unpleasant personalized advertising on self-concept. A moderated mediation model with bootstrapping was specified using the PROCESS macro for SPSS (Hayes, 2013). Dummy variables denoting advertising personalization (non-personalized (0)

or personalized (1) advertising) and favorability (unfavorable (0) or favorable (1)) as well as their interaction term were entered as predictors of both perceived publicness (the mediating variable) and appearance self-concept (the outcome), and perceived personalization was entered as a covariate. The model specified explained 36.2% of the variance in perceived publicness, $F(94) = 11.80$, $p < .001$, and 8.3% of the variance in appearance self-concept, which was not significant, $F(94) = 1.87$, $p = .11$. There was no significant main effect of ad personalization on perceived publicness ($\beta = .44$, $p = .19$), no significant main effect of ad favorability on perceived publicness ($\beta = .14$, $p = .63$), and no significant interaction effect ($F(94) = 1.24$, $\beta = -.51$, $p = .27$). Only perceived personalization predicted perceived publicness, $\beta = .58$, $p < .001$. Furthermore, perceived publicness did not predict appearance self-concept, $\beta = .04$, $p = .62$. Finally, there were no conditional indirect effects of personalization on appearance self-concept via perceived publicness whether people saw unpleasant ($\beta = .02$, 95% CI [-.06, .11]) or neutral ads ($\beta = .00$, 95% CI [-.06, .05]). Thus, there was no evidence of mediation,[6] and Hypothesis 2 was not supported.

OBA & Perceptions of Source of Publicness

Research question 1 asked about perceptions of audience. That is, *who* or *what* would participants expect to serve as an audience for their personalized advertisements? Specifically, participants were asked to rate the extent to which they felt a number of entities would associate them with the ad they saw, including: the ad personalization system,

[6] The perceived publicness variable used to test H2 was computed by taking the average of perceptions of the presence of different types of audiences: the ad personalization system, Instagram employees, advertisers, the researchers, and their Instagram followers. However, I also ran the model using only the item measuring perceptions of "*the ad system*", and the results were similar. In addition, in an exploratory analysis I tested whether perceived publicness *moderated* rather than mediated the effect of the manipulation on appearance self-concept, but once again the effect was non-significant.

Instagram employees, advertisers, the researchers, and their Instagram followers. To answer this question, I filtered the dataset to only include those in the personalized advertising conditions. Next, I conducted a one-way ANOVA with blocking in which perceived publicness was the outcome and the entity that participants were rating (the ad system, Instagram employees, etc.) was the blocking variable. The effect of the blocking variable on perceived publicness was not significant, $p = .45$. Furthermore, Tukey's post-hoc test did not reveal any significant pairwise comparisons between these items. These findings suggest that those who saw personalized advertisements did not feel more observed by one particular entity over another.

Trait Self-Esteem as a Moderator

Hypothesis 3 posited that self-esteem would moderate the effect of advertising personalization and favorability on self-concept. A moderation analysis was conducted using the PROCESS macro for SPSS (Hayes, 2013) to determine the effect of ad personalization and favorability on appearance self-concept at different levels of trait self-esteem. Ad personalization (not personalized = 0 or personalized = 1), favorability (unfavorable = 0 or favorable = 1, and a trait self-esteem composite score (centered) were entered as predictors of appearance self-concept, as were their interaction terms. Perceived personalization was included as a covariate. The model explained 25.2% of the variance in appearance self-concept, $p < .001$. However, trait self-esteem was the only variable in this model that significantly predicted appearance self-concept ($\beta = .64$, $p < .05$). The effects of ad personalization ($\beta = -.45$, $p = .08$) and favorability ($\beta = -.32$, $p = .17$) on appearance self-concept were not significant, nor was their interaction effect significant ($\beta = .51$, $p = .14$). Furthermore, there were no significant interactions between self-esteem and ad

personalization (β = -.06, p = .91) or between self-esteem and ad favorability (β = .29, p = .51). Finally, there was no 3-way interaction between ad personalization, ad feedback, and self-esteem (β = -.33, p = .64). Hypothesis 3 was not supported.

Causal Attributions as a Moderator

Hypothesis 4 predicted that self-concept change would be greater among those who made more internal attributions about why they saw a certain ad. A moderation analysis was conducted using the PROCESS macro for SPSS (Hayes, 2013) to determine the effect of ad personalization and favorability on appearance self-concept at different levels of trait self-esteem. Ad personalization (not personalized = 0 or personalized = 1), favorability (unfavorable = 0 or favorable = 1, and a measure of the extent to which one made internal attributions about the ad (centered) were entered as predictors of appearance self-concept, as were their interaction terms. Perceived personalization was included as a covariate.[7] The model explained 11.5% of the variance in appearance self-concept, which was not a significant portion, p =.08. The effects of attributions (β = .11, p = .39) and ad favorability (β = -.36, p = .17) on appearance self-concept were not significant, but the effect of ad personalization was significant (β = -.68, p < .01). Once again, there was a significant

[7] Given the significant bivariate correlation of r = .40, p < .01, between perceived personalization and causal attributions, indicating some redundancy between these variables, the conditional moderation model was respecified without including perceived personalization as a covariate. This model specified explained 9.9% of the variance in appearance self-concept, $F(94) = 2.29$, p = .05. There was a significant main effect of ad personalization on appearance self-concept (β = -.67, p < .01) and a significant interaction effect of ad personalization and favorability (β = .74, $F(94) = 5.24$, p < .05). In this case, there was a conditional effect of the manipulation on self-concept depending on people's causal attributions such that among those who saw the unfavorable ad, those who were told it was personalized for them rated their appearance self-concept significantly lower than those who were told the ad was not personalized *only if* they made slightly internal (β = -.67, p < .01) or very internal attributions (β = -.78, p < .01) about why they saw that particular ad, but not if they made more external attributions (β = -.55, p = .06).

interaction effect of ad personalization and favorability on self-concept (β = .77, p = .05). However, there were no significant interactions between attributions and ad personalization (β = -.17, p = .23) or between attributions and ad favorability (β = -.12, p = .50). Finally, there was no 3-way interaction between ad personalization, ad feedback, and attributions (β = .25, p = .41). The model did not reveal any conditional effects of ad personalization and message on self-concept among people with different attributions (evaluated at the scale mean and M ± 1SD). In sum, attributions did not have an additive effect on self-concept on top of the manipulation and perceived personalization, and hypothesis 4 was not supported.

Advertising Effects on Privacy Concern

Hypothesis 5 proposed that those exposed to more unpleasant personalized advertising would be more concerned about their online privacy and more likely to desired greater privacy control than those exposed to less unpleasant personalized advertising. A 2 (*advertising unpleasantness:* unpleasant or neutral) x 2 (*personalization expectations:* expect non-personalized or personalized advertising) Bonferroni-corrected MANCOVA was used to test for between-group differences in privacy concern and each of the three opinion statements about online privacy while controlling for perceived personalization. None of the dependent variables differed significantly between experimental conditions, p > .1, which meant that Hypothesis 5 was not supported.

Discussion

This study attempts to build upon previous research on self-concept change following exposure to personalized advertising (Summers et al., 2018) and other forms of algorithmically-generated feedback on self-concept (Carr & Hayes, 2018; French, 2018; Walther et al., 2011). Historically this work has focused on the effect of positive or neutral

feedback rather than negative feedback. The primary goal of this study was to determine whether unfavorable personalized advertising, even that which does not clearly reference a recent instance of self-presentation, can also shift self-concept (H1). Given the wealth of potentially ego-threatening or "negative" feedback one might receive online (e.g. weight loss products, skin treatments, counseling services, etc.), this test was intended to expand the external validity of the identity shift construct while also testing perceived publicness as a mediator of this effect (H2). In addition, I tested two individual differences as moderators of H1: trait self-esteem (H3) and the extent to which people made internal attributions for why they saw a particular advertisement (H4). Finally, I tested whether exposure to unfavorable OBA had a greater effect on online privacy concern compared to favorable OBA (H5).

Before summarizing the results of this study, it is important to note the null result of the ad personalization manipulation check among those who saw the unfavorable version of the ClassPass ad. That is, in the unfavorable ad conditions, perceptions of ad personalization did not differ between those who were told that the advertisements were personalized for them and those who were told that the advertisements were not personalized. Therefore, any analyses ran using the ad personalization manipulation for the unfavorable ads are exploratory in nature, and it is not possible to conclude with certainty that the differences between conditions were a result of the manipulation. It should also be noted that my approach to dealing with the failure of the manipulation check was a bit unconventional. Specifically, I decided to include the manipulation check measure of perceived personalization as a covariate when testing my hypotheses. According to a recent review article on the use of manipulation checks in experimental social psychology in the past few years have used manipulation checks, only 1.2% of studies sampled had included a

manipulation check as a covariate in statistical analyses (Ejelöv & Luke, 2020). Rather, manipulation checks were more commonly used as exclusion criteria; however, this was not an option in the present study due to the small sample sizes in each condition and time constraints on data collection. Nevertheless, if I were to remove perceived personalization as a covariate in my analyses, the results would be largely unchanged apart from the test of Hypothesis 3, which I will describe later in more detail.

Given the caveats mentioned above, the results yielded mixed support for my central hypothesis—those who saw a personalized advertisement which featured unfavorable feedback about their appearance rated their appearance self-concept lower than those who saw a similar ad but were told that it was not personalized. In contrast to previous research, however, this study did not find an effect of favorable feedback in the form of OBA on self-concept. As I will discuss later in more detail, the reason for this null result may have had to do with the experimental stimuli used for the favorable feedback conditions.

Theoretical Implications

The significant effect of ad personalization on self-concept, albeit only for those exposed to the unfavorable advertisement, bolsters the findings by French (2018). Taken together, these studies demonstrate that algorithmically-personalized messages about the entirety of one's online self-presentation on a digital platform such as Instagram or Facebook can constitute a form of feedback and shift one's self-concept. Previous research has found that perceptions of an audience can enhance the effect of strategic, contextually-motivated self-presentation on self-concept in the absence of interpersonal feedback (Gonzales & Hancock, 2008; Kelly & Rodriguez, 2006; Schlenker et al., 1994; Tice, 1992). Receiving feedback strengthens these audience effects, even when the feedback is computer-generated

(Carr & Hayes, 2019; Walther et al., 2011). The present study and the study by French (2018) contribute to this body of literature by showing that self-concept change can occur in the absence of clear awareness of one's own self-presentation, but merely in response to feedback about how one presumably presented oneself at some point in the past. This is consistent with recent studies which have found that, upon reviewing the ways in which Google and Facebook categorized them for advertising purposes, people believed that surprising and incorrect inferences made about them must be connected to their online behavior in some way, even if they could not fully articulate a precise explanation (Eslami et al., 2018; Hautea et al., 2020). Thus, it was hypothesized that even when people view OBA that does not clearly reference a specific self-presentation, people might still envision having had an audience for their past online behavior and experience a shift in the self-concept. These findings yield mixed support for that hypothesis.

This study makes another novel contribution to a body of literature on people's responses to algorithmic interferences by measuring reactions to unfavorable OBA, given that negative feedback may have a different effect on self-concept than positive feedback. Self-enhancement leads people to often interpret negative self-presentations and feedback in ways that minimize the threat to one's ego (Baumeister, 1982; Gonzales et al., 2018; Kunda, 1987; Sedikides, 1993; Shrauger & Lund, 1975), which can involve distorting the feedback or discrediting its source (Crary, 1966; Kunda, 1987). For instance, in a study of people's responses to explanations for personalized advertising on Facebook, Eslami et al. (2018) found that people were dissatisfied when the algorithm made incorrect and unflattering inferences about them, but not when incorrect assumptions were flattering (although the authors did not determine whether these experiences resulted in self-concept change). In the

present study, people who saw an unfavorable advertisement did not feel that it was more personalized or relevant when they were told that it was personalized for them compared to when they were told that it was not personalized. Perhaps, given the tendency towards self-enhancement, people might have been motivated to view the unfavorable advertisement as a random or dysfunctional output of the advertising system rather than viewing it as negative feedback about themselves. Nevertheless, those who were told that the unfavorable advertisement was personalized rated their appearance self-concept significantly lower than those who were told that it was not personalized.

Given the pervasive algorithmic systems that make inferences about who we are, it is critical to understand the effects of this process on one's sense of self. These systems have been known to generate incorrect inferences or pick up on less favorable aspects of one's identity (Eslami et al., 2018; Hautea et al., 2020), both of which can lead people to encounter messages implying that they possess any number of undesirable traits. Encountering these types of messages is all the more likely considering how algorithms observe some of our most private and sensitive online behavior, often without our conscious awareness and thoughtful discretion. The present research shows that even when people are unsure why exactly an algorithm has made an assumption that it did, the algorithm might still influence how they see themselves in some contexts. Perhaps due to the belief that algorithms know us better than we know ourselves, people sometimes defer to the judgement of the algorithms even when doing so contradicts their pre-existing self-views (Eslami et al., 2018).

As a result of frequent exposure to presumably personalized messages that temporarily activate particular self-concept, this self-concept may become more *accessible,* or more readily used for social information processing (Barg et al., 1986; Higgins & King,

1981; Higgins, 1987). For instance, if some people are constantly exposed to personalized messages telling them that they are out of shape, and these messages lead to a negative shift in their appearance self-concept, this self-view will become chronically accessible to them and have a lower threshold of activation compared to other self-views when processing ambiguous stimuli in the future (Barg et al., 1986; Higgins, 1987). Therefore, even when these people encounter messages about fitness and weight loss that are not clearly personalized or are less decidedly negative, their negative view of their appearance would be activated more easily than it would in people who do not regularly encounter unfavorable OBA about their appearance. Such chronic activation could have the beneficial outcome of motivating someone to begin exercising more often, or alternatively, it could damage one's body image or encourage extreme weight-loss behavior.

Practical Implications

The insight gained from this study also has important implications for marketers and designers of consumer-facing algorithmic content curation systems who wish to optimize their digital marketing strategy and mitigate negative responses from their audiences. Online platforms containing advertising might want to be careful when featuring products that highlight less favorable traits. Numerous studies on consumer behavior have found that marketing messages which make salient a more favorable aspect of one's identity generally tend to have a positive effect on brand attitudes and purchase intentions (e.g. Deshpandé & Stayman, 1994; Reed, 2004; Thomas et al., 2015), yet few have focused on the negative implications of messages that appeal to more negative or unfavorable aspects of one's identity (Thomas et al., 2015), and none to my knowledge have looked at this in the context of OBA. Thomas et al. (2015) found that people who viewed a traditional advertisement

which “aired their dirty laundry” viewed the advertised brand less favorably, but only when viewing the ad in the company of others. As noted, Summers et al. (2016) found effects of perception when OBA confers a social label, but only tested this theory using OBA that conferred positive or neutral traits (e.g. sophisticated, environmentally conscious) rather than more negative traits (e.g. out of shape). In sum, it is not entirely clear how consumers’ identification with unfavorable OBA would impact brand attitudes and purchase intentions. Therefore, given my findings, digital marketers might think twice before they behaviorally target people with advertisements that might be ego-threatening or induce negative self-conscious emotions. In fact, Facebook already prohibits advertisers from displaying personal health-related advertisements that are “likely to trigger a negative reaction, such as an ad that gives an individual a negative perception of their body or figure” (Facebook Business Help Center, 2020), although this policy seems rather vague and subjective, and the degree to which Facebook actually enforces it is unknown.

Additionally, platforms should develop features that let people play a more active role in determining the types of personalized advertisements that they see. Some tools do exist currently to regulate one’s exposure to OBA, which include features developed by specific platforms such as Google or Facebook and those designed to work across the web via one’s browser (e.g., AdBlock Plus). While these tools allow people to block certain formats of personalized advertising, many of them fall short of allowing people to block certain categories of advertisements or ones that reference particular aspects of their identity (Agarwal et al., 2013). Many consumers recognize that there are some benefits to OBA such as more useful advertising (Ur et al., 2012), so blocking OBA entirely would not be an ideal solution for them. Tools such as Facebook’s ad settings do in fact allow users to remove

certain ad categories from their profiles, but many users are not aware of this tool, some are reluctant to use this tool due to uncertainty about how it will impact the ads they see on Facebook, and others doubt that editing their profiles will have any effect on the ads they see on the platform (Rao et al., 2015). Clearly the digital advertising industry still has substantial progress to make towards enabling self-regulation on the part of consumers.

Finally, these findings point to the reasons for raising consumer awareness of platform privacy control options. This might include making platform-specific features more prominent, nudging users to adjust their features, or providing more information about how to use these features and how they actually affect the advertising one sees. Encouraging people to review or edit their ad categories could potentially mitigate the effect of OBA on their self-concept. When Eslami et al. (2018) asked participants to review the ways in which websites such as Google and Facebook categorized them for advertising purposes, many people's first instinct was to justify inaccurate or surprising inferences made about them by these algorithms. Eventually, however, encountering enough inaccurate inferences led some people to experience *algorithmic disillusionment*, or "the realization that advertising algorithms were not as perceptive and powerful as users thought" (Eslami et al., 2018, p. 9). Those who have had this realization might be more critical of algorithmic feedback and less susceptible to the effect of negative algorithmic feedback on their self-concept. In other words, once people learn that algorithmic personalization systems have limitations and do not always magically generate spot-on inferences about them, the outputs of these algorithms (e.g., OBA) might have less of an impact on people's self-views.

No Effect of Mediators or Moderators on Self-Concept Change

Prior studies on self-concept change following exposure to OBA (Summers et al., 2016) or other types of algorithmic feedback (French, 2018) had not clearly identified a mechanism explaining why these effects occurred. Summers et al. suggested that OBA conferred a social label from an external agent (i.e., marketers), and French (2018) argued that an algorithmically-generated career recommendation based on one's Facebook data constituted a form of algorithmic feedback. Nevertheless, neither study tested whether people's perceptions of marketers, algorithms, or other agents observing their past online behavior actually mediated the effect of OBA or algorithmic feedback on self-concept. Previous experiments manipulating the extent to which people perceive that their self-presentation is public (versus private) have found that self-concept change is more likely to occur when people perceived a greater sense of publicness (Gonzales & Hancock, 2008; Kelly & Rodriguez, 2006; Schlenker et al., 1994; Tice, 1992). Therefore, it seemed highly possible that the effect of OBA or algorithmic feedback on self-concept are driven by perceived publicness. In fact, Walther et al. (2011) had tested whether receiving feedback from an algorithm about a recent, otherwise-private self-presentation would heighten perceptions of publicness compared to those who engaged in private self-presentation but did not receive feedback. Although this study did not find a significant effect of algorithmic feedback on perceived publicness, I decided to retest this possibility using a revised publicness measure and predicted that perceived publicness would mediate the effect of feedback delivered in OBA on one's self-concept (H2).

This study did not find any evidence that perceptions of publicness mediate the effect of OBA on self-concept, however. The reason for this null result is not entirely clear. It is possible that there is another mechanism responsible for the self-concept shift observed in

this study. Perhaps OBA heightens *private self-awareness*, or the amount of attention paid to one's innermost thoughts and feelings or personal memories (Duval & Wicklund, 1972), to a greater degree than non-personalized advertising, but OBA does not affect *public self-awareness*, or how one believes oneself might appear to others, which is what our measures of publicness seem to be operationalizing. Another possibility is that algorithmic feedback does create some sense of having had an audience for one's past behavior, but this feeling eludes measurement because having a non-human audience does not constitute publicness in the traditional sense. Since previous research has found a relationship between heightened self-awareness or the presence of an audience, and greater physiological arousal (Carver & Sheier, 1981), future studies might substitute or supplement a self-report measure of this mediating variable with a measure of physiological arousal (e.g., Galvanic skin response).

Nonetheless, the findings generated from the refined measure of perceived publicness used in the present study generated some additional insight about this phenomenon. Specifically, the results of this study do not seem to support the claim made by Summers et al. (2016) that people envision advertisers or marketers as the source of feedback in OBA—participants exposed to OBA were no more likely to feel that advertisers had associated them with a certain advertisement than had Instagram employees, their Instagram followers, the researchers, or the ad personalization system (RQ1). Understanding the audience that people imagine when exposed to OBA is important in light of previous work showing that characteristics of one's audience or the source of interpersonal feedback can affect self-perception (Carr & Foreman, 2019; Fedor et al., 2001; French & Hancock, 2016). Future work might address how various perceived attributes of the audience people imagine when

viewing OBA (e.g., audience size, trustworthiness, how much they value people with a certain trait) moderate subsequent self-concept change.

This study also examined the role of trait self-esteem (H3) and causal attributions (H4) as moderators of the effect of personalization on self-concept. Previous research utilizing self-verification theory had found that people with low self-esteem are more likely to accept negative feedback than those with high self-esteem because it is consistent with their pre-existing self-views (Swann, 1983). Furthermore, Kim & Gonzales (2018) found that those with low self-esteem were more likely to accept negative feedback when they perceived a higher degree of publicness. Therefore, I expected that unfavorable OBA would have a greater effect on the self-concept of people with lower self-esteem compared to those with higher self-esteem, and self-concept change would be greater among people with low self-esteem when they encountered unfavorable personalized advertising (which was presumed to heighten perceived publicness) compared to unfavorable advertising that was not personalized. Unsurprisingly, there was a significant main effect of trait self-esteem on self-concept, yet there were no interaction effects between trait self-esteem, ad personalization, and ad favorability. H3 was not supported: there was no evidence that unfavorable OBA has a greater effect on the self-concept of people with lower self-esteem compared to those with higher self-esteem.

It is possible that the predictions made by self-verification theory about people's responses to negative feedback depending on their self-esteem levels only apply when the feedback comes from another person and not an algorithm (e.g. Mishra, 2006). In other words, it might be more difficult for people to discount algorithmically-generated negative feedback than is it for them to dismiss negative interpersonal feedback. In a study that

compared the effects of performance feedback from a human versus a computer on self-perception, Mishra (2006) argued that, "people accept feedback from the computer at face value. In the case of receiving feedback from humans, people are more interpretive, and seek to understand the context of the feedback and this is not something they do when working with computers" (pp. 125-126). In other words, whereas people might possess or seek out a depth of knowledge about the individuals who supply them with negative feedback and therefore more easily form explanations for why they received the feedback they did, people might find it challenging to gain adequate knowledge about the algorithms that supply negative feedback and have more difficulty forming explanations about how these systems operate.

Furthermore, also contrary to my expectations, causal attributions did not moderate the effect of ad exposure on self-concept, and H4 was not supported. This was surprising, as French (2018) had found that those who made more internal attributions to explain why they received a certain algorithmically-generated career recommendation were more likely to experience self-concept change. French argued that those who formed more internal attributions might be more aware of how their online behavior would be perceived by an external agent, and this awareness is presumed to cause a shift in their self-concept according to the literature on self-perception. Prior research in other domains had also found that people's causal attributions for favorable and unfavorable outcomes predicted their self-views (Cadinu et al., 2006; Kanoy et al., 1980). The lack of a moderating effect of attribution in this study was in part to the fact that I included perceived personalization as a covariate when testing this hypothesis due to the failure of the manipulation check, but perceived personalization correlated highly with the attributions measure and created a redundancy

effect. When including either perceived personalization or causal attributions in the model, each variable had a significant effect on self-concept, but this effect was non-significant when including both variables. Although causal attributions and perceived personalization were not perfectly correlated, perhaps the two are not conceptually distinct, in which case it would not be necessary to include both in a single predictive model. After all, those who believe that an advertisement has been personalized for them should also feel that the reason they are seeing that particular ad has something to do with their own traits or behavior, and vice versa. Additional research is needed to disentangle the effects of causal attributions and perceived personalization on self-concept and determine whether there is a distinction between the two.

No Effect of the Advertising Manipulation on Privacy Concern

Finally, this study did not find any significant effects of ad personalization or favorability on online privacy concern. To my knowledge, little research has been done to simultaneously study the effects of OBA on self-concept as well as perceptions of privacy. However, I felt it was important to test and expand upon this relationship given the link between OBA and privacy concern identified consistently in previous research (Agarwal et al., 2013; Boerman et al., 2018; Eslami et al., 2018; Phelan et al., 2016; Ur et al., 2012; Zhu & Chang, 2018). Zhu & Chang (2018), for instance, found that self-awareness mediates the relationship between perceived OBA relevance and privacy concern such that those who perceived greater ad relevance and were more self-aware had less privacy concern. They argued that when people are made to feel more self-aware, as is the case when they received presumably personalized advertising, they are more willing to accept responsibility for positive or negative outcomes. However, the authors did not appear to use advertisements

that were particularly unfavorable or offensive, nor did they measure advertising favorability. Therefore, it was possible that unfavorable personalized advertising would increase privacy concern whereas favorable personalized advertising would decrease privacy concern. Additional research is needed here to further explore these findings.

Ultimately, it is difficult to determine whether there is truly no relationship between these factors or if the lack of findings was due to improper measurement. For instance, most of the privacy-related items which I adapted from other studies (Boerman et al., 2018; Pew Research Center, 2019) specifically involved privacy concerns related to Instagram, despite the fact that the company was not directly involved with the ad-personalization process used in this study. As a result, people might not have been concerned about Instagram per say having access to the data they have shared with the platform, but instead might have been more concerned with other parties gaining access to their Instagram data. In addition, online privacy concern and desire for greater privacy control were both fairly high across conditions, indicating a possible ceiling effect. Moreover, given the privacy paradox, or the tendency for people to report more concern for their online privacy than is reflected in their behavior (Norberg et al., 2007), future research might consider measuring a privacy-related behavioral outcome. For example, in addition to measuring privacy attitudes, studies might test whether participants modify their privacy settings or minimize their self-disclosure online following exposure to unfavorable personalized advertising.

Limitations

As with any research, this study had limitations that must be addressed. First, it should be noted that only about half of the participants who completed the initial survey actually provided us with their Instagram data. Although some participants may have been

confused about how to download their Instagram data or simply did not feel it was worth the trouble, others may have felt uncomfortable doing so because of the sensitive, personal nature of that information. Indeed, a few participants communicated these privacy concerns with us via email after they were sent a reminder to download and send us their Instagram data, and ultimately, these individuals chose not to participate further. In other words, those who were most sensitive to perceptions of publicness concerning their Instagram data removed themselves from the study. Therefore, it is possible that ad favorability and personalization would have had a greater effect on those who chose not to participate in this study due to privacy concerns.

Second, although this study deals with the effects of personalized advertising, and participants were led to believe that they were seeing personalized ads, I did not actually show participants advertisements that had been personalized specifically for them as individuals. It is not uncommon for researchers studying the effects of message personalization to manipulate perceived personalization rather than actual personalization. For example, Summers et al. (2016) had participants generate a search history which was then used to deliver advertisements that were ostensibly personalized, when in fact the ads were the same for all participants. However, the study by Summers et al. showed participants ads presumably related to their recent search history, and the majority of identity shift studies have asked participants to perform some task in which they present themselves as possessing a certain trait such as extroversion before providing seemingly personalized feedback (e.g. Carr & Hayes, 2019; Walter et al., 2011), which is also based in precedent from public commitment, biased scanning, and self-perception studies whence the identity shift literature originated (e.g. Kelly & Rodriguez, 2006; Schlenker et al., 1994; Tice, 1992). In contrast, I

did not elicit any self-presentation from the participants in the present study. My procedure was inspired by that used by French (2018), who led participants to believe that participants were receiving career recommendations based upon the entirety of their Facebook data, when in fact these recommendations were generated randomly. This procedure had the benefit of providing enhanced external validity by more closely replicating the conditions of a typical encounter with OBA, which often does not reference a specific and recent act of self-presentation (Andreou et al., 2018). However, it is possible that the results of this study would have been different if I had used a procedure more similar to that used by Summers et al. (2016) or previous identity shift studies (Carr & Hayes, 2019; Gonzales & Hancock, 2008; Walther et al., 2011).

Next, there are some issues with the experimental manipulations used in this study. First, although the manipulation of ad favorability was successful, the favorable version of the ad may not have actually referenced a positive trait as intended, despite the fact that people reacted more positively to this version compared to the unfavorable version. Whereas the unfavorable version of the ad seemed to clearly provide negative feedback by insinuating that the viewer was lazy and out of shape, the favorable version was more subtle and did not explicitly provide positive feedback or necessarily insinuate that the viewer was in good shape. In other words, the "positive" feedback, was likely, in retrospect, *neutral* feedback. It did not do harm, but it did not actively enhance any particular self-concept. Given this shortcoming then, it is perhaps unsurprising that ad personalization did not have an effect on self-concept for those who saw the favorable version of the ad (i.e. those who saw the personalized ad did not feel better about their appearance than those who saw the non-personalized ad). Future research looking at the effect of OBA on self-concept should ensure

that experimental stimuli unambiguously reference a specific trait, be it favorable or unfavorable. Second, it is important to note the failure of the perceived personalization manipulation check among those who saw the unfavorable advertisement in the experiment, despite significant differences in perceived personalization during pilot testing. This is interesting given that the personalization manipulation check yielded significant results among those who saw the favorable version of the advertisement. These findings suggest that people may be more reluctant to acknowledge that they are part of the intended audience for a personalized advertisement conferring an unfavorable trait compared to an ad conferring a more favorable trait. Qualitative research could be useful to further disentangle this issue.

Another limitation of this study is that it relied upon several newly developed or revised measures of key concepts which have yet to be rigorously tested and validated. The newly developed measures included those of perceived personalization, perceived publicness, and desire for greater online privacy control, and the revised measures were those of causal attributions for algorithmically-generated messages and online privacy concern. There were some clear issues of construct validity with some of these measures. First, the measure of online privacy concern pertained to Instagram's collection and use of one's personal data for advertising purposes, yet participants were aware that Instagram was not involved in this experiment. Thus, it would have been more appropriate to measure privacy concern about OBA in general. Second, the measure of ad favorability might not have been appropriate given that the items pertained more to how much people liked the advertisements rather than how the advertisements made them feel about themselves (although I supplemented this with a measure of negative self-conscious emotions experienced in response to the advertising). There were also some potential issues of content validity. For

example, the items measuring desire for greater online privacy control did not correlate highly with one another, and therefore they could not be combined into a composite. Additional items may have been needed to measure this construct more robustly. Also, the measure of perceived publicness included the extent to which people felt identifiable by a number of entities (i.e., the ad personalization system, the researchers, Instagram employees, one's followers on Instagram, and advertisers), but this list is not exhaustive, nor does it include an item assessing how identifiable people felt in general.

To compensate for these measurement shortcomings, future research using these measures might conduct cognitive interviews to determine how people currently interpret each item and revise them accordingly. There are a couple of outstanding questions that such research could address. First, when people make causal attributions about why they saw a particular advertisement online, can their attributions reflect something internal (i.e., their own behavior or traits) as well as something external (i.e., the ad system, advertisers), or are internal and external attributions mutually exclusive? Second, does it make sense to compute a composite measure of perceived publicness encompassing how identifiable people felt to a number of different agents or, is feeling identifiable to an algorithm phenomenologically different than feeling identifiable to a person or group of people? These are just some of the questions that need to be addressed in future work that continues to test the boundary conditions of identity shift research by testing the construct in new digital contexts.

There is a final substantial limitation of these data that is important to note: On average, the subjects in the personalized-unfavorable ad condition had significantly lower trait self-esteem compared to those in the other conditions, despite random assignment. This is a problem because there is a well-documented positive correlation between trait self-

esteem and satisfaction with one's appearance (Harter, 2006). Thus, those who saw the unfavorable personalized ad might have had lower appearance self-concept ratings prior to the experimental manipulation. Nevertheless, among those who saw the unfavorable ad, the effect of the ad personalization manipulation remained significant after controlling for the effect of trait self-esteem. In short, it is difficult to conclude with complete certainty whether ad personalization truly had any effect on self-concept, or whether the findings were a result of between-condition differences in self-esteem. In order to rule out this possibility, I am currently collecting more data to hopefully achieve balanced levels of trait self-esteem within each condition.

Future Directions

The present study failed to establish whether heightened perceptions of publicness act as a mechanism for the effect of OBA on self-concept. That is, although people who saw personalized ads containing negative feedback about their appearance rated their appearance lower than those who were told that the ads were not personalized, it is unclear *why* exactly this effect occurred. One possibility is that the measures used to gauge perceptions of publicness were not adequate; for instance, this study measured the extent to which people felt that a number of different entities (e.g. the ad system, the researchers, advertisers) would associate them with a given advertisement without measuring perceptions of publicness in general. This was intentional in this study because the format of publicness questions from previous public commitment and identity shift research did not lend itself to the objectives of the present study (e.g. "To what extent do you think your presentation in this study was publicly identifiable?" (Tice, 1992, p. 439)). However, it is possible that my adaptation of this measure was poorly operationalized. A similar problem was noted by Walther et al. in

their 2011 study: “It may be that the measures we used to assess publicness were not sensitive to the presence of a small audience rather than a large one” (pp. 19-20). As a result, I opted not to measure perceptions of publicness in general, but rather perceptions of publicness with respect to several specific entities. This decision, however, meant that I was left testing a new question structure that had not been validated by use in previous contexts. Moving forward, researchers should carefully develop and test new ways to measure perceived publicness for one’s online behavior, as well as how perceived audiences might change depending on the context. Given a host of research identifying psychological implications of perceived publicness, a validated measure of this phenomenon that is not platform dependent yet maintains external validity would be valuable for continued research in this area. For instance, it would be interesting to see whether people imagine different audiences for the entirety of their past online behavior compared to when they are engaging in acts of self-presentation online, or whether people report less concern about their online privacy when they envision an algorithm as the audience for their online behavior rather than a human or a corporation.

Future studies on the effects of OBA on self-concept should also disentangle, if possible, feedback from publicness. Studies to-date on this topic have not yet done so (French, 2018; Summers at al., 2016). Instead of building on these studies by disentangling this particular confound, I chose to elaborate on whether public commitment is indeed the mechanism behind these findings, and whether these phenomena yield the same pattern of results when advertising is favorable versus unfavorable. Because research by Walther and colleagues (2011) demonstrated that feedback can have a magnifying effect of public commitment on self-concept change, a possible future study could manipulate whether

participants are reminded of the public nature of OBA systems (e.g. “Your data are being collected and analyzed in order to most effectively personalize your advertising in the future”). At the same time, one could also manipulate whether participants receive OBA or advertising that has not been personalized for them. Thus, this would be a 2 (publicness cue or no publicness cue) x 2 (personalized or non-personalized advertising) experiment. This experimental design could show whether cueing publicness moderates the effect of OBA on self-concept. Although testing this hypothesis was outside the scope of the current study, future work might tease apart the effects of feedback and publicness on self-concept using a design similar to the one mentioned here.

In future work, researchers might also consider how other individual differences moderate the effect of unfavorable OBA on self-concept. In the present study, advertisements were designated as objectively favorable or unfavorable depending on the reactions they gleaned from viewers in general. Nonetheless, the extent to which participants rated the favorability of these advertisements varied somewhat within each condition. It is possible that among some individuals, the advertisement highlighted a larger discrepancy between one’s actual self and one’s ideal self, or the person who one is presently versus who one aspires to be (Higgins, 1987; Markus & Nurius, 1986). When this discrepancy is larger, or when it is more accessible (e.g., when someone is prone to more negative self-perceptions of their appearance and the desire to improve it), self-concept change might be greater following exposure to OBA which highlights that discrepancy. Future studies in this area could also examine whether unfavorable OBA confirms or disconfirms people’s prior beliefs about their online self-presentation. For instance, Carr and Hayes (2019) found that identity-confirming or disconfirming feedback moderated the effect of online self-presentation on

self-perception. However, this effect has not yet been tested in a context where the feedback is delivered in the form of OBA and does not reference a specific instance of self-presentation. If people believe that their past behavior online reflects a desire to improve their appearance, for example, then they might be more likely to experience self-concept change in response to unfavorable OBA that confers negative feedback about their appearance.

Finally, additional research is needed to replicate the effect of unfavorable personalized feedback generated by an algorithm on self-concept using a wider variety of stimuli. This study only focused on the effect of OBA on appearance self-concept due to the prevalence of advertising that points out favorable or unfavorable aspects about one's appearance and the success of this particular manipulation during pilot testing. Future research should examine the effects of unfavorable OBA personalization on other self-concepts to determine whether this is a robust phenomenon, not limited to the self-concept of interest in this study. For instance, one might test whether exposure to favorable or unfavorable personalized advertising can impact people's perceptions of their social competence, health, intelligence, or taste level. Another avenue could involve testing the effects of OBA on self-concept when it is viewed in a stream of content (similar to one's Instagram or Facebook feed) rather than in isolation, which would enhance the ecological validity of these findings. One might also test whether other forms of unfavorable or undesirable algorithmically-generated feedback, such as personalized recommendations of content to consume or products to buy, have a similar effect on self-concept. I suspect that if a wider range of stimuli were tested, then I would be able to more successfully determine the effect of OBA on self-concept.

Conclusion

Ultimately, this study found some initial support for the claim that people can perceive OBA as a source of feedback, even when it does not reference a specific self-presentation (as is usually the case when people encounter online advertising in their natural environments). Rather than instruct participants to engage in a specific act of self-presentation, I asked participants in this study to supply all of their Instagram data, which constitutes a nebulous record of their self-presentation over time, before viewing personalized or non-personalized advertising. In conjunction with a 2018 study by French which used a similar procedure, there is growing evidence that the algorithmically-personalized messages to which we are frequently exposed online, but are rarely able to explain with complete certainty, have the potential to serve as feedback on our self-presentation and shape how we see ourselves. This insight contributes to previous work examining how perceptions of an audience for one's behavior can affect self-concept (Gonzales & Hancock, 2008; Kelly & Rodriguez, 2006; Schlenker et al., 1994; Tice, 1992) and how feedback amplifies this effect (Carr & Hayes, 2019; Walther et al., 2011). This study also attempted to extend previous research on the effect of OBA (e.g. Summers et al., 2016) and other types of algorithmically-personalized messages (French, 2018) on self-concept by examining perceptions of publicness as a potential mechanism and exploring who or what constitutes the perceived audience experienced with OBA. Although this study fell short of clearly identifying the mechanism by which OBA affects self-concept, it contributed to this body of literature by testing perceived publicness as a mediator of self-concept change following exposure to algorithmic feedback.

Furthermore, this study lends some support to the claim that unfavorable personalized ads can influence self-concept just as favorable or pleasant personalized ads can. This suggests that self-enhancement processes do not necessarily protect the ego from negative feedback generated by an algorithm, potentially because algorithms are often viewed as powerful and all-knowing (Eslami et al., 2018), and algorithmic feedback is instead taken at face-value (Mishra, 2006). In other words, even when people do not feel that an allegedly personalized advertisement is well-suited to them, they may still experience self-concept change as a result of this algorithmic feedback. Interestingly, this study did not find that those with low self-esteem were more likely than those with high self-esteem to experience self-concept change following negative feedback, despite what previous studies have found when testing the effects of negative interpersonal feedback (e.g. Kim & Gonzales, 2018; Swann, 1983). This null finding might suggest that it is more difficult for people with high self-esteem to form self-enhancing explanations for negative feedback when it comes from an algorithm rather than a human, although future research is needed to test this proposition. It is crucial to understand the impact of negative algorithmic feedback given the possibility that if one continues to encounter unfavorable messages which temporarily activate a negative self-concept, that self-concept could become chronically accessible when processing information in the future.

OBA is a ubiquitous feature of many online platforms which can at times be unpleasant for people, and it is unlikely to disappear from the digital landscape anytime soon. Thus, having a better understanding of the disadvantages of OBA can help online platforms and their advertisers greatly improve user experience and advertising effectiveness. Moreover, the digital advertising industry as a whole should enable and encourage

consumers to exercise more control over the types of advertisements they see across the web. In the meantime, by understanding the shortcomings of ad personalization algorithms and engaging with tools to manage their online advertising preferences, people might be able to proactively mitigate some of the negative effects of OBA and improve their experiences online.

References

Agarwal, L., Shrivastava, N., Jaiswal, S., & Panjwani, S. (2013). Do not embarrass: Re-examining user concerns for online tracking and advertising. In *Proceedings of the Ninth Symposium on Usable Privacy and Security* (pp. 1-13). https://doi.org/10.1145/2501604.2501612

Alicke, M. D. & Sedikides, C. (2009). Self-enhancement and self-protection: What they are and what they do. *European Review of Social Psychology, 20*(1), 1-48. https://doi.org/10.1080/10463280802613866

Ananny, M. & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society, 20*(3), 973–989. http://dx.doi.org/10.1177/1461444816676645

Andreou, A., Venkatadri, G., Goga, O., Gummadi, K., Loiseau, P., & Mislove, A. (2018). Investigating ad transparency mechanisms in social media: A case study of Facebook's explanations. In *Proceedings from Network and Distributed System Security Symposium (NDSS 2018)* (pp. 1-15). http://dx.doi.org/10.14722/ndss.2018.23204

Auxier, B. (2020). Most Americans support right to have some personal info removed from online searches. *Pew Research Center*. Retrieved from https://www.pewresearch.org/fact-tank/2020/01/27/most-americans-support-right-to-have-some-personal-info-removed-from-online-searches/

Auxier, B., Rainie, L., Anderson, M., Perrin, A., Kumar, M., & Turner, E. (2019). Americans and privacy: Concerned, confused and feeling lack of control over their personal information. *Pew Research Center*. Retrieved from

https://www.pewresearch.org/internet/2019/11/15/how-americans-think-about-privacy-and-the-vulnerability-of-their-personal-data/

Bargh, J. A., Bond, R. N., Lombardi, W. J., & Tota, M. E. (1986). The additive nature of chronic and temporary sources of construct accessibility. *Journal of Personality and Social Psychology, 50*, 869-878.

Baumeister, R. F. (1982). Self-esteem, self-presentation, and future interaction: A dilemma of reputation. *Journal of Personality*, *50*(1), 29-45. https://doi.org/10.1111/j.1467-6494.1982.tb00743.x

Boerman, S. C., Kruikemeier, S., & Borgesius, F. J. Z. (2018). Exploring motivations for online privacy protection behavior: Insights from panel data. *Communication Research*. http://dx.doi.org/10.1177/0093650218800915

Cadinu, M., Maass, A., Lombardo, M., & Frigerio, S. (2006). Stereotype threat: The moderating role of locus of control beliefs. *European Journal of Social Psychology, 36*(2), 183-197. http://dx.doi.org/10.1002/ejsp.303

Carr, C. T., & Foreman, A. C. (2016). Identity shift III: Effects of publicness of feedback and relational closeness in computer-mediated communication. *Media Psychology*, *19*(2), 334-358. https://doi.org/10.1080/15213269.2015.1049276

Carr, C. T. & Hayes, R. A. (2019). Identity shift effects of self-presentation and confirmatory and disconfirmatory feedback on self-perceptions of brand identification. *Media Psychology, 22*(3), 418-444. http://dx.doi.org/10.1080/15213269.2017.1396228

Carver, C. S., & Scheier, M. F. (1981). The self-attention-induced feedback loop and social facilitation. *Journal of Experimental Social Psychology*, *17*(6), 545-568.

Cooley, C. H. (1964). *Human nature and the social order*. Schocken Books. (Original work

published 1902)

Crary, W. G. (1966). Reactions to incongruent self-experiences. *Journal of Consulting Psychology, 30*(3), 246–252. https://doi.org/10.1037/h0023347

Deshpandé, R., & Stayman, D. M. (1994). A tale of two cities: Distinctiveness theory and advertising effectiveness. *Journal of Marketing Research*, *31*(1), 57-64.

Ditto, P. H., & Boardman, A. F. (1995). Perceived accuracy of favorable and unfavorable psychological feedback. *Basic and Applied Social Psychology, 16*(1-2), 137-157. https://doi.org/10.1080/01973533.1995.9646106

Ejelöv, E., & Luke, T. J. (2020). "Rarely safe to assume": Evaluating the use and interpretation of manipulation checks in experimental social psychology. *Journal of Experimental Social Psychology*, *87*, 103937. https://doi.org/10.1016/j.jesp.2019.103937

Eslami, M., Kumaran, S. R. K., Sandvig, C., & Karahalios, K. (2018). Communicating algorithmic process in online behavioral advertising. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. http://dx.doi.org/10.1145/3173574.3174006

Facebook Business Help Center. (2020). Prohibited content for personal health products or services. Facebook for business. https://www.facebook.com/business/help/388992385313478?id=689788528177451

Fedor, D. B., Davis, W. D., Maslyn, J. M., & Mathieson, K. (2001). Performance improvement efforts in response to negative feedback: The roles of source power and recipient self-esteem. *Journal of Management, 27*(1), 79–97. https://doi.org/10.1177/014920630102700105

French, M. R. (2018*). Algorithmic mirrors: An examination of how personalized recommendations can shape self-perceptions and reinforce gender stereotypes* (Unpublished doctoral dissertation). Stanford University, Palo Alto, CA.

French, M., & Hancock, J. T. (2016, June). *Addressing the audience: Effects of evaluation on identity shift in computer-mediated environments*. Paper presented at the 2016 meeting of the International Communication Association, Fukuoka, Japan.

Frijda, N. H. (1987). Emotion, cognitive structure, and action tendency. *Cognition and Emotion*, *1*(2), 115-143.

Goffman, E. (1959). *The presentation of self in everyday life*. Anchor Books.

Gonzales, A. L., & Hancock, J. T. (2008). Identity shift in computer-mediated environments. *Media Psychology, 11*(2), 167-185. http://dx.doi.org/10.1080/15213260802023433

Gonzales, A. L., & Hancock, J. T. (2011). Mirror, mirror on my Facebook wall: Effects of exposure to Facebook on self-esteem. *Cyberpsychology, Behavior, and Social Networking, 14*(1-2), 79-83. http://dx.doi.org/10.1089/cyber.2009.0411

Gonzales, A. L., Kwon, E.Y., Lynch, T., & Fritz, N. (2018). "Better everyone should know our business than we lose our house": Costs and benefits of medical crowdfunding for support, privacy, and identity. *New Media & Society*, 20(2), 641-658. http://dx.doi.org/10.1177/1461444816667723

Harter, S. (2006). The self. In W. Damon & R. M. Lerner (Series Eds.) & N. Eisenberg (Vol. Ed.), *Handbook of Child Psychology: Vol. 3. Social, Emotional, and Personality Development* (6th ed., pp. 505-570). Wiley.

Hautea, S., Munasinghe, A., & Rader, E. (2020). 'That's not me': Surprising algorithmic inferences. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in*

Computing Systems (pp. 1-7). http://dx.doi.org/10.1145/3334480.3382816

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. The Guilford Press.

Higgins, E. T. (1987). Self-discrepancy: A theory relating self and affect. *Psychological Review*, *94*(3), 319-340.

Higgins, E. T., & King, G. (1981). Accessibility of social constructs: Information-processing consequences of individual and contextual variability. In N. Cantor & J. E Kihlstrom (Eds.), *Personality, cognition, and social interaction* (pp. 69-121). Erlbaum.

Holbrook, M. B., & Batra, R. (1987). Assessing the role of emotions as mediators of consumer responses to advertising. *Journal of consumer research*, *14*(3), 404-420. https://doi.org/10.1086/209123

Johnson, B. K., & Van Der Heide, B. (2015). Can sharing affect liking? Online taste performances, feedback, and subsequent media preferences. *Computers in Human Behavior*, *46*, 181-190. http://dx.doi.org/10.1016/j.chb.2015.01.018

Joinson, A., & Banyard, P. (2002). Psychological aspects of information seeking on the Internet. In *Aslib Proceedings*. MCB UP Ltd. https://doi.org/10.1108/00012530210435220

Kanoy III, R. C., Johnson, B. W., & Kanoy, K. W. (1980). Locus of control and self-concept in achieving and underachieving bright elementary people. *Psychology in the Schools*, *17*(3), 395-399. https://doi.org/10.1002/1520-6807

Kelly, A. E. & Rodriguez, R. R. (2006). Publicly committing oneself to an identity. *Basic and Applied Social Psychology, 28*(2), 185-191. https://doi.org/10.1207/s15324834basp2802_8

Kim, Y. W. & Gonzales, A. L. (2018). You don't know me: Negative self-views interact with publicness and feedback to shape interpersonal impressions online. *Computers in Human Behavior, 88*, 195-204. http://dx.doi.org/10.1016/j.chb.2018.06.025

Kunda, Z. (1987). Motivated inference: Self-serving generation and evaluation of causal theories. *Journal of Personality and Social Psychology, 53*(4), 636-647. https://doi.org/10.1037/0022-3514.53.4.636

Leary, M. R., & Baumeister, R. F. (2000). The nature and function of self-esteem: Sociometer theory. In M. P. Zanna (Eds.), *Advances in experimental social psychology* (Vol. 32, pp. 1-62). Academic Press.

Lewis, M. (1995). Self-conscious emotions: Embarrassment, pride, shame, and guilt. *American Scientist*, *83*(1), 68-78.

Li, C. (2016). When does web-based personalization really work? The distinction between actual personalization and perceived personalization. *Computers in Human Behavior*, *54*, 25-33. https://doi.org/10.1016/j.chb.2015.07.049

Markus, H., & Kunda, Z. (1986). Stability and malleability of the self-concept. *Journal Of Personality and Social Psychology*, *51*(4), 858-866. http://dx.doi.org/10.1037//0022-3514.51.4.858

Markus, H., & Nurius, P. (1986). Possible selves. *American Psychologist*, *41*(9), 954-969.

Markus, H., & Wurf, E. (1987). The dynamic self-concept: A social psychological perspective. *Annual Review of Psychology*, *38*(1), 299-337.

McAuley, E., Duncan, T. E., & Russell, D. W. (1992). Measuring causal attributions: The revised causal dimension scale (CDSII). *Personality and Social Psychology Bulletin, 18*(5), 566–573. http://dx.doi.org/10.1177/0146167292185006

Mead, G. H. (1934). *Mind, self, and society*. University of Chicago Press.

Mishra, P. (2006). Affective feedback from computers and its effect on perceived ability and affect: A test of the computers as social actor hypothesis. *Journal of Educational Multimedia and Hypermedia*, *15*(1), 107-131.

Neemann, J., & Harter, S. (2012). *Self-perception profile for college people: Manual and questionnaires.* University of Denver Press. Retrieved from: https:// portfolio.du.edu/downloadItem/223543.

Norberg, P. A., Horne, D. R., & Horne, D. A. (2007). The privacy paradox: Personal information disclosure intentions versus behaviors. *Journal of Consumer Affairs*, *41*(1), 100-126.

Oeldorf-Hirsch, A., Birnholtz, J., & Hancock, J. T. (2017). Your post is embarrassing me: Face threats, identity, and the audience on Facebook. *Computers in Human Behavior, 73*, 92-99. http://dx.doi.org/10.1016/j.chb.2017.03.030

Park, L. E. & Crocker, J. (2008). Contingencies of self-worth and responses to negative interpersonal feedback. *Self and Identity, 7*, 284-203. http://dx.doi.org/10.1080/15298860701398808

Phelan, C., Lampe, C., & Resnick, P. (2016). It's creepy, but it doesn't bother me. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems,* 5240-5251. http://dx.doi.org/10.1145/2858036.2858381

Putoni, S., de Hooge, I. E., & Verbeke, W. J. (2015). Advertising-induced embarrassment. *Journal of Advertising*, *44*(1), 71-79. https://doi.org/10.1080/00913367.2014.935899

Rainie, L. (2019). Facebook Algorithms and Personal Data.

https://www.pewinternet.org/2019/01/16/facebook-algorithms-and-personal-data/

Rao, A., Schaub, F., & Sadeh, N. (2015). What do they know about me? Contents and concerns of online behavioral profiles. Retrieved from https://arxiv.org/pdf/1506.01675.pdf

Reed, A. (2004). Activating the self-importance of consumer selves: Exploring identity salience effects on judgments. *Journal of Consumer Research, 31*, 286-295.

Rosenberg, M. (1965). *Society and the adolescent self-image.* Princeton University Press.

Scheier, M. F., & Carver, C. S. (1980). Private and public self-attention, resistance to change, and dissonance reduction. *Journal of Personality and Social Psychology*, *39*(3), 390-405. https://doi.org/10.1037/0022-3514.39.3.390

Schlenker, B. R., Dlugolecki, D. W., & Doherty, K. (1994). The impact of self-presentations on self-appraisals and behavior: The power of public commitment. *Personality and Social Psychology Bulletin, 20*(1), 20-33. https://doi.org/10.1177/0146167294201002

Sedikides, C. (1993). Assessment, enhancement, and verification determinants of the self-evaluation process. *Journal of Personality and Social Psychology*, *65*(2), 317.

Shrauger, J. S., & Lund, A. K. (1975). Self-evaluation and reactions to evaluations from others. *Journal of Personality, 43*(1), 94-108. https://doi.org/10.1111/j.1467-6494.1975.tb00574.x

Shrauger, J. S., & Rosenberg, S. E. (1970). Self-esteem and the effects of success and failure feedback on performance. *Journal of Personality*, *38*(3), 404-417. https://doi.org/10.1111/j.1467-6494.1970.tb00018.x

Snyder, M., Tanke, E. D., & Berscheid, E. (1977). Social perception and interpersonal

behavior: On the self fulfilling nature of social stereotypes. Journal of Personality and Social Psychology, 35, 656–666.

Snyder, M. & Swann, W. B. (1978). Behavioral confirmation in social interaction: From social perception to social reality. Journal of Experimental Social Psychology, 14, 148–162.

Summers, C. A., Smith, R. W., & Reczek, R. W. (2016). An audience of one: Behaviorally targeted ads as implied social labels. *Journal of Consumer Research, 43*(2016), 156-178. https:/10.1093/jcr/ucw012

Swann, W. B. (1983). Self-verification: Bringing social reality into harmony with the self. In J. Suls & A. G. Greenwald (Eds.), *Social Psychological Perspectives on the Self* (Vol. 2, pp. 33-66). Erlbaum.

Swann, W. B., Pelham, B. W., & Krull, D. S. (1989). Agreeable fancy or disagreeable truth? Reconciling self-enhancement and self-verification. *Journal of Personality and Social Psychology*, *57*(5), 782. https://doi.org/10.1037//0022-3514.57.5.782

Swann, W. B., Wenzlaff, R. M., Krull, D. S., & Pelham, B. W. (1992). Allure of negative feedback: Self-verification strivings among depressed persons. *Journal of Abnormal Psychology, 101*(2), 293–306.

Tabachnick, B. G., & Fidell, L. S. (2007). *Experimental designs using ANOVA*. Thomson/Brooks/Cole.

Thomas, T. C., Trump, R. K., & Price, L. L. (2015). Advertising as unfavorable self-presentation: The dirty laundry effect. *Journal of Advertising, 44*(1), 58-70. http://dx.doi.org/10.1080/00913367.2014.935898

Tice, D. M. (1992). Self-concept change and self-presentation: The looking glass self is also

a magnifying glass. *Journal of Personality and Social Psychology, 63*(3), 435-451. http://dx.doi.org/10.1037//0022-3514.63.3.435

Toma, C. L., Hancock, J. T., & Ellison, N. B. (2008). Separating fact from fiction: An examination of deceptive self-presentation in online dating profiles. *Personality and Social Psychology Bulletin, 34*(8), 1023-1036. http://dx.doi.org/10.1177/0146167208318067

Ur, B., Leon, P. G., Cranor, L. F., Shay, R., & Wang, Y. (2012). Smart, useful, scary, creepy: Perceptions of online behavioral advertising. In *Proceedings of the Eighth Symposium on Usable Privacy and Security (SOUPS)* (pp. 1-15).

Walther, J. B. (1996). Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction. *Communication Research, 23*, 3–43. https://doi.org/10.1177/009365096023001001

Walther, J. B., Liang, Y. J., DeAndrea, D. C., Tong, S. T., Carr, C. T., Spottswood, E. L., & Amichai-Hamburger, Y. (2011). The effect of feedback on identity shift in computer-mediated communication, *Media Psychology, 14*(1), 1-26. http://dx.doi.org/10.1080/15213269.2010.547832

Weber, R., & Popova, L. (2012). Testing equivalence in communication research: Theory and application. *Communication Methods and Measures*, *6*(3), 190-213. https://doi.org/10.1080/19312458.2012.703834

Wiesenfeld, B. M., Swann, W. B., Jr. Brockner, J., & Bartel, C. A. (2007). Is more fairness always preferred? Self-esteem moderates reactions to procedural justice. *Academy of Management Journal, 50*(5), 1235–1253.

Yao, Y., Re, D. L., & Wang, Y. (2017). Folk models of online behavioral advertising. In

Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17), 1957-1969. http://dx.doi.org/10.1145/2998181.2998316

Zhu, Y. Q., & Chang, J. H. (2016). The key role of relevance in personalized advertisement: Examining its impact on perceptions of privacy invasion, self-awareness, and continuous use intentions. *Computers in Human Behavior*, *65*, 442-447. http://dx.doi.org/10.1016/j.chb.2016.08.048

Table 1.

Pilot Test Ad Believability & Relevance Ratings

	Believability		Relevance	
	M	*SD*	*M*	*SD*
Filler advertisements				
Postmates food delivery ad (*N*=102)	5.70	1.02	5.30	1.53
Chipotle restaurant ad (*N*=102)	5.45	1.32	5.44	1.49
Everlane clothing ad (*N*=102)	5.72	1.30	4.95	1.80
Nike sneakers ad (*N*=102)	5.22	1.14	4.74	1.45
Experimental manipulation advertisements				
ClassPass fitness ad (*N*=62)	5.18	1.52	4.46	1.68
Curology skincare ad (*N*=62)	5.52	1.42	4.60	1.76
Stop COVID-19 PSA (*N*=62)	4.49	1.46	3.79	1.64
Protein powder ad (*N*=62)	4.49	1.50	3.67	1.80
Mentorship/tutoring program ad (*N*=102)	3.71	1.38	3.57	1.57

Note. Sample sizes differ because advertisements were tested during two separate rounds of pilot testing using the same procedure.

Table 2.

Descriptive Statistics by Condition

	Favorable Caption				Unfavorable Caption			
	Not personalized (*N*=26)		Personalized (*N*=21)		Not personalized (*N*=25)		Personalized (*N*=23)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Trait self-esteem	2.82	0.51	2.93	0.46	2.88*	0.38	2.59*	0.46
Perceived personalization	3.33*	1.18	4.38*	1.60	4.08	1.23	3.74	1.31
Ad relevance	3.93**	1.34	5.05**	1.47	4.32	1.46	4.34	1.42
Ad favorability	4.48	0.74	5.02	1.23	4.25	1.20	4.40	1.18
Negative emotions	2.06	1.22	2.06	1.23	2.72	1.56	2.62	1.54
Perceived publicness	3.55	1.08	4.10	1.56	3.85	1.26	4.09	1.40
The ad personal-ization system	3.62*	1.53	4.57*	1.75	3.88	1.56	4.17	1.88
Insta employees	3.42	1.42	3.71	1.68	3.68	1.49	3.83	1.72
Advertisers	4.08	1.41	4.15	1.57	4.08	1.44	4.35	1.50
The researchers	3.65	1.23	4.29	1.74	3.96	1.37	4.39	1.44
Your Instagram followers	3.00	1.55	3.95	2.20	3.64	1.35	3.70	1.77
Appearance self-concept	2.30	0.91	2.37	0.63	2.62**	0.72	2.01**	0.72
Causal attributions	4.31*	1.69	5.27*	1.38	4.69	1.55	4.77	1.91
Privacy concern	4.62	0.99	4.68	1.05	4.31	1.05	4.75	1.15

Note. *p < .05, **p < .01; Trait self-esteem and appearance self-concept were measured on a four-point scale; All other items were measured on a seven-point scale. Higher causal attribution values = more internal (vs. external).

Table 3.

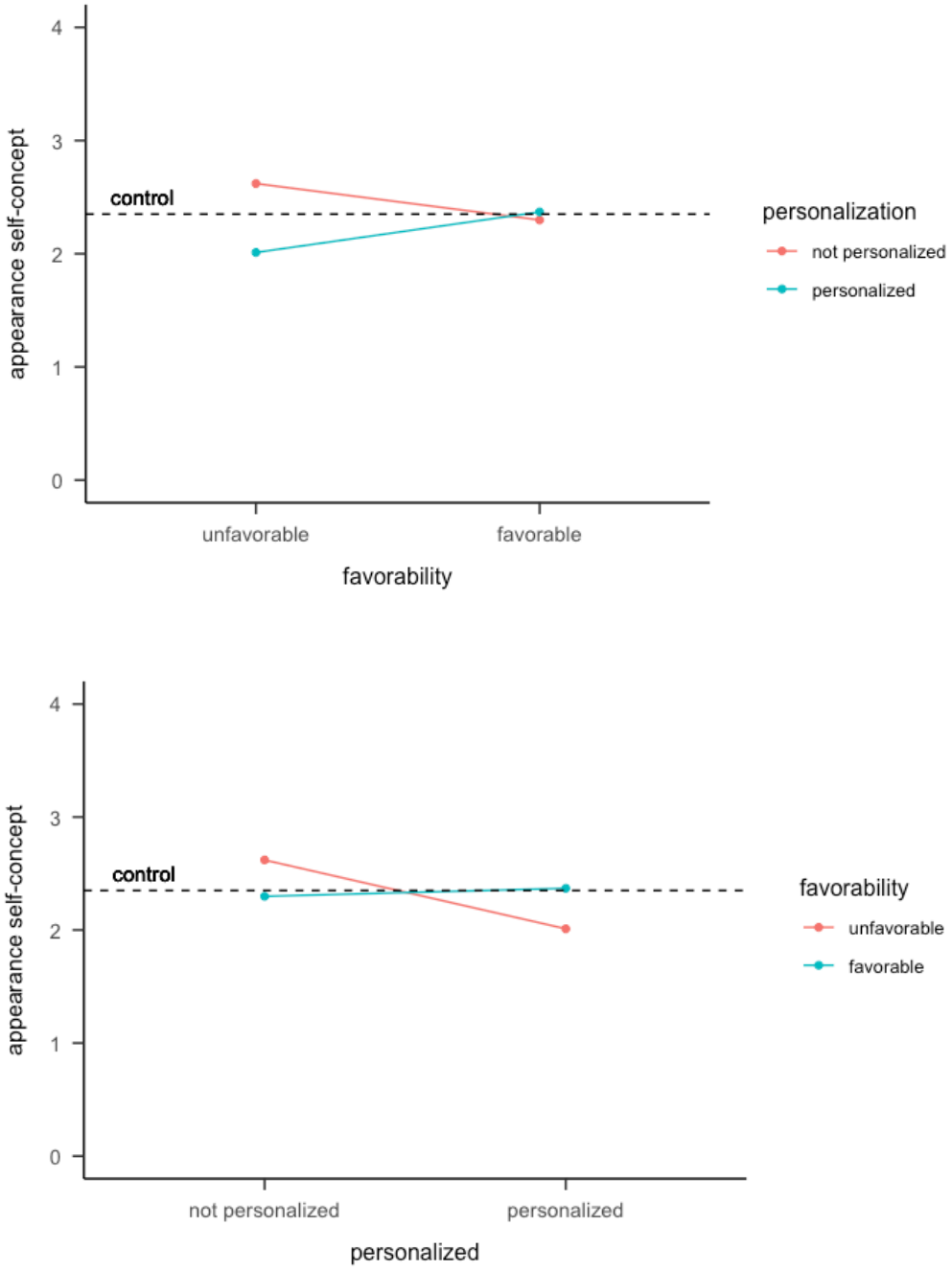
Bivariate Correlations and Descriptive Statistics for Variables

	M	***SD***	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1. Trait self-esteem	2.85	.46								
2. Perceived personalization	3.86	1.36	.12							
3. Ad relevance	4.38	1.45	.19	.76**						
4. Ad favorability	4.52	1.11	.26*	.47**	.57**					
5. Negative emotion	2.37	1.41	-.32**	.02	.03	-.13				
6. Perceived publicness	3.88	1.32	.04	.59**	.59**	.41**	.07			
7. Appearance self-concept	2.33	.78	.46**	-.01	-.03	-.09	-. 28**	.00		
8. Causal attributions	4.73	1.66	.21*	.40**	.55**	.34**	.07	.49**	.00	
9. Privacy concern	4.58	1.06	-.07	.00	.06	.00	.10	.09	.10	-.05

Note. *p < .05, **p < .01; Trait self-esteem and appearance self-concept were measured on a four-point scale; All other items were measured on a seven-point scale. Higher causal attribution values = more internal (vs. external).

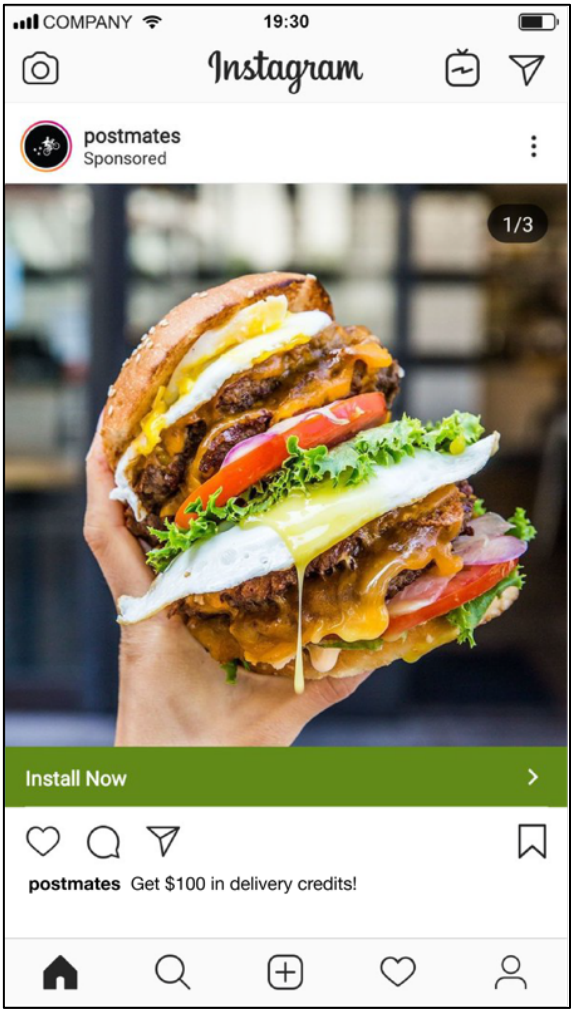
Figure 1.

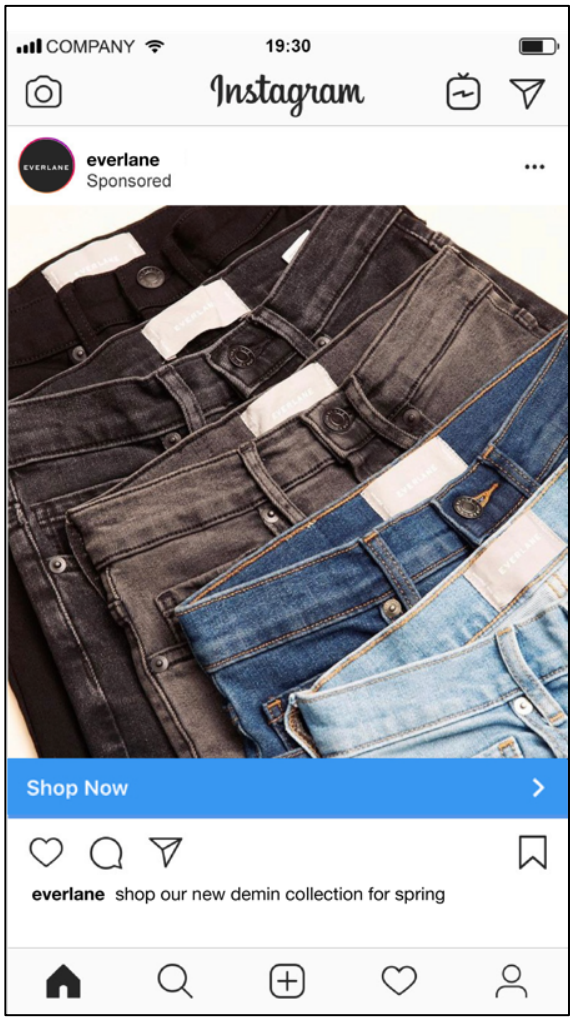
Interaction Effect of Ad Personalization and Favorability on Appearance Self-Concept



Appendix A

Filler Advertisements

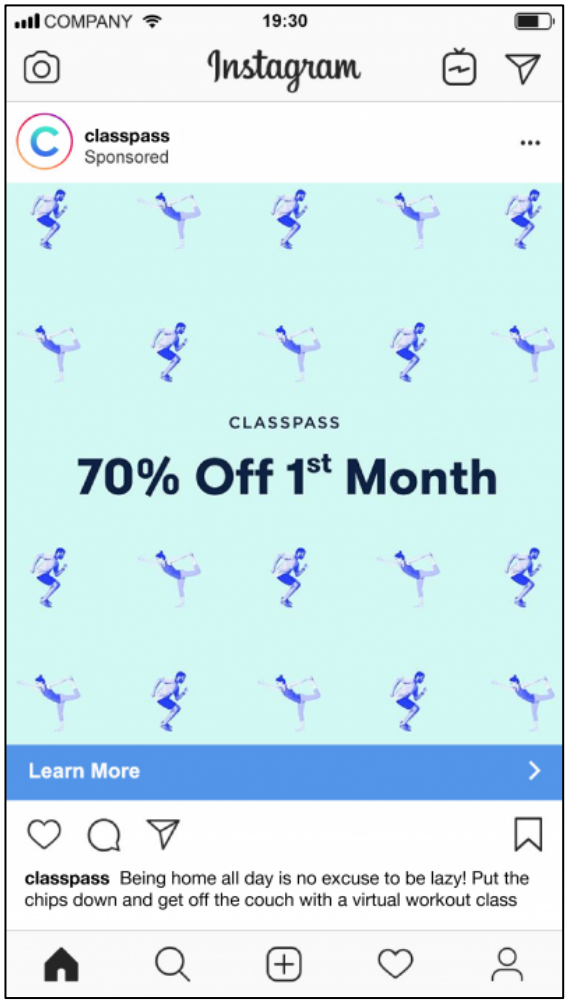




Note: Each participant saw both advertisements in the same order.

Manipulated Advertisement

Favorable version



Unfavorable version

Note: Each participant saw one version of this advertisement.

Appendix B

Downloading a copy of your data on Instagram

If you want a copy of everything you've shared on Instagram, you can request a download of your data in a machine readable (JSON) format. You'll need your Instagram account password to request this information. Learn more if you've forgotten your password or can't log in.

From Instagram on the Web:

1. Go to your profile and click .
2. Click **Privacy and Security**.
3. Scroll down to **Data Download** and click **Request Download**.
4. Enter the email address where you'd like to receive a link to your data and enter your Instagram account password.
5. You'll soon receive an email titled **Your Instagram Data** with a link to your data. Click **Download Data** and follow the instructions to finish downloading your information.

From iOS or Android:

1. Go to your profile and tap .
2. Tap **Settings**.
3. Tap **Security** > **Download Data**.
4. Enter the email address where you'd like to receive a link to your data and tap **Request Download**.
5. Enter your Instagram account password.
6. You'll soon receive an email titled **Your Instagram Data** with a link to your data. Click **Download Data** and follow the instructions to finish downloading your information.

Note: It may take up to 48 hours for us to email you a download link. Some data you have deleted may be stored temporarily for safety and security purposes, but will not appear when you access or download your data.

If you can't access your Instagram Account and still want to download a copy of your data, you can contact us.

Appendix C

Rosenberg Self-Esteem Scale.

Instructions: Below is a list of statements dealing with your general feelings about yourself. Please indicate how strongly you agree or disagree with each statement.

1. On the whole, I am satisfied with myself.
2. At times I think I am no good at all.
3. I feel that I have a number of good qualities.
4. I am able to do things as well as most other people.
5. I feel I do not have much to be proud of.
6. I certainly feel useless at times.
7. I feel that I'm a person of worth, at least on an equal plane with others.
8. I wish I could have more respect for myself.
9. All in all, I am inclined to feel that I am a failure.
10. I take a positive attitude toward myself.

Note. Participants rated each statement on a four-point scale (1 = *strongly disagree*, 2 = *disagree,* 3 = *agree,* and 4 = *strongly agree*); Items 2, 5, 6, 8, 9 are reverse scored.

Appendix D

Self-Perception Profile for College Students (Neemann & Harter, 2012).

Instructions: The following are statements that allow people to describe themselves. There are no right or wrong answers since people differ markedly. Please read the entire sentence across. First decide which one of the two parts of each statement best describes you; then go to that side of the statement and check whether that is just sort of true for you or really true for you. You will just check ONE of the four boxes for each statement.

Appearance Subscale

Really true for me	Sort of true for me				Sort of true for me	Really true for me
(1)	(2)	Some people are not happy with the way they look	BUT	Other people are happy with the way they look	(3)	(4)
(4)	(3)	Some people are happy with their height and weight	BUT	Other people wish their height or weight was different	(2)	(1)
(1)	(2)	Some people wish their body was different	BUT	Other people like their body the way it is	(3)	(4)
(4)	(3)	Some people like their physical appearance the way it is	BUT	Other people do not like their physical appearance	(2)	(1)

Note. The second and fourth statements are reverse scored.