Performative or authentic?

How affordances signal (in)authentic digital allyship

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Kyla Noni Elizabeth Brathwaite, MA

Graduate Program in Communication

The Ohio State University

2024

Dissertation Committee

Dr. David DeAndrea, Advisor

Dr. Joseph Bayer

Dr. Teresa Lynch

Abstract

Millions of people use social media platforms to curate digital representations of themselves through the creation of profiles and sharing images and narratives about their daily lives. Social media are also increasingly used to amplify offline prosocial causes that grow and evolve into prosocial movements. Although social media platforms have been used to bring people together in solidarity and push for change in society through causes such as #MeToo, #BoPo, and #BLM (Jackson et al., 2020), some scholars contend that social media increasingly provide venues for users to engage in performative allyship (Wellman, 2022) that prioritizes the personal interests of users rather than the social causes they purport to advance. The vast number of sources—with different agendas and messaging strategies—who exist on social media platforms can create complexities related to promoting prosocial causes authentically and effectively. The current study examines how differences in two perceived affordances—persistence and visibility—affect how women viewers evaluate the authenticity of men who post allyship messages on Instagram. Additionally, I examine how viewer perceptions of comment deletion and comment restriction impact source evaluations. Results indicate that the more women viewers found posts to be persistent, the more they rated the men who posted as authentic allies, as having prosocial reasons for posting, and as having less self-interested reasons for posting. Greater perceived visibility of posts had a similar impact on these outcome variables. Moreover, comment deletion and restriction were found to partially affect how viewers evaluated the authenticity and motives of sources, providing additional support to warranting theory (DeAndrea, 2014; Walther & Parks, 2002). Implications for promoting prosocial messaging and digital allyship online are discussed.

Dedication

This work is dedicated to the memory of Jacqueline Y. Jackson (“Dah”), my grandmother who is forever with me.

Acknowledgements

I would like to first thank my incredible advisor, Dr. David DeAndrea for his dedication and guidance throughout my time in the School of Communication. His support, advice, and encouragement have never wavered. In the next phase of my career as a communication scholar and (eventually) a faculty member, I hope to exemplify his dedication to graduate students and junior scholars within the field.

I would like to also acknowledge my committee members Drs. Joseph Bayer and Teresa Lynch. I am grateful for the time you both have taken to help guide me through my doctoral program. From my comprehensive exams, to now – I have been able to apply your feedback and have become a better scholar in the process.

To Dr. Megan Vendemia ("Dr. V") – thank you for taking me, a first year at the time, under your wing and becoming not only a mentor and collaborator but also a friend.

Thank you to my friends from the School of Communication, I am thankful for our time together in Columbus.

I would also like to acknowledge my closest friend, Elizabeth Gonzalez. From the first day of college at Cornell, to now, as we both finish up our doctoral programs – I am incredibly thankful for you.

I would like to acknowledge my family for their everlasting support. Thank you to my aunts, uncles, and cousins for cheering me on throughout the years and encouraging me continuously. Thank you to my grandmother, Brenda. To my partner, Omar, thank you for your continuous love, encouragement, perspective, and motivation. And finally, to my parents, Claude and Jamila – THANK YOU. I would not be here without you. You have raised me to see the value of education, learning, and exploration. You both have supported me through all my academic and life endeavors, being my greatest and wisest cheerleaders and forever best friends. We did it.

Vita

2014……Hackley School

2018……B.S Communication, Cornell University

2020……M.A Communication, The Ohio State University

2020 to Present……...Graduate associate, School of Communication, The Ohio State University

Publications

Brathwaite, K.N., DeAndrea, D.C., & Vendemia, M.A. (2024). Non-sexualized images and body-neutral messaging foster body positivity online. *Social Media + Society.* 9(4). https://doi.org/10.1177/20563051231207852

Brathwaite, K.N., & DeAndrea, D.C. (2022). BoPopriation: how self-promotion and corporate commodification can undermine the Body Positivity (BoPo) movement on Instagram. *Communication Monographs.* 89(1), 25-46. https://doi.org/10.1080/03637751.2021.1925939

Vendemia, M. A., Brathwaite, K. N., & DeAndrea, D. C. (2022). An intersectional approach to evaluating the effectiveness of women's sexualized body-positive imagery on Instagram. *New Media and Society. 0*(0). https://doi.org/10.1177/14614448221143345

Vendemia, M.A., DeAndrea, D.C., & Brathwaite, K.N. (2021). Objectifying the body positive movement: the effects of sexualizing and digitally modifying body-positive images on Instagram. Body Image. 38, 137-147. https://doi.org/10.1016/j.bodyim.2021.03.017

Fields of Study

Major Field: Communication

Area of Emphasis: Communication Technology

Table of Contents

List of Tables

List of Figures

Introduction

Social media platforms are powerful tools that allow users incredible reach through the creation and reception of original user-generated content. Popular social media platforms such as Instagram, TikTok, Twitter/X, and Snapchat are centered around self-expression facilitated through personal profiles, selfies, and other digitized personal narratives about people's daily lives (e.g., accomplishments, personalities, relationships, etc.,) or personal interests (e.g., hobbies, politics, pets). While these platforms allow for diversity in self-expression, they are also used as tools for impression management; properties of these platforms allow users great control over what other users are able to see and contribute towards on their profiles (Walther & Jang, 2012). As a result of this strategic control, users are able to curate their online identities and showcase themselves to others in the most desirable fashion (Chua & Chang, 2016; Ellison et al., 2006; Verduyn et al., 2020; Walther, 1996).

As the use of social media platforms continues to expand, the utility of these platforms has diversified as well. The decreased barriers of access that social media platforms such as Twitter/X, Instagram, and TikTok provide have afforded users a low stakes environment where they can make observations about power imbalances in the world around them and further challenge them. To that end, groups of African Americans, women, members of the LGBTQ community, and other individuals belonging to disadvantaged or marginalized identities have begun to make "identity-based cultural and political demands" pushing issues concerning their

communities into the digital public sphere (Jackson et al., 2020, pp. xxv). While people may use social media to identify and communicate about injustices on an individual level, characteristics of these platforms can help to facilitate "meaning making, resource mobilization, and coalition building" (Mundt et al., 2019, pp. 3) scaling up movements. To that point, hashtags such as #BlackLivesMatter (BLM), #BodyPositivity, #MeToo, and #WomensMarch, have become strategic in their use on sites such as Twitter/X or Instagram to associate with a variety of content related to prosocial movements advocating for social change (Jackson et al., 2020).

As a whole, social media platforms allow users to create and exchange user-generated content; such exchanges can both positively and negatively exert social influence due to the mass number of sources, messages, and agendas existing online (Walther & Jang, 2012). To that end, the spread of user-generated prosocial content may be beneficial for eventual collective action efforts. Computer-mediated communication can facilitate eventual collective action efforts through features of communication technology (Hampton & Wellman, 2003). For example, other social media users, or movement allies, can create, share or reproduce others' content using movement hashtags and features native to platforms (Wellman, 2022). However, one caveat of the increased user control of prosocial messaging is the potential for the messages themselves or features of communication technology to be exploited or misappropriated by other social media users who claim to support these issues.

The effects of inauthentic adoption of prosocial messaging is exemplified in the events following the murder of George Floyd and massive nationwide protests against police brutality in 2020, allies of Black Lives Matter (BLM) began to use #BLM as well as sharing black squares and other BLM related imagery on both Instagram and Twitter/X to show support for the black

community (Wellman, 2022). However, the use of movement hashtags and other movement imagery by allies may have been more harmful than helpful, causing information about crucial resources and protests to be suppressed and calling into question social media users' ingenuine motivation for posting BLM content (Wellman, 2022).

The circumstances of the events of the 2020 Black Lives Matter movement speak to a larger phenomenon concerning performative allyship, or the adoption of prosocial messaging and imagery by social media users as an impression management tactic to gain favor with their online audiences, rather than to advance the goals of the prosocial issues in question (Kutlaca & Radke, 2020; Wellman, 2022). With the increase in popularity of social media activism, it begs the question of whether assumed prosocial social media activities (e.g., creating or resharing prosocial messaging, or utilizing hashtags related to social movements) can effectively promote prosocial issues of interest, especially if they are believed to be ingenuine by those the activities are purported to help.

Given the user-generated nature of social media platforms and the use of these social media platforms to promote prosocial issues and movements, this dissertation will examine the paradoxical relationship between sharing prosocial content on social media and the performative and self-promotional nature of many social media platforms. Social media platforms provide low-cost settings for social advocacy and provide users with ample opportunity to utilize its' tools for beneficial self-representation purposes. Source characteristics of online allies and their use of features of social media platforms to disseminate prosocial messaging can cue audience attributions about the users' (in) genuine motives for posting. To that end, this dissertation will adopt an affordance-based approach to test how features related to content sharing on Instagram

affect how viewers evaluate social media users and the prosocial issues they promote. The overall goal of this dissertation research is to aid interested allies, prosocial movement stakeholders, and communication strategists on how to effectively use social media as a tool to show genuine or authentic allyship for historically marginalized groups online.

Chapter 1: Authenticity

Defining authenticity

In today's digital environment, with the prominence of user, company, and bot-generated content, the ability of viewers to discern what is and what is not authentic is important. But how do we define authenticity, and how do computer mediated environments influence authenticity perceptions? Merriam-Webster's Dictionary (2022) defines authenticity as being "true to one's personality, spirit or character," whereas the Oxford Learner's Dictionary (2022) defines authenticity as being "true and accurate." However, academic conceptualizations of authenticity are much more complex.

Some definitions of authenticity conceptualize it as the possession of certain traits. For example, Enli (2015) defined authenticity in mediated communication as trustworthiness (i.e., accuracy and correctness), originality (i.e., being genuine and real), and spontaneity (being true to oneself). Other conceptualizations focus on the consistency between individuals' internal wishes and external behavior or influence. For example, Beer and Brandler (2020) define authenticity as the extent that an individual's daily life reflects their core values and qualities. To be authentic, however, individuals must draw upon their core values and reject external influence, even if the result is unflattering (Beer & Brandler, 2020). Harter (2002, p. 383) suggests that the behavioral component of authenticity involves individuals acting in "accordance" with their true selves and expressing themselves consistently with their thoughts and feelings. Other conceptualizations describe authenticity along the same lines as an enactment

of the true self (Kernis & Goldman, 2004), expression of the genuine self (Cranton, 2004), or genuine expression ignoring pressure from others' expectations (Tisdell, 2003). To that end, many scholars conceptualize authenticity as the consistency between individuals' internal wishes and external behavior.

However, the dimensions that consist of these internal wishes and external behaviors or influences vary. For example, Kernis and Goldman (2006) suggest that authenticity consists of four dimensions: (1) individuals' *awareness* about their inner motives, feelings, goals, strengths and weaknesses, and interest in learning more about themselves. (2) Individuals' ability for *unbiased processing of self-relevant information*, in that people can assess positive and negative aspects of themselves objectively. (3) Individuals' *behavior*, which refers to acting per one's needs and goals. (4) And, finally, individuals' *relational orientation*, which describes an individual's motivation to act honestly with their peers.

Wood and colleagues (2008) also presented a multidimensional conceptualization of authenticity. Authenticity involves consistency between individuals': (1) *primary experience*, which refers to their physiological, emotional, and cognitive state. (2) Individuals' *symbolized awareness*, or their consciousness of their physiological, emotional, and cognitive state. And (3) individuals' outward *behavior and communication*. Wood et al. (2008) suggests that perfect congruence between an individual's internal state (*primary experience*) and their actual experience (*symbolized awareness*) is nearly impossible. Incongruence leads to *self-alienation*, the first dimension of authenticity, which is the experience of not feeling true to oneself. The second dimension of authenticity is the congruence between actual experience and behavior, known as *authentic living*. *Authentic living* involves expressing oneself in a way that is

consistent with physiological states, emotions, beliefs, and cognitions. Thus, the person is living in a manner consistent with their values and beliefs. Lastly, the third dimension of authenticity refers to the extent to which individuals place weight on the expectations of others, known as *accepting external influence*. Together, *self-alienation*, *authentic living*, and *accepting external influenc*e are the tripartite person-centered perspective of authenticity.

Finally, Knoll et al. (2015) proposed that authenticity has only two dimensions: (1) *authentic self-awareness* and (2) *authentic self-expression*. *Authentic self-awareness* refers to an individual knowing their own identity (e.g., feelings, motives, and commitments). *Authentic self-expression* refers to individuals presenting themselves in a way that follows their values, preferences, and needs. Although Kernis and Goldman (2006), Wood et al. (2008), and Knoll et al. (2015) differ in the number of dimensions of authenticity, taken together they have some commonality with unidimensional conceptualizations from Beer and Brandler (2020), Harter (2002), and others; authenticity is the awareness of and consistency between individuals' internal state (e.g., thoughts, feelings, motivations, and goals) and their outward behavior (e.g., how one acts or expresses themselves). Continuing, authenticity may also consider the impact of external influences or expectations and how motivated individuals are to succumb to these forces in expressing themselves genuinely (Beer & Brandler, 2020; Kernis & Goldman, 2006; Tisdell, 2003).

Additionally, some literature has focused on how authenticity may be perceived by audiences in mediated contexts. Lee (2020) extended prior impression management scholarship on authenticity by focusing on how communicative acts are perceived as authentic. Authentic communication, as such, consists of three dimensions: (1) *authenticity of the source*, which refers

to the validating that the communicator is whom they claim to be. In mediated contexts, the source of a message can vary from real individuals to unreal auto-generated bots. Thus, how individuals perceive the source of communication may lead to variability in perceived authenticity. (2) *Authenticity of the message* is based on whether the communication is true. Lee (2020) suggested that if a message is consistent with people's expectations of the real world, it should be considered authentic. (3) *Authenticity of the interaction* refers to whether the interaction between communicating parties was real. Lee (2020, p. 63) suggests CMC may enable individuals to "virtually participate" in an interaction that isn't physically happening or that they are not a part of – yet still feel intimacy towards. Features of CMC that render engagement (e.g., comments, likes) may erode the distinction between public and private communication. Unlike previously discussed conceptualizations of authenticity that solely focused on feelings of genuineness attributed to some source, Lee (2020)'s perspective cross-examines the source, the medium, and the message which may all interact in a digital environment. In sum, authenticity can be defined as the consistency between an individual's inner self and their outward persona. However, viewpoints on how to conceptualize authenticity as a construct vary. Although some scholars suggest that authenticity is a unidimensional construct, others contend that the construct is multidimensional and is tied to both inner values and outward behavior. With the rise of computer-mediated communication, new conceptualizations of authenticity have focused on the communication act, rather than centered on the consistency of the identity of the communicator.

Authenticity research

Authenticity can be studied from the perspective of both the sender and the receiver. To clarify, authenticity can be examined from the perspective of how authentic or genuine *one* feels or how authentic or genuine an individual perceives *others* to be. However, how authentic or genuine an individual is perceived to be by others may be obscured by their attempts to control or influence how others evaluate them. *Impression management*, or self-presentation, can be used to describe an individual's attempt to control impressions made about them by others (Leary & Kowalski, 1990). People may be motivated to influence the impressions of others for a variety of reasons including: the potential for social and material benefit, self-esteem, and identity development (Leary & Kowalski, 1990). Discussion of self-presentation and impression management traces back to sociologist Goffman (1959)'s dramaturgical metaphor in *The Presentation of Self in Everyday Life.* Goffman (1959) characterizes people as performers who attempt to leave a particular impression on audience members (Picone, 2015). Individuals create a 'front,' or a public persona, that they perform when interacting with their audience. Individuals' 'frontstage' persona is idealized, and may involve behaving in a particular manner, or relaying (mis)information to control how their audience perceives them, or to influence audience reaction (Goffman, 1959; Solomon et al., 2013). While Goffman (1959) intended his dramaturgical metaphor to apply to face-to-face interactions, the principles of impression management can be applied to interactions in computer-mediated contexts. Just as people may attempt to put forth an inauthentic 'front' in face-to-face contexts, in computer-mediated environments, authentic expression can be variable due to individuals use of attributes of computer-mediated communication (CMC), which may augment how authenticity is perceived. Authentic expression online consists of an individual's online self-matching their real-world or

offline self (Salisbury & Pooley, 2017). However, attributes of computer-mediated communication (CMC) may cause users to potentially selectively self-present to reveal only socially desirable aspects of themselves to their online audiences (Choi & Bazarova, 2020; DeVito et al., 2017; Hollenbaugh, 2021; Walther, 1996; Papacharissi, 2002).

Some literature suggests that authentic presentation may be closely tied to the affordances of CMC. For example, examining authenticity from the perspective of the source of social media content, both Choi et al. (2020) and Kreling et al. (2022) emphasize the role affordances have in online self-effects. Self-effects are defined as the impact that messages have on senders; posting information about oneself on social media can facilitate the internalization process and open senders up to self-related feedback (Valkenburg, 2017). According to Valkenburg (2017), affordances of CMC encourage self-effects as the asynchronous nature of online social platforms allows users to craft more constructed versions of themselves, all while reaching larger audiences and opening themselves up to feedback due to the scalability of their platform of choice. The ability to receive more self-related feedback thus can lead to heightened self-effects (Valkenburg, 2017).

Both Choi et al. (2020) and Kreling et al. (2022) took an affordance approach in understanding how features of online social platforms can impact impression management concerns. Popular social media platforms such as Instagram, WhatsApp, and Snapchat have adopted features that have default deletion attributes in their design. “Story” features typically have a 24-hour time window until the post disappears from public view. Thus, “Stories” are ephemeral as compared to traditional social media “Posts” which remain permanent and visible on profiles (Choi et al., 2020; Kreling et al., 2022). Interviews suggest that ephemerality on

social media encourages more authentic behavior from users by reducing self-consciousness and increasing spontaneity (Bayer et al., 2015; Xu et al., 2016). Focusing on the impact of self-concept change on social media, Choi et al. (2020) found that ephemerality positively impacted individuals self-symbolizing efforts (i.e., individuals' motivation to attain aspects of an identity they want to acquire). Self-symbolization efforts are considered identity-constructing, in that individuals may acquire tangible symbols related to that identity and participate in identity-relevant activities (Choi et al., 2020). Specifically, Choi et al. (2020, p. 1) found that ephemerality not only enhanced self-symbolizing efforts but also led to increased internalization of self-symbolized identity by "affording non-strategic self-presentation and reducing impression management concerns" affording people the ability to display their true selves. Research from Kreling and colleagues (2022) found similar results. Comparing Instagram Story features to Posts, Kreling et al. (2022) found that Instagram Stories produced (slightly) higher authenticity perceptions. Kreling et al. (2022) suggested that users differentially evaluated themselves as authentic based on the Instagram feature they used. Instagram Stories may be viewed as a more spontaneous form of expression, and thus drive more authentic self-presentation.

Selective self-presentation can have social implications on perceivers of online content. Previous research has suggested that viewers are often skeptical of the extent to which online self-presentations are accurate representations of individuals' offline selves (Hancock & Toma, 2009). To further explain the receiver effects of authenticity in online self-presentations, Vendemia and DeAndrea (2018) utilized warranting theory to provide a framework to explain the viewer effects of authenticity in online self-presentation. Warranting theory suggests that online viewers are more likely to believe information genuinely reflects an offline reality (i.e., is

authentic) when they perceive information is not controlled or strategically modified by the proprietor of the information (DeAndrea, 2014; Walther and Parks, 2002). Vendemia and DeAndrea (2018) noted that a large amount of warranting research has examined how written information made about a target influences authenticity evaluations. However, in the context of online social media platforms, Vendemia and DeAndrea (2018, p. 119) suggested that the perceived authenticity of visual self-presentations can be affected by the degree to which viewers thought social media users edited their visual content. Additionally, perceptions of inauthenticity have the potential for negative evaluations of sources. In the context of Vendemia and DeAndrea (2018), the more viewers perceived photographs shared online to be digitally modified the more likely viewers attributed negative evaluations to the source of the photo. A study by DeAndrea et al. (2018) found that the more individuals perceived the target of online reviews (i.e., user-generated reviews) to be in control of information about them, the less influential those reviews were in impression formation.

Considering Credibility

A construct similar to but distinct from authenticity is credibility. According to Oxford Learner's Dictionary (2022), credibility is the "quality that someone has that makes people believe or trust them." Hovland et al. (1953) define source credibility as the "believability of the communicator." However, there is still debate on the distinct dimensions of source credibility with the inclusion of factors such as source reliability, composure, sociability, goodwill, dynamism, safety, qualification, competence, objectivity, likeability and the receiver's perceived similarity to the source (Choi & Stvilia, 2015; Flanagin & Metzger, 2020). Generally speaking, conceptualizations of the construct commonly include the dimensions of expertise/competence

(i.e., the extent that a perceiver believes a source knows the truth) and trustworthiness (i.e., the degree to which a perceiver is being honest), with the addition of goodwill (i.e., the degree to which a perceiver thinks that a source has their best interest in mind; Choi & Stvilia, 2015; Hovland et al., 1953, Flanagin & Metzger, 2020; Westerman et al., 2012)

Similar to authenticity, computer-mediated channels pose an issue in understanding credibility judgments. Unlike decades before, where information collection and dissemination were high-cost and controlled by a limited number of providers, the Internet provides low-cost, unrestricted access to information (Metzger, 2007; Metzger & Flanagin, 2013). Pre-digital age information providers were deemed highly credible (e.g., doctors, government entities, librarians, etc.,) having high expertise and serving as gatekeepers of information (Metzger et al., 2010). However, the Internet disrupts the top-down dynamic of information dissemination with user-generated content through social media platforms (e.g., Twitter, Instagram, Facebook) or wikis (e.g., Wikipedia) and the sheer propensity and swiftness that information can be created, stored, and consumed online.

Along with assessing viewers' perception of the credibility of the source of communication, credibility researchers have also begun addressing how specific attributes of online resources (i.e., cues) can be used to make systematic or objective credibility assessments (Choi & Stvilia, 2015; Fogg et al., 2003). Given the vast amount of information available online, scholars suggest that due to the perceived cost of "information search and overload" internet users seek to minimize their cognitive effort and time by using heuristics (i.e., cues; Metzger et al., 2010, p. 417). There are several perspectives related to how individuals employ heuristics to make credibility judgments online. For example, Metzger et al. (2010) suggest internet users rely

on six cognitive heuristics to evaluate credibility: reputation (i.e., name recognition), endorsement (i.e., the site is known to others, aggregated reviews, ratings), consistency (i.e., checking source credentials., last update, considering bias), self-confirmation (i.e., confirmation of pre-existing beliefs), expectancy violation (i.e., the site fails to meet their expectations), and persuasive intent (i.e., advertising is present). Work by both Fogg et al. (2003) and Choi and Stvilia (2015) show that that are several types of cues that can be indicative of credibility (e.g., design look, advertising tone, writing, affiliations, site sponsor identity, etc.,). However, Choi and Stvilia (2015) categorized cues concerning if they were a feature of the operator (i.e., source), content (i.e., message), and design (i.e., medium) and mapped them onto Hovland et al.'s (1953) two-dimensional conceptualization of credibility (i.e., trustworthiness and expertise) to understand credibility assessment.

How do credibility and authenticity differ, and why is it important to differentiate authenticity from credibility? First, credibility focuses on the perceived competence, trust and goodwill individuals have in some entity (Lee, 2020). Authenticity, on the other hand, focuses on the realness or the genuineness of a source, the message, or the interaction a user has online (Lee, 2020). Lee (2020) asserts that while (information) credibility research primarily focuses on the spread and influence of information through conventional channels of both mass and computer-mediated communication, authenticity research applies to a wider variety of computer-mediated channels and can have both socio-emotional and relational influence (e.g., how a celebrity interacts with fans online). With the use of social media platforms such as Instagram to not only promote individual identity but also promote awareness of social causes and displays of

presumed allyship – verification of information as well as the (in)genuineness, or *performative*, nature of the source should be evaluated.

Chapter 2: (Performative) allyship in digital spaces

Online prosocial movements such as #MeToo, #BlackLivesMatter, #SayHerName and more have become calls to action, shows of solidarity and launching points for social change in digital spaces. While these hashtags and affiliated prosocial movements have brought groups of people together to broker change, they have also launched critical discussion about how the adoption of prosocial messaging on social media platforms has been used as an impression management strategy, rather than with genuine interest in advancing diversity, equity and inclusion, putting movements at risk (Wellman, 2022). So, how do we understand what (in)authentic displays of allyship, or prosocial behavior look like in online prosocial movements?

Allyship

Prosocial behaviors cover a variety of voluntary actions that an individual can take to benefit people other than themselves (Batson & Powell, 2003). Prosocial behaviors may include helping others in need (e.g., returning a lost item, or being an active bystander), donating time or money, sharing resources, giving advice, abiding by rules and laws, volunteering, or acting sustainably (Batson & Powell, 2003; Labroo et al., 2022). Thus, acting as an *ally,* might be considered a form of prosocial behavior. Allies are members of advantaged groups committed to supporting, creating opportunities for, and improving the status of marginalized or disadvantaged groups (Knowlton et al., 2023). Allyship can occur in support of a number of marginalized identity groups including gender minorities, sexual minorities, racial minorities, or those who are treated inequitably based on age, religion, disability status or other social identities (Collier-

Spruel & Ryan, 2022). Examples of advantaged groups advocating for disadvantaged groups may include men demonstrating support for gender equality initiatives or abortion rights, heterosexuals supporting marriage equality, or White people supporting Black Lives Matter efforts. Presently, allied efforts can manifest in many forms, for example people may show solidarity by demonstrating for a common cause in person, or with the rise of social media platforms and digital activism, individuals may use social media platforms such as Facebook, Instagram, Twitter and more to show solidarity for disadvantaged groups.

Conceptualizations of allyship vary. Some conceptualizations of allyship have focused on how members of disadvantaged groups qualify allyship. Chen and colleagues (2023) center the meaning of allyship from the perspective of members of disadvantaged groups. Focusing on how members of the lesbian, gay, bisexual, and transgender (LGBT) community perceived self-identified allies, Chen et al. (2023) found that LGBT participants considered there to be multiple components of allyship. In conceptualizing allyship, LGBT participants considered allies to include individuals acting nonprejudiced, acting against perceived discrimination and inequality, and advantaged individuals having humility about their perspective. In another study Collier-Spruel and Ryan (2022) asked participants belonging to marginalized groups (e.g., racial, and ethnic minorities, sexual and gender minorities) how they defined effective allyship. Effective allyship was characterized by participants as increased inclusion, acceptance, empowerment, advocacy and decreasing bias.

Other work has suggested that allyship goes beyond advocating for the improvement of the disadvantaged group members' status and is instead inclusive of leveraging one's positional power to challenge the status quo. For example, Salter and Migliaccio (2019) suggest that a core

component of allyship behaviors is communication and confrontation. Allies should not only be willing to be (self) reflective and (self) aware concerning disadvantage group members' experiences, but also communicate about and confront prejudice and discrimination. Continuing, Erskine and Bilimoria (2019) center their conceptualization of allyship within the group dynamics of Afro-Diasporic women and their White colleagues in the workplace. Erskine and Bilimoria (2019, pp 319) define white allyship as an individual "leveraging [their] position of power and privilege" by disrupting "the status quo." Additionally, Erskine and Bilimoria (2019) suggest that in organizational settings, White individuals in positions of power should engage in prosocial behaviors to support Afro-Diasporic women's advancement. Other conceptualizations have characterized allyship as the use of power and privilege to disrupt inequitable systems (Philips et al., 2009; Rosette & Thompson, 2005) or the commitment to the use of racial privilege to promote equity (Goodman, 2011).

Taking on allyship, thus, involves advantaged individuals taking on a level of interpersonal risk between themselves and their social context by challenging the status quo (Pyram, 2023). Advantaged group members may risk losing power, privilege or social capital afforded to them by challenging dominant forces or opinion (Thai & Nylund, 2023). For example, challenging a peer may lead to social consequences within friend groups, whereas challenging a coworker or a higher-up in a workplace setting may lead to financial consequences (Erskine & Bilimoria, 2019). Additionally, advantaged group members risk experiencing scrutiny, or critical feedback from disadvantaged group members for their efforts and may internalize as being a personal reflection of their ability or judgement (Knowlton et al., 2023). These risks even translate to computer mediated environments. Roden (2023) in their study note

that White interviewees viewed there to be social risks in posting about racial justice online. Participants cited wanting to avoid destructive conversation, and fearing losing family, friends, and followers as a reason to refrain from posting online. To that end research suggests disadvantaged group members consider "allyship-associated costs and rewards" when making attributions about allies (Thai & Nylund, 2023, pp. 131). Allies who appear to be facing more costs in acting are perceived to be more authentic in their allied behaviors as compared to those perceived to be facing no cost or gaining rewards. Altogether, allyship can be conceptualized as individuals working in support of disadvantaged groups and using their positional power to disrupt or challenge the status quo; however, as Louis and colleagues (2019) note, conceptual attributes of allyship do not include self-focused motives that can ultimately impact the effect of allied efforts.

Performative allyship in online settings

Central to the current study is the examination of how displays of prosocial behavior may be perceived as inauthentic, or performative displays of allyship. Allies, or members of advantaged groups may be motivated to engage in action for marginalized groups due to self-interests such as improving their image, popularity, or financial status (Radke et al., 2020). For example, research on motivation to volunteer suggests that while people may conduct service to fulfill humanitarian values, some prosocial behaviors may be extrinsically motivated*;* people volunteer to enhance their self-esteem, gain experience that may be beneficial to their careers, gain skills, or increase their social network (Konrath et al., 2012; Kutlaca & Radke, 2022; Omoto & Snyder, 1995). As such, some scholars contend that prosocial behaviors that are extrinsically motivated are comparable to a form of impression management (Lavertu et al., 2020) — the act

of individuals expressing themselves in a constructed manner to control the perceptions others form of them (Leary & Kowalski, 1990). In fact, some scholars contend that ingenuine expressions of care about social issues are strategic and are a result of evaluating the social context one is in to determine how to represent oneself in the best light (Jordan & Rand, 2019).

The proliferation of virtual participation and active displays of allyship online (e.g., rainbow filters to support the LGBTQ+ movement, black squares for BlackLivesMatter), have raised concerns about the potential for performative, or ingenuine displays of allyship by dominant group members online (Wellman, 2022). *Performative allyship* is characterized as an ingenuine display of allyship enacted with the goal of enhancing popularity or self-image rather than contributing to a social cause (Katalina, 2020; Kutlaca & Radke, 2022). Social media platforms provide ample opportunity for individuals to manage their positive self-presentations and acquire feedback from virtual audiences. Online virality metrics (i.e., following, likes, reposts) or feedback (i.e., comments) may serve as social rewards, or behavioral motivators, to engage in self-presentation related activities. Likes, comments, and follows may all serve as cues indicating how other users have engaged with or evaluated content posted online (Lee et al., 2021). Thus, reception of positive feedback (i.e., likes, comments) from peers may induce positive self-perception and affect, fulfilling users' need for belonging from their peers online (Baker et al., 2019; Chua & Chang, 2016; Grinberg et al., 2017; Zell & Moeller, 2018). Related to online displays of allyship or solidarity, performative allyship may be exemplified by one-off social media posts, utilization of hashtags, or the profession of support in order to receive praise from their online audiences (Spielman et al., 2023). Accordingly, performative allyship can be characterized as being motivated by an individual's "need for validation and acceptance" without

sacrificing social capital and can be perceived as empty (Erskine & Bilimoria, 2019;Wellman, 2022, pp. 2).

Research in psychology, communication, and political science have identified various forms of inauthentic prosocial behavior both offline and online (Kutlaca & Radke, 2020). For example, research on politics and activism have examined 'slacktivism,' which is defined as "feel-good but useless Internet activism" (Morozov, 2009, pp. 13). Some scholars contend that some digital activism is of low cost and risk to the actor, and that at its core slacktivism lacks demonstrable impact in solving the issue at large (Smith et al., 2019). Continuing, the phrase 'virtue signaling' has been used to describe "public expression of moral values or support for a cause" with the goal to increase one's image or notability (Kutlaca & Radke, 2022, pp. 3; Westra, 2021). Another concept, moral grandstanding, was examined by Grubbs et al. (2019) in their study which explored status-seeking motives as a potential explanatory mechanism for moral grandstanding. Moral grandstanding is defined as the "use of public moral discourse for self-promotion and status attainment" (Grubbs et al., 2019, pp. 3). Individuals may take part in public discussions about topics pertaining to morality, or politics in order to impress others. Grubbs et al. (2019) found across a set of politically diverse samples that moral grandstanding was associated with status-seeking personality traits. Continuing, Wellman (2022) showed that social media influencers often mask credibility concerns with (ingenuine) displays of allyship online. Wellman (2022) specifies that during the 2020 #BlackLivesMatter movement, White wellness influencers exhibited empty online efforts (e.g., creating, or re-producing content with the #BlackLivesMatter hashtag) without substantially contributing to any other forms of advocacy. Through extensive interviews, Wellman (2022) suggests that White wellness creators

viewed presenting as an ally as desirable as it increased their credibility and managed audience expectations, indicating that displays of allyship were inauthentic.

Finally, social psychology research has also experimentally explored indications of self-interest and prosocial behavior in online settings. For example, Lavertu and colleagues (2020) found that increased salience of online audiences led participants to show greater intention to engage in prosocial behavior offline, mediated by higher public self-awareness and extrinsic motivations (e.g., maintaining self-image, fitting in with social norms, displaying social capital, morality, or time spent). In another study, Choi and Seo (2017) found that individuals who were status-seeking exhibited a high level of prosocial behavior online as compared to non-status-seeking individuals (Choi & Seo, 2017). Cox et al. (2018) found similar results; after analyzing the prosocial behavior of users of a virtual crowdfunding platform, their findings showed that individuals who had a publicly visible profile tended to lend more, as compared with those with a private profile. Across multiple domains, scholarship has identified a common phenomenon of inauthentic displays of allied concern, or support for disadvantaged groups for the purposes of enhancing one's own image. As social media platforms such as Instagram have been characterized as venues for users to take part in a form of visibility labor, or strategically conducting profile work to be viewed in a desirable manner by others, (Abidin, 2016; Duffy & Hund, 2016), online activism efforts may be especially vulnerable to those looking to act prosocially in pursuit of their own self-interest, rather than genuine concern for disadvantaged groups.

Chapter 3: Features of communication technology

In the context of prosocial movements, features native to social media platforms such as Twitter, Instagram, and Facebook give individuals the opportunity to advance movement impact through the connection building, mobilization, expanding resources, and amplifying key narratives (Mundt et al., 2019). However, as research on the intersection of communication technology and social activism has identified variety in the use of features of social media to amplify prosocial messages, research has also alluded that properties of these platforms can also be exploited for impression management concerns.

Research examining prosocial movements and communication technology have examined various ways social media platforms can be used to amplify prosocial movements. Some studies have focused on the typology of prosocial content online. For example, Cornet et al. (2017) in their study examined how different image based social media platforms support various social movements. Results showed that prosocial content online takes a variety of forms including photographs taken during rallies or other offline political activities, selfies, user-generated infographics and art, repurposed media, and commercial images that users will post themselves or reshare from others (Cornet et al., 2017). Other studies have focused on the substance of content generated under popular hashtags associated with movements. For example, Jackson et al. (2020) examines how features of Twitter, (now X), such as hashtagging, allow users to connect and mobilize around several causes focused on disenfranchised populations. Einwohner and Rochford (2019) in their study sought to understand how Instagram

was used to maintain the presence of the Women's March through examining users' public posts that included the hashtag #womensmarch and posts from the official Instagram account of the organization Women's March. Other research related to prosocial movements online have focused on the use of specific features and their influence on social media users. For example, Li (2022) surveyed Black Instagram users to examine their use of Instagram and its specific features (e.g., Instagram Carousel, Live, Story, Feed Posts, and comment section) and how it influenced their Black activism orientation, in app activism, and identity.

While some scholars have examined how specific movements exist online, others have explored how features of social media platforms can be used not only to promote prosocial movements but also social media users themselves, bringing into question the authenticity of allyship online. Specifically, scholars have identified that social media users' intent to share prosocial messaging may be partially self-serving; content creators may want to maintain good appearance with their followings or promote themselves (Abidin, 2022; Riedl et al., 2021; Wellman, 2022). For example, Riedel et al. (2021) introduces the phenomena of 'political influencing,' where content creators focus their online presence on more meaningful political content such as issues of sustainability, immigration, or LGBTQ+ rights. However, Riedel et al. (2021) note that while content creators make political statements about issues of concern, they simultaneously embed self-promotive or advertising hashtags (e.g., #sponsored, #ad) within their content – creating a paradox between the promotion of a prosocial cause or issue, and the use of self-promotive features of the app for explicit personal gain. In another study, Abidin (2022) examines the phenomenon of grief "hypejacking" on Instagram, where individuals misappropriate high-visibility hashtags in order to enhance their profile publicity. Abidin (2022)

notes that users often capitalize on trending hashtags, display themselves publicly grieving, use unrelated hashtags to boost engagement, and copy or edit photos from other sources to bring attention to themselves in proximity to the tragic event in question. Abidin (2022, pp. 184) alleges that individual's pursuit of social capital on Instagram – most commonly thorough gaining likes, positive comments, and shares – may drive users' attempts to self-promote and "game" the platform. Taken together, these works suggest that although social media platforms may be venues for the promotion of prosocial movements and issues, bad actors exploit the features of these platforms to promote themselves or maintain their image.

Instagram, in particular, has been characterized as fostering 'visibility labor,' in that users partake in a level of attention seeking and profile work to be viewed desirably by others and for personal gain (Abidin, 2016). Users may exploit features of the platform such as hashtags, @mentions or tags to increase their visibility, or may strategically post content in order to minimize criticism from their audience (Abidin, 2016; Duffy & Hund 2016). As evidenced by the work of Wellman (2022), Riedel and colleagues (2021) and Abidin (2022), presumed allies of prosocial movements may have differing motives related to their online participation, which may be signaled through the use of platform specific features. Although extant scholarship does somewhat cover the use of features native to social media platforms, less work seeks to tease out how characteristics, or affordances, of these platforms may signal different meaning to audiences about the intent of the prosocial content creator, subsequently impacting the success of the movement. Thus, it may be appropriate to explore the implications of the use of specific features native to social media platforms using an *affordance* based approach, which not only allows scholars a systematic way to examine differences between the use of different channels of

communication technology, and the features they host, but also how the use of features may differ across contexts leading to differential social and psychological outcomes.

Defining Affordances

Gibson (1979, 1986) first defined the term *affordance* to describe how the utility of an object could be understood differently by various animals due to their perception of the function of the object (Fox & McEwan, 2017; Treem & Leonardi, 2013). According to Gibson (1979, 1986) different animals perceive a particular set of uses that an object provides, coining the term *affordance*, or the perceived utility of an object (Treem & Leonardi, 2013). Gibson (1979, 1986) suggested that affordances hold different meaning to different people, regardless of an objects' features being common to all (Fox & McEwan, 2017; Treem & Leonardi, 2013). Other perspectives of the nature of affordances and technology suggest that affordances stem from user's experience with the technology itself (Norman, 1990) or from adaptation and experimentation with technology over time (Gaver, 1991; Leonardi, 2011; see Evans et al., 2017; Fox & McEwan, 2017). Keeping both perspectives in mind, Evans and colleagues (2017) suggest that the neither the materiality (e.g., features of an object) of technology nor human agency alone influence technology use. Thus, taking an affordance approach to communication technology can be advantageous to researchers as they can examine the relationship between people and the technological features they come into contact with. As people may interact with technology with different goals in mind, they may perceive technology use differently across contexts, leading to differential social and psychological outcomes (Treem & Leonardi, 2013).

While efforts to spread awareness via social media to promote prominent social issues is lauded, as noted earlier in this manuscript, Wellman (2022) highlighted that aspects of online

activism may be more performative than authentic. Instagram in particular is not only a venue for online activism, but also self-promotion, making it an opportune platform to study. Using an affordance based approach may be appropriate to understanding how Instagram features used to promote prosocial content can impact audience perception of prosocial movements. While much literature on prosocial movements has focused on the use of hashtags related to movement messaging, more work is needed examining the channel of communication used to share content on the platform and what meaning its use may convey to viewers. Instagram Stories and Feed Posts are both widely used features on the platform and are distinct in their duration and presentation of content. These features may trigger different attributions in viewers related to the perceived authenticity of the content creator. Thus, this dissertation aims to compare viewer perceptions of prosocial post(er) authenticity via two social media platform features native to Instagram – Stories vs. Posts – along a spectrum of two affordance categories: persistence and visibility (Treem & Leonardi, 2013) as well as perceived source control over audience feedback.

Persistence

Persistence is defined as the relative permeance or ephemerality of a communicative message (Fox & McEwan, 2017). Communication can be considered persistent if it remains accessible, in its original form and does not expire after its author has published it (Treem & Leonardi, 2013). The opposite of persistence, ephemerality, is reflected in popular social media platforms such as Snapchat and the Stories feature on Instagram. Snapchat is characterized by its use of temporary messaging; messaging and images delete upon opening. Instagram Stories allow users to share photos and videos in the moment or posts from other users for 24 hours before they permanently disappear to the public (Kreling et al., 2022).

Research on platforms such as Snapchat or features like Instagram Stories have associated ephemerality with reduced self-presentational concerns (Bayer et al., 2020). For example, work from Bayer et al. (2016) utilized experience sampling data, and in-depth interviews to understand college students' perception of social and emotional experiences related to Snapchat use. Quantitative results showed that participants perceived Snapchat interactions as more enjoyable and associated with positive mood as compared to other communication platforms. Qualitative results illuminated that participants saw the ephemeral platform as a place to share "mundane experiences with close ties," and had less concern for self-presentation (Bayer et al., 2016, pp. 956). Along the same lines, McRoberts et al. (2017) interviewed participants and reviewed 14 days of participants Snapchat Stories, a feature similar to Instagram Stories. Results showed that participants viewed Snapchat Stories as a way to share about themselves "without any specific planning" and in an "improvisational manner" (McRoberts et al., 2017, pp. 6906-6907). However, while communication technology such as Instagram Stories or Snapchat allow for increased spontaneity in posting, temporary sharing may leave negative impressions in observers (Bayer et al., 2020). Hofstetter et al. (2017) in a series of experiments found that temporary sharing led participants to share more risky selfies of themselves as compared to sharing permanently. As a result, Hofstetter et al. (2017) note, reduced inhibition in sharing from sharers, led observers' impressions of sharers to be more negative.

In contrast to ephemeral channels of communication, such as Instagram Stories, Instagram Feed Posts are categorized as more persistent forms of communication. Instagram Feed Posts remain permanently visible on users' profiles unless they are archived or deleted. Content that is more persistent, thus may be more accessible long after its initial conception (Fox

& McEwan, 2017). With this in mind, when posting content online, social media users may be aware that their content, and thus their self-presentations, are observable to others and may be subject to scrutiny. In fact, research on self-presentation suggests that awareness of publicness of social media content can impact how users present themselves online. For example, research has shown that both adults and adolescent social media users effortfully, and selectively present self-related information and images in order to manage how their audience (i.e., followers) perceive them (Chua & Chang, 2016; Huang & Vitak, 2022; Yau & Reich, 2018). Thus, when social media users opt to use more persistent channels of communication, there is not only an awareness that content is reviewable by online audience members, but there is subsequently more effort placed in strategic self-presentation levied by the potential for audience feedback.

Given discourse from scholars on the phenomena of social media activism and its efficacy (Jackson et al., 2020; Wellman, 2022), social media allies choice in how to share prosocial content may (mis)communicate perceived goals related to prosocial movements to their online audiences. Wellman (2022, pp. 7), for example notes that in the context of White allies sharing content online related to BLM, the use of features such as Stories, connoted "surface-level attempts at allyship" where users "blindly re-post[ed] without engaging [with content]" to observers online. When prosocial content is posted on ephemeral channels, it stands to reason that viewers may presume a lack of time or effort placed in posting content as compared to a persistent channel like Instagram Feed Posts. This, coupled with previous findings that suggest that features associated with ephemerality are linked with reduced self-presentational concerns through more "in the moment" or spontaneous posting (i.e., less planning) (Bayer et al., 2020; McRoberts et al., 2017; Kreling et al., 2022) leads to the following hypothesis:

H1: The more persistent prosocial posts are perceived to be, the less viewers will perceive the sources of the content as posting spontaneously.

Additionally, while digital social activism may be viewed as being low cost, sharing social or political commentary with audiences can be seen as risky (Thorson, 2014). Engaging in political discussion (of which many prosocial movements encompass) can leave communicators vulnerable to encountering disagreement, or experience discomfort, while additionally posing the risk of disrupting otherwise amicable relationships (Mutz, 2006; Thorson, 2014). For example, Thorson (2014) found via in-depth interviews that social media users are often concerned about the social repercussions of their political posts, citing uncertainty in how posts may be interpreted and how their online audiences may (un)favorably react (Thorson, 2014). With this in mind, participants described strategizing ways to make posts more palatable to their online audiences to avoid conflict in the comments or unfriending (Thorson, 2014). Along the same lines, Wellman (2022) observed in her study that White allies who posted about BLM reported having to navigate audience expectations, political beliefs and opinions regarding content they 'should' be posting related to BLM.

As persistent communication can be "searched, browsed, replayed, annotated, visualized, restricted, and recontextualized" by both the source of content and third-party observers online (Erickson & Kellogg, 2000,p. 68; Treem & Leonardi, 2013) – it stands to reason that online allies who post prosocial content face not only a greater chance of other users viewing their content, but making subsequent judgements about their content. Thus, online allies' choice in using a more persistent channel of communication (e.g., Instagram Feed Posts) may signal to viewers that users are committed to the prosocial issue they are sharing about online due to its

permeance on their profiles. When individuals commit to *something* (e.g., a set of ideals, a goal, an organization, another person, a prosocial movement etc.,) they are "pledging or binding" themselves to something else, and agreeing to a framework that they will follow and can be judged on (Schlenker et al., 1994, pp. 637). In the context of support of prosocial movements online, social media content posted serves as a public digital archive, of a presumed allies' support and commitment of a cause or movement. However, viewers may question content creators' level of commitment if prosocial messaging only remains for a short duration (i.e., is ephemeral). Thus, I hypothesize the following:

H2: The more prosocial posts are perceived to be persistent, the more viewers will perceive the sources of the content as being committed to the prosocial movement/issue.

Additionally, viewer perceptions of commitment may subsequently impact how authentic viewers perceive the source of prosocial content to be. In the context of prosocial behavior, research suggests that factors that shape observers' judgements of a prosocial actors' true motives include the perceived cost of acting prosocially (e.g., sacrifice of time, money, effort) and if observers suspect individuals are acting in self-interest (Berman & Silver, 2022). Specifically, when observers perceive prosocial behavior to be costly, they make more positive attributions about the prosocial actor (McAndrew & Perilloux, 2012; Olivola & Shafir, 2018). In contrast, research shows if observers perceive the prosocial behavior as being motivated by selfishness (e.g., profits or reputation) they may deem it insincere and inauthentic (Lin-Healy & Small, 2012; Newman & Cain, 2014; Silver et al., 2020). Along the same lines, prosocial posters who post on more persistent channels of communication may be viewed as more genuine proponents for a cause, and less self-serving, as their posts signal their longstanding public

association with an issue, thus opening themselves up to the potential of criticism from members of their online audience. As such, I hypothesize the following:

> **H3:** The more prosocial posts are perceived to be persistent, the more viewers will perceive the sources of the content as being committed to the prosocial movement/issue, and thus will rate the sources as **(a)** an authentic ally, as having **(b)** more prosocial reasons for posting and **(c)** less self-interested reasons for posting.

Visibility

Visibility is defined as the ability for communication technology to showcase users' online activity, and communication network to other users (Treem & Leonardi, 2013; Treem et al., 2020; Fox & McEwan, 2017). According to Treem and Leonardi (2013, pp. 150) social media platforms gives users the ability to showcase their "behaviors, knowledge [and] preferences," that once were invisible to a broader audience. Communication technology afford users visibility in several ways, giving users agency in how accessible they can make communication to third parties (Treem and Leonardi 2020). Communication technology such as texting, phone, or email are considered to be less visible, or more private, channels of communication as information is only shared with specific individuals (Fox & McEwan, 2017). In contrast, communication posted on public websites or social media platforms are considered more visible due to their accessibility on the web. For example, on social media, visibility of activity may include users' posts, like metrics, comments, status updates, friending, editing history, pictures, and other actions that are visible to other users on the platform (Treem & Leonardi, 2013). Thus, the visibility of activity allows social media users not only to display

information about themselves and their interests, but also for third parties to observe and make judgements about content they consume.

Visibility is bidirectional, in that it is characterized by the relationship between the activity of one social media user and their followers or observers. Treem and Leonardi (2020) suggest that there is thus an interdependence between users' efforts to be seen and make their activity visible, third party users' efforts to see others online activities, and the extent to which the platform makes that visibility easier or more difficult. Thus, communication visibility is partially determined by third parties (Leonardi, 2014; Treem et al., 2020). For example, DiMicco et al. (2009) found that IBM employees who used an internal social networking site, Beehive, used visible information from the site (e.g., profile information, photos, personal information) to learn about the backgrounds, interests, and activities of their coworkers. Leonardi (2014) through their field study implementing a new social networking platform within a large corporation found similar results. Members of the corporation were able to make inferences about other coworkers' knowledge and expertise based on messages sent and who their coworkers regularly communicated with; this in turn led to more innovation in products and services developed and less duplication of ideas within the company. Thus, Leonardi (2014) suggests that making once private communication visible to third parties, can help third-party observers gain information and make inferences about those they are observing.

However, features on social media platforms such as Instagram offer users agency in how visible they desire their content and activity to be. According to Kreling et al. (2022) visibility is higher for Posts as compared to Stories as it relates to (1) how viewable content is to other users and the (2) how viewable interactions between the content creator and other users are. With

respect to content visibility, users are able to adjust Story privacy settings to be seen by all followers, a select few on a customizable list (i.e., Close Friends) or specifically hiding Stories from certain followers. Posts, however, offer less flexibility in visibility. Posts can either be seen by all followers, or virtually any Instagram user if an individual's profile is public. With respect to visibility of interaction, Instagram users are able to publicly view interaction between Instagram posters and their followers on Posts. Other users can publicly like and comment on Posts, which can be viewed by third-party users. In contrast, Stories offer low visibility in interaction; Story replies are not able to be viewed by third parties (e.g., followers) – reactions to Stories are sent via private message and are one-on-one interactions between the user and the specific individual reacting to their Story.

Kreling et al. (2022) suggest that the distinction between the visibility of Stories and Posts is related to perceptions of *privacy control*. Privacy, the authors define as the "selective control of access to the self or one's group" (Altman, 1975 pp. 18; Kreling et al., 2022). Social media platforms, as compared to face-to-face interactions, give users greater ability to control what thoughts and emotions are disclosed to others through the implementation of privacy settings (Trepte & Reinecke, 2011). Continuing, Trepte and Reinecke (2011) argue that privacy control and authenticity on social media platforms are inherently linked. If users feel that they have increased control over their audience – they are able to be more authentic. Kreling et al. (2022), as a result, suggest that higher levels of perceived privacy, are related to higher levels of authenticity; thus, Stories should induce more authenticity in content creators.

A subset of cross sectional (Vranken et al., 2022), and qualitative work (Trieu & Baym, 2020; Li et al., 2021) seems to agree with this assertion. Vranken et al. (2022) conducted a

cross-sectional survey examining the relationship between social media use and disclosure of alcohol related behaviors. Using an affordance based approach across a continuum of two dimensions: temporality (ephemeral to persistent) and accessibility (private to public) they found that individuals were more likely to make extreme alcohol references (e.g., binge drinking) on both ephemeral-private (e.g., private stories) and persistent-private (e.g., private posts) as compared to both public ephemeral and public persistent content. In another study, Li and colleagues (2021) found that female athletes use of the Stories feature empowered them to act more authentically and share more about their private lives. The authors suggest that having more rigid control over interactions with followers as conversations were kept private from third parties, led to more comfortability sharing (Li et al., 2021). Finally, Trieu and Baym (2020) conducted semi-structured interviews to explore how Instagram users understand the process of sharing through Stories. Users indicated that they experienced less pressure to self-present, thus they could be more authentic. Altogether, these studies suggest that there is a distinct relationship between the nature of visibility, privacy control, and how producers of social media content perceive their self-expression online.

While the previously discussed research highlights how visibility may influence what content creators share online, as noted by Trieu and Baym (2020), users' choice in what features to use to share content may have implications with respect to third party viewers. Specifically, Trieu and Baym (2020) suggest that an implication of sharing content via Stories as compared to Instagram Feed Posts is that that while the Story itself is public, third-party observers are not aware of how the content is being received by other audience members. Thus, it stands to reason that if viewers are not able to view feedback from other users – which may be negative or

positive – decreased visibility in relevant content may lead viewers to make attributions about the motives or how authentic the source of the content posted is.

According to Lee (2020), online interaction is a key contributor to perceptions of authenticity in computer-mediated environments. Lee (2020, pp. 63) notes that computer-mediated environments increase opportunities for "vicarious interactions" through the use of online communities, webpage comment sections, and social media platforms. Characterized as a type of parasocial experience or interaction, Lee (2020) suggests that similar to imagined interactions in a mass communication context – where media audiences feel a reciprocal relationship or pseudo-friendship with media figures without direct contact – observation of online exchanges can induce a sense of "real conversational engagement" in viewers (Lee, 2020, pp. 63; Lee & Jang, 2013, Lee & Shin, 2014). Thus, these venues allow people to "unobtrusively observe" other activities online through the erosion of "public and private" (Lee, 2020, pp. 63). As noted by Kreling et al. (2022) a key distinction between Instagram Stories and Instagram Feed Posts is the visibility of interactions between the poster and other users.

In the context of the promotion of prosocial issues or movements, viewers may make attributions about the intent of the content creator when there are perceived limitations of how visible the prosocial content is. As previously mentioned, discussion of social or political commentary with a public audience can be considered risky as it leaves communicators vulnerable to negative feedback or interactions (Mutz, 2006; Thorson, 2013). Users might not only be concerned with self-presentational aspects, but also the social repercussions of their posts (Lane & Cin, 2017; Thorson, 2013). As Instagram Feed Posts offer more visibility with respect to how integrated into the Feed of the platform (-- by default posts are viewable to all

followers), and with respect to the publicness of third-party interactions (-- by default third-party comments and likes are viewable to other users), content creators may be more susceptible not only to their content being viewed third party observers, but also facing potential public feedback from audience members. As previous research suggests that viewers are often skeptical of the extent to which online self-presentations are accurate representations of individuals' offline selves (Hancock & Toma, 2009), it stands to reason that if viewers perceive prosocial content to be less visible, and thus less open to public feedback –which may be negative or positive – viewers may make more negative attributions about genuine motives of the source. As such, I hypothesize the following:

> **H4:** The more prosocial posts are perceived to be visible, the more viewers will rate the sources of the content as **(a)** an authentic ally, as having **(b)** more prosocial reasons for posting and, **(c)** less self-interested reasons for posting.

Comment visibility, warranting, and authenticity

Additionally, while different channel features of social media platforms vary in their degree of visibility by default setting, social media platforms such as Instagram further allow users to individually manage not only content visibility but also to restrict how other social media users can publicly interact with their social media content (Instagram, 2023). In the context of Instagram, Instagram Feed Posts are variable in their degree of visibility due to users' ability to turn off, limit, block, filter and delete comments on posts. For example, if users want to avoid spam, or unfavorable comments on their posts they are able to turn off other users' ability to comment on their posts, and as a result make any original comments invisible. Continuing, instead of turning off comments, users can also temporarily limit comments to restrict comments

from non and recent followers. Instagram also allows profile owners to block comments from specific users making them invisible to other viewers, delete comments, and curate a custom list of words that can be filtered out so that they are invisible to other users. While these features are intended to use in instances of online harassment and spam (Instagram, 2023), given impression management concerns related to online self-representation (Chua & Chang, 2016), there is also the potential for these features to be used instead to buffer users from third-party feedback that may be unflattering. To that end, given that some online displays of care about social issues online are ingenuine, and rather strategic evaluations of social context by individuals to represent themselves in the best light (Abidin, 2016; Jordan & Rand, 2019; Wellman, 2022), it stands to reason that social media users who have insincere or self-enhancing motives may attempt to avoid unflattering feedback from others. Thus, it may be fruitful to understand how viewers interpret the use of comment management features on social media by presumed digital allies, and how this might impact attributions made toward the source of prosocial content.

Warranting theory is a useful framework to explain how source control over third-party interaction and communication visibility impacts viewer attributions. In computer mediated contexts, people rely on online information with greater warranting value (Walther & Parks, 2002); warranting value represents the degree to which individuals perceive the target of information to have manipulated, controlled, or shaped information about themselves (DeAndrea, 2014; Walther & Parks, 2002). The warranting principle proposes that individuals put greater weight on information about a person when it cannot be manipulated by the person it describes (DeAndrea, 2014; Walther et al., 2009). Put another way, the less information is

perceived to be manipulated or controlled by the target of information, the more influence the information about the target has in impression formation.

Warranting theory has garnered both direct and indirect support. Walther et al. (2009), for example, over the course of two experiments, examined the impact of self-generated versus friend-generated (i.e., other-generated) statements made about a target on Facebook wall postings. Study 1 tested perceptions of extroversion, while study 2 tested perceptions of physical attractiveness. Walther et al. (2009) expected that third-party claims made on Facebook, regardless of their valence, would have a higher degree of warranting value attributed to them, and thus be seen as more authentic by viewers. Results suggested a boundary condition; there may be domains with which the warranting principle is heuristically useful. For example, results from study 2, which examined physical attractiveness information, showed that participants perceived greater warranting value in other-generated statements. Physical attractiveness, unlike introversion and extroversion (study 1), has clear social desirability implications. Walther et al. (2009) suggested that perceivers may be more likely to seek third-party verification for claims made that are viewed to benefit the target; if claims do not appear to benefit target, the warranting principle may not hold.

In another study on the warranting principle, DeAndrea and Vendemia (2016) found that people tend to accept positive claims posted on social media about a pharmaceutical company to a greater extent when they believe the company is not associated with the users producing the positive claims. Similarly, DeAndrea et al. (2018) found that audiences believe positive online reviews about a company more when they are hosted on a third-party site as compared to when the same reviews are hosted on the company site. Both works above garner support for the

warranting principle; the more information is perceived to be controlled or manipulated by the target who may benefit from claims made, the less people are influenced by the message. Other works have applied warranting theory in a variety of contexts (DeAndrea & Carpenter, 2018). For example, warranting theory has been applied to understanding social networking sites (e.g., Antheunis & Schouten, 2011), product rating sites (e.g., Flanagin & Metzger, 2013; Willemsen et al., 2012), and determining the authenticity of visual depictions online (Johnson et al., 2015; Vendemia & DeAndrea, 2018) amongst others.

While a large portion of warranting research has focused on how people evaluate the authenticity of third-party claims relative to self-claims, DeAndrea (2014) points out other factors may impact perceptions of warranting value. Specifically, DeAndrea (2014) suggests that control over the dissemination of third-party information by the source may impact viewers perception of warranting value. While people who control participatory websites (e.g., social media platforms, blogs, YouTube channels) might not control what others contribute to their pages, page proprietors do have the ability to control what user contributions are able to remain visible on their webpages. Social media users are able to control their online self-presentations by privatizing their pages, deleting comments, and untagging themselves in undesirable pictures (DeAndrea, 2014). As DeAndrea (2014) points out, peoples' ability to manage, or control, their self-presentation on these platforms leaves room for the restriction of third-party content, which may subsequently impact viewer attributions about the intent of the source.

A study by Waddell (2017) explored this claim by testing how readers of online news reacted to online comments that were moderated by a journalist. Participants were instructed to read a news article accompanied by negative or positive comments that either appeared in the

comment section or were embedded into the article itself. The condition where comments were embedded in the article served to indicate moderation, or comment gatekeeping, by the journalist who wrote the article. Results showed that viewers in the comment gatekeeping condition perceived articles to be less credible and more biased as compared to viewers who viewed the article and accompanying comment section. Waddell (2017) suggests that these findings build on warranting theory as perceptions of journalistic gatekeeping contributed to lower warranting value of comments, reflected by decreased perceptions of source and article credibility. Vendemia and colleagues (2019) took a similar approach examining how viewers evaluated political organizations that strategically hide comments on their political posts. Results showed that viewers had less trust in political messages when they perceived that political organizations were strategically controlling public comments (Vendemia et al., 2019). Finally, another study by Shin and Dai (2022) similarly examined the concept of dissemination control in the context of e-commerce. Specifically, Shin and Dai examined how claims by viewers that sellers delete comments impact viewers perceptions of sellers' perceived control over comments, as well as how perceptions of control influenced product evaluation and purchase intention. Results show that heightened perceptions of seller control led viewers to make negative attributions towards the product and the seller and made them less likely to want to buy the product.

Germane to the present study, while Instagram Feed Posts may lend prosocial messaging greater visibility, content creators are still able to manage not only in who can respond to content they post, but also the visibility of those responses. Instagram allows dissemination control of third-party feedback through comment limitation, allowing page proprietors the ability to exert control over *who* (e.g., which third party users) can comment on their content. Additionally, page

proprietors are able to exert dissemination control by deleting comments, allowing them to restrict the substance of third-party comments that accompany their content. Warranting theory suggests that comments may be perceived as less authentic if viewers believe an individual is selectively controlling their dissemination to curate a desired image (DeAndrea, 2014). Thus, prosocial content creators should be perceived as less authentic the more viewers believe that they are controlling the dissemination of comments. As such, I hypothesize the following:

H5: The more viewers perceive the sources of the prosocial content to be controlling what comments exist on Instagram Feed Posts, the less viewers will rate the sources as **(a)** an authentic ally, having **(b)** less prosocial reasons for posting and having **(c)** more self-interested reasons for posting.

H6: The more viewers perceive the sources of the prosocial content to be controlling who can comment on Instagram Feed Posts, the less viewers will rate the sources as **(a)** an authentic ally, having **(b)** less prosocial reasons for posting and having **(c)** more self-interested reasons for posting.

Chapter 4: Method

Research design overview

A 2 (Message Persistence) x 4 (Comment Type) nested between-subjects experiment was used to investigate the proposed hypotheses. The message persistence factor had two conditions: low persistence (i.e., Instagram Stories) and high persistence (i.e., Instagram Feed Posts). The comment type factor was a nested condition within the high persistence condition. The comment type factor included four conditions: no comments, no comments deleted, comments deleted, and comments limited.

Sample

A sample of 850 women participants were recruited from Connect via CloudResearch. For inclusion in the final analyses, participants needed to (a) identify as a woman, (b) be an adult at least 18 years of age or older, and (c) use Instagram. I removed four participants for failing to meet qualifications for participating in this study, for a final sample of 846 self-identified women. All participants received $2.00 in compensation in exchange for their participation. Participants ranged from age 18 to 78 (M = 38.58, SD = 11.85), and identified as "White/Caucasian" (n = 610), "African American/Black" (n = 106), "American Indian/ Indigenous American" (n = 6), "Asian/Asian American" (n = 68), "Native Hawaiian or Pacific Islander" (n = 1), "Multiracial" (n = 45), and "Other" (n= 10).

Procedure

Procedures were approved by The Ohio State University's Institutional Review Board. Once individuals chose to participate in the study, they were directed to a Qualtrics survey and consent was obtained. Participants were told that they would view 5 Instagram Feed Posts or Stories from several Instagram users who are sharing content in support of women's empowerment. Each participant was randomly assigned to one of five experimental conditions: (a) low persistence, (b) high persistence, no comments, (c) high persistence, no comments deleted, (d) high persistence, comments deleted, and (e) high persistence, comments restricted. After viewing the posts in their randomly assigned condition, participants were instructed to fill out a questionnaire. After submitting their responses, participants were directed back to the original host site for compensation.

Stimuli

Message Persistence. The stimuli for each condition was designed to appear to be actual Instagram Stories or Feed Posts (see Appendix B for examples). To account for confounds related to any endorsement metrics on Instagram Feed Posts, as Instagram Stories did not show metrics, Instagram Feed Posts had information related to the number of likes and identities of users providing feedback omitted. Stimuli were designed to be as equal as possible across conditions and reflect how the platform features naturally appear online.

User attributes. In all conditions the message creator was identified by an assumably masculine username (e.g., username @adam_mack). Additionally, each stimulus image included a profile picture of an individual who physically appears to be male. Images for the profile picture were sourced from The Chicago Face Database which provides high-resolution, standardized photographs of self-identified males and females of various races and ethnicities

between the ages of 18-40 years old (Ma et al., 2015). To heighten the visibility of the identity of the poster, before viewing the stimulus image participants were shown the username and profile photo of the Instagram user.

Prosocial Stimulus. The prosocial component of the stimuli consisted of a message designed as an infographic typical to online activism efforts and an accompanying caption. To achieve this, five prosocial messages were adapted from existing social media content related to the modern-day feminist movement (See Appendix C for stimuli examples). These messages served to indicate to viewers that the posts are prosocial in nature. Scholars have suggested that women and girls and their allies have begun to use social media platforms as an activism tool to address feminist issues through collaboration, personal disclosure, networking, organizing and challenging modern-day sexism, misogyny, and rape culture (Linavary et al., 2019; Mendes et al., 2019). Thus, content created and shared online with regards to feminist issues aims to "expose, critique, and educate the public about sexism and offer counter discourses" to misogyny (Mendes et al., 2019, pp. 3). For the current study, messages created were intended to serve to uplift and empower women (e.g., "Women's Rights are Human Rights").

In addition to viewing the prosocial graphic and message, participants viewed captions accompanying the image. The caption served to explicitly indicate that the user is an ally for women's empowerment (e.g., "I'm always an ally for women. #humanrights #womensrights;" "I'm fighting, are you? #womensequality #fightforwomen"). The prosocial messages and captions used were standardized across all Instagram Story and Instagram Feed Post conditions. To specify, while the Instagram feature differed, the prosocial messages and captions viewed were the same across all conditions.

Comments. All conditions with comments had posts with between 3 and 6 user comments. To achieve more realism, comment numbers varied between posts shown to viewers. Comments were positive in valence and were complementary of the user posting.

Measures

Unless otherwise noted, all items were measured on seven-point scales that ranged from strongly disagree (1) to strongly agree (7).

Visibility. Participants were asked the extent to which they perceive the Instagram content to be visible to others. Visibility was measured using bipolar statements measured along a 7-point semantic differential scale adapted from Monge et al. (2024). The stem read "How many people do you think have seen this post" followed by four statements such as "not that many people – many people," and "a small amount of people – a large amount of people, " (ω= 0.97).

Persistence. Persistence was measured using bipolar statements measured along a 7-point semantic differential scale adapted from Monge et al. (2024). The stem read "The users shared messages that support women in a way that will last…" followed by five statements such as "a short duration of time – a long duration of time" and "an insubstantial amount of time – a substantial amount of time," ($\omega = 0.98$).

Dissemination Control. The degree to which participants believed content was controlled by the source of the content was measured by adapting measures from DeAndrea and Carpenter's (2018) dissemination control scale. In the analyses dissemination control was operationalized in two ways: comment deletion (i.e., the extent to which viewers perceived users to delete comments from others, $\omega = 0.95$) and comment restriction (i.e., the extent to which

viewers perceived users to restrict who was able to comment, ω = 0.94). Examples of items include: “The poster made sure viewers only saw what he wanted viewers to see” and “Only content approved by the poster is viewable to others” (entire scale, ω = 0.96).

Authenticity. Two items were adapted from a previous study (Pittman et al., 2022) and two were developed following conceptual definitions of authenticity. As such, four items measured participants authenticity perceptions of content viewed on Instagram: “I think the Instagram users who posted are genuine,” and “The Instagram users who posted seem to care about women’s empowerment issues” (ω = 0.94).

Prosocial Reasons for Posting. This scale measured the degree to which participants believe content was shared for prosocial reasons. Research suggests that participation in online collective action efforts stems from individuals' use of social media as an information source, a tool for mobilization, and a medium to express their opinions (Hong & Kim, 2021). The stem of this measure stated “The men posted this content…” followed by nine reasons related to sharing on social media as an information source (e.g., to generate awareness*)*, to express opinions (e.g., to show support)*,* and as a tool for mobilization (e.g., to mobilize online support for an issue, ω = 0.91).

Self-interested Reasons for Posting. This scale measured the degree to which participants believed content was shared for self-serving reasons. I adapted items from Vendemia and DeAndrea (2018) and Vendemia et al. (2021). The stem stated “The men posted this content…” followed by nine reasons for sharing related to social feedback (e.g., to gain likes/shares/endorsement), attention-seeking (e.g., to show off), and audience scale/reach (e.g., to gain popularity, ω = 0.94).

Spontaneity. Spontaneity refers to how improvisational or without any planning participants perceive the source of the prosocial content to be. Spontaneity was measured using a semantic differential scale that started with the stem "The men posted these images…" followed by three bipolar adjectives describing participants' perception of the sources sharing habits (e.g., not giving it much thought – with careful deliberation, $\omega = 0.94$).

Commitment. Commitment refers to how dedicated to the cause or issue participants perceived the users to be. Commitment was measured using a semantic differential scale that started with the stem "The men who posted the images are…" followed by three bipolar adjectives capturing participating's perception of the sources commitment to the prosocial issue (e.g., dedicated – not dedicated, $\omega = 0.95$).

Chapter 4: Results

Analysis plan

SPSS Statistics (Version 29) was used for all analyses. The message persistence factor contained two conditions: low persistence (Instagram Stories) and high persistence (Instagram Posts). The comment type factor contained four conditions nested within the Instagram Posts condition (no comments, no comments deleted, comments deleted, and comments restricted). PROCESS (Hayes, 2017) was used to conduct simple mediation analyses (Model 4) for H1, H2, H4, H5, and H6 and serial mediation analyses (Model 6) for H3. For all tests, effect estimates are reported with percentile bootstrap confidence intervals based on 10,000 resamples. For H1-H4, I compared the low persistence condition (Stories) with the high persistence no comments condition to avoid confounds (Stories are not accompanied with comments). The experimental conditions were dummy coded as follows: 1 = low persistence, 2 = high persistence no comments, 3 = high persistence comments, 4 = high persistence comments limited, 5 = high persistence comments deleted.

Persistence

Although the following tests examine differences between the low persistence condition and high persistence no comment condition, comparing the low persistence condition to the other high persistence conditions reflected similar findings. These results can be found in Appendix A. Prior to probing my hypotheses, I conducted a one-way ANOVA to determine if participants meaningfully differentiated between the low persistence and high persistence factor. Results

showed that participants in the low persistence condition, M = 3.69, SD = 2.10, rated the content as less persistent than participants in the high persistence no comment condition, M = 4.75, SD = 1.87, $F(1,333) = 23.76$, $p < .001$, $\eta^2 = .067$, indicating that participants did in fact differentiate between the low persistence and high persistence, no comment conditions. For the following analyses (H1-H2) X is the experimental condition, M is perceived persistence, and Y is the outcome variable. H1 stated that the more persistent prosocial posts are perceived to be, the less viewers would perceive the source of the content as posting spontaneously. Supporting H1, a significant indirect effect was detected in the predicted direction, b = 0.40, (SE = 0.09), 95% CI = [0.22, 0.59]. H2 predicted that the more prosocial posts are perceived to be persistent, the more viewers will perceive the sources of the content as being committed to the prosocial movement or issue. Supporting H2, a significant indirect effect was found in the predicted direction for the outcome of commitment to the prosocial movement or issue, b = 0.30, (SE= 0.07), 95% CI = [0.16, 0.45].

H3 predicted that the more prosocial posts are perceived to be persistent, the more viewers would perceive the sources of the content as being committed to the prosocial movement or issue, and thus would rate the sources as (a) an authentic ally, as having (b) more prosocial reasons for posting and (c) less self-interested reasons for posting. To probe H3 I used PROCESS Model 6 (Hayes, 2017) to test if there was serial mediation. For the outcome of perceptions of authentic allyship (H3a), the serial mediation was significant (experimental condition → persistence → commitment → authentic allyship), b = 0.19, (SE = 0.05), 95% CI = [0.10,0.30]. For the outcome of perceptions of prosocial reasons for posting (H3b), the serial mediation was significant (experimental condition → persistence → commitment → prosocial

reasons for posting), b = 0.14, (SE = 0.03), 95% CI = [0.07, 0.21]. And finally, for the outcome of perceptions of self-interest (H3c), serial mediation was significant (experimental condition → persistence → commitment → perceptions of self-interest), b = -0.12 (SE=0.03), 95% CI = [-0.20, -0.06].

Visibility

To probe H4, I again used Model 4 in the PROCESS macro (Hayes, 2017) to estimate the indirect effect of the persistence factor (X) on the outcome variables (Y) through perceptions of visibility (M). H4 predicted that the more prosocial posts are perceived to be visible, the more viewers would rate the sources of the content as (a) an authentic ally, as having (b) more prosocial reasons for posting and, (c) less self-interested reasons for posting. First, I examined whether the low persistence condition was viewed as less visible than the high persistence no comment condition. The results indicated a significant difference but in the opposite direction as predicted, b = -0.49 (.18), p = 0.007. Participants perceived the high persistence no comment post to be significantly less visible (M = 3.43, SD = 1.68) than the low persistence post (M = 3.92, SD = 1.56). Nonetheless, given the significant difference in visibility perceptions across conditions, I ran mediation analyses to examine if greater visibility influenced the outcome measures as predicted. The mediation analyses indicated that greater visibility led to higher perceptions that sources were authentic allies (H4a), b = -0.14 (SE = 0.05), 95% CI = [-0.26, -0.04], greater perceptions that the user had prosocial reasons for posting (H4b), B = -0.09 (SE = 0.03), 95% CI = [-0.16, -0.02], and less self-interested reasons for posting, b = 0.04 (SE = 0.02), 95% CI = [0.01., 0.10]. To summarize, the effect of the manipulation (low persistence condition

vs. high persistence, no comment) on visibility was significant in the opposing direction but the predicted effects of greater visibility held for each outcome.

A post hoc test was conducted using PROCESS Model 4 (Hayes, 2017) to compare the indirect effect of the comment type factor (X) on the outcome variables (Y) through perceptions of visibility (M). The post hoc tests were conducted given the possibility that the absence of comments signaled lower viewership and allowed to further test if posts with lower or greater visibility influenced the outcome variables as anticipated. Although the following details results between the high persistence no comments condition and high persistence comments condition, the other high-persistence comment conditions reflected similar results across all outcomes. Full results of the post hoc tests can be found in Appendix A. For the perceptions of users being authentic allies, results of the post hoc test showed a significant indirect effect, b = 0.17 (SE = 0.05), 95% CI = [0.07, 0.27]. Relative to the high persistence no comments condition, participants in the high-persistence comments condition perceived posts to be more visible, and thus considered the source of the content as an authentic ally. For the outcome of prosocial reasons for posting, results showed a significant indirect effect, b = 0.10 (SE = 0.03), 95% CI = [0.04, 0.17]. Relative to the high persistence no comments condition, participants in the high persistence comments condition perceived posts to be more visible, and thus considered the source of the content to have more prosocial reasons for posting. Finally, for the outcome of self-interest, results showed a significant indirect effect, b = -0.04 (SE = 0.02), 95% CI = [-0.09, -0.002]. Relative to the high persistence no comments condition, participants in the high-persistence comments condition perceived posts to be more visible, and thus considered the

source of the content to have less self-interested reasons for posting. To that end, results of the post hoc test suggest partial confirmation of H4.

Dissemination control

Finally, to probe H5 and H6, Model 4 in the PROCESS macro was used (Hayes, 2017). For H5 I estimated the indirect effect of the comment type factor (X) on the outcome variables (Y) through perceptions of comment deletion (M). For H6 I estimated the indirect effect of the comment type factor (X) on the outcome variables (Y) through perceptions of commenter restriction. H5 compared the comments condition and the comments deleted condition whereas H6 compared the comments condition and the comments restricted condition. H5 predicted that the more viewers perceive the sources of the prosocial content to be controlling what comments exist on Instagram Feed Posts, the less viewers would rate the sources as (a) an authentic ally, having (b) less prosocial reasons for posting and having (c) more self-interested reasons for posting. Supporting H5a and H5c, significant indirect effects were found in the predicted direction for the outcome of perceptions of authentic allyship, b = -0.09 (SE = 0.03), 95% CI = [-0.17, -0.03] and perceptions of self-interest, b = 0.18 (SE = 0.03), 95% CI = [0.10, 0.26]. However, the results indicated that there was no significant indirect effect of the comment type factor on perceptions of prosocial reasons for posting (H5b), b = -0.04, (SE = 0.04), 95% CI = [-0.09, 0.01], leaving H5 partially supported.

H6 predicted that the more viewers perceive the sources of the prosocial content to be controlling who can comment on Instagram Feed Posts, the less viewers would rate the sources as (a) an authentic ally, having (b) less prosocial reasons for posting and having (c) more self-interested reasons for posting. Results indicated that there was no significant indirect effect of

the comment type factor on perceptions of authentic allyship (H6a), b = -0.14, (SE = 0.07), 95% CI = [-0.30, 0.01], or perceptions of prosocial reasons for posting (H6b), b = -0.04, (SE = 0.05), 95% CI = [-0.15, 0.06]. However, supporting H6c, significant indirect effects were found in the predicted direction for the outcome of perceptions of self-interest, b = 0.25 (SE = 0.09), 95% CI = [0.07, 0.46], leaving H6 partially supported.

Chapter 5: Discussion

The current study examined how perceived affordances—persistence and visibility—in addition to comment dissemination control, influenced how women evaluated the authenticity of men who posted allyship messages on Instagram. My results provide evidence that the more prosocial content is perceived to be persistent or visible, the more positively it impacts viewer attributions towards the motives and authenticity of presumed digital allies. Notably, a serial mediation analysis indicated that the more viewers perceived posts to be persistent, the more committed they viewed the source of the content to be, and consequently the more they saw the source as an authentic ally, having prosocial reasons for posting, and being less self-interested. Additionally, the results suggest that perceived comment deletion and comment restriction (i.e., limiting who can comment) can also impact viewer authenticity attributions under certain conditions. Results showed that the more viewers believed users were deleting comments, the less authentic and more self-interested they found the purported allies to be. Likewise, the more viewers thought users restricted who could comment on their prosocial content, the more they thought the users had self-interested motives for sharing prosocial content.

Theoretical and Practical Implications

Although there is a growing body of work examining the use of social media platforms such as Instagram, X/Twitter and TikTok and their impact on digital allyship efforts and prosocial movements, few have taken an affordance based approach to this subject. Adopting an affordance approach to research on digital allyship is imperative, as it allows scholars to examine

the relationship between people, technology, and context. To that end, a central aim of this work was to understand the social and psychological impact that perceived affordances have on viewer attributions towards prosocial content posted by digital allies on social media. Overall, my findings suggest that differences in perceived affordances do in fact impact viewer attributions towards the authenticity and motives of sources of prosocial content.

My results suggest that the more persistent channel features are on Instagram, the more members of disadvantaged groups (i.e., women) perceive digital allies (i.e., men) as putting in more care or deliberation towards posting. Existing research suggests that there are norms related to users' choice in the use of features such as Posts, which afford more persistence, and Stories, which are less persistent (i.e., ephemeral). On the one hand, previous research suggests that behaviors like sharing Posts on Instagram are associated with greater selective self-presentation and effort to manage audience impressions (Chua & Chang, 2016; Huang & Vitak, 2022; Yau & Reich, 2018). On the other hand, research on ephemeral elements of social media platforms, such as Snapchat and Instagram Stories, have suggested that these tools are associated with a reduction in users' impression management concerns due to platform specific norms. Users tend to use these features to communicate in a spontaneous and casual manner about life in the moment, or the mundane (Bayer et al., 2016; McRoberts et al., 2017; Xu et al., 2016). To that end, the results suggest that these norms may apply to the use of these features to share prosocial messaging. The use of Instagram Stories, as compared to Instagram Posts, garnered more negative attributions from viewers about the level of spontaneity (e.g., less careful planning put into sharing prosocial messaging). This may be due to a perceived violation of platform norms of use. As low persistence, or ephemeral media is not typically associated with deliberate curation

and attention to presentation as high-persistence media, it stands to reason that digital allies' use of the ephemeral element signaled to viewers that users are treating prosocial content casually or sharing content without much thought.

Likewise, I detected a significant effect between persistence, commitment, and the authenticity and motive outcomes. The more viewers perceived content to be persistent, the more they viewed digital allies as being committed to the prosocial cause, leading them to believe users were more authentic allies, had more prosocial reasons for posting, and were less self-interested. These findings might speak to an underlying connection between commitment, authenticity, and lay understanding of allyship. Contemporary conceptualizations of allyship suggest that the behavior involves members of advantaged groups taking on a degree of interpersonal risk (i.e., loss of power, privilege, or social capital) in challenging the status quo for the betterment of disadvantaged groups (Erskine & Bilimoria, 2019; Pyram, 2023; Salter & Migliaccio, 2019; Thai & Nylund, 2023). Continuing, research suggests that powerful individuals are granted more respect when they are perceived to be acting prosocially and not for self-enhancement (Park et al., 2022; Willer et al., 2012). Additional research on allyship and authenticity has found that advantaged group members are viewed as having more genuine motives if they are perceived to face more costs, as compared to rewards, in helping (Thai & Nylund, 2023). To that end, an ally's perceived commitment to the prosocial issue, in spite of the costs or obstacles—such as backlash from trolls or adversarial social media users—may signal to disadvantaged group members that their investment is long-term and may subsequently lead to more favorable attributions. Although digital allyship efforts may be low in physical or monetary cost, individuals who choose to speak out about social or political issues online face

the potential for audience related disagreement, (critical) feedback, and social repercussions (Thorson, 2014; Wellman, 2022). Unlike content shared on ephemeral elements of platforms, prosocial content that is shared using a high-persistence feature, such as Instagram Posts, affords a public and remaining archive of the users' display of solidarity. The prosocial content remains on digital allies' profiles, showing not only permeant commitment to integrate digital allyship efforts into their profile self-presentation, but also leaving them suspectable to third-party views and feedback—ultimately signaling to viewers that users have genuine motives for posting.

Relatedly, the findings show a similar effect of perceived visibility on the authenticity and motive outcomes. What would be seen as more or less visible was not consistent with initial expectations. It was expected that Stories would be viewed as less visible than Posts. The results indicated that Posts without comments were viewed as less visible than Stories or Posts with comments. Nonetheless, the effect of greater visibility on the predicted outcomes held as anticipated. Similar to comparing Posts without comments to Stories, the post hoc analyses showed that the presence of comments positively influenced visibility perceptions. Increased visibility perceptions ultimately led viewers to rate users as more authentic allies, having more prosocial reasons for posting, and being less self-interested. Like the role of persistence, the heightened visibility that comments afford Posts may cue viewers to consider not only the risk posters take in sharing, but to also consider digital allies' openness to (critical) feedback, leading to more positive attributions. To further contextualize my findings, Knowlton et al.'s (2023) relational perspective on effective allyship can be applied. Knowlton et al. (2023) suggests that allies are most effective and cultivate more trust between themselves and disadvantaged group members through exhibiting *behavioral humility*. Behavioral humility is described as an

interpersonal characteristic rooted in social contexts that exemplify an individuals' (a) willingness to view oneself accurately through seeing one's own short comings and blind spots, (b) displayed appreciation for others' strengths and contributions, and (c) desire to learn from others (Owens et al., 2013). Knowlton and colleagues (2023) suggest that one way allies can achieve behavioral humility is through seeking feedback prior to and after helping disadvantaged group members. By doing so, allies show that they are self-reflective, interested in how their behavior impacts disadvantaged group members, and not just how their behaviors appear to others (Knowlton et al., 2023). In the context of digital allyship, individuals' willingness to receive or display feedback may positively lend to viewer attributions of their genuineness as they may be susceptible to both (critical) feedback and dissent from both advantaged in group members, and disadvantaged group members. Finally, the results provide confirmation to the role of source control over the dissemination of third-party information and its impact on viewer attributions about authenticity. Warranting theory proposes that online viewers are more likely to believe information genuinely reflects an offline reality (i.e., is authentic) when they perceive the information is not controlled or strategically modified by the entity it describes (DeAndrea, 2014; Walther & Parks, 2002). The results of this experiment provide support to DeAndrea (2014)'s proposed extension of warranting theory which suggests that perceived source control over third-party information can impact viewer perceptions of warranting value. On Instagram in particular, users have the functionality to delete comments from others, and to restrict, or "limit" what other users can comment on their Posts, typically to protect users from online hate and harassment. My results indicated that when viewers perceived comments to be deleted, they perceived the source as less of an authentic ally, and as having more self-interested reasons for

posting. Additionally, when viewers perceived comments to be restricted, they perceived the source as having more self-interested reasons for posting. As previous work suggests that viewers are often skeptical of the authenticity of others online self-presentations (Hancock & Toma, 2009), in the context of this study, viewers may perceive the presumed control of audience feedback as a signal that users' may be concealing information. Viewers might muse users are deleting comments or restricting who can comment because feedback was unfavorable to users' self-presentation goals. Notably, my results showed that perceptions of comment restriction only impacted how self-interested viewers perceived digital allies to be. This may suggest that while comment restriction may negatively impact audience evaluations, in comparison to comment deletion, which completely removes evidence of commentary, comment restriction may be viewed as exerting less manipulation or control over third-party contributions. Future work may seek to further refine the notion of dissemination control by comparing different methods of influencing/manipulating third-party content that might lead to meaningful differences in the perceived warranting value of information and thus its effect on impressions.

The findings also may contribute to establishing boundary conditions for when perceptions of warranting value may have stronger, lesser, or no effect on evaluations. DeAndrea and Vendemia (2019) contend that the warranting principle holds the strongest when sources are viewed as directly benefiting from the claims they make or exert control over. On the one hand, the findings can be interpreted as consistent with that theoretical assertion; viewers believe that some users seek to gain social cache on Instagram by claiming to be allies and the strategic control of feedback—which might suggest otherwise—invites skepticism regarding their authenticity. On the other hand, viewers who might be less inclined to question the sincerity of

online allies might have viewed the restriction of comments in a positive light, as a way of ensuring supportive messages are broadcast without the accompaniment of alternative views or toxic messages from trolls/bots. Future work should continue to explore these dynamics to better understand how different ways of strategically controlling third-party content can hinder or bolster the authenticity of sources and their messages.

With the rise of prosocial movements on social media platforms such as Instagram, X/Twitter, Tik Tok and even Snapchat, my work may provide practical implication to those interested in participating in digital activism efforts. The user-generated nature of social media platforms and prosocial movement efforts allows users great control over the content they want to share about the issues they are passionate about. However, with increased user control comes the possibility that those hoping to participate for self-enhancement (i.e., performative allies) may negatively impact genuine initiatives concerning disadvantaged groups. Although other studies (Abidin, 2022; Wellman, 2022) have discussed how bad actors may exploit features of social media platforms to enhance their self-image via prosocial messaging—this study elucidates specifically how perceived affordances of platform features signal different meaning to audience members about the intent of the prosocial content creator.

To that end, individuals interested in participating in digital activism should mind what they post but what features they use to promote prosocial content, given what their choice might signal to viewers. My results suggest that using features that are perceived to be more persistent, such as Instagram Posts, lend to more favorable viewer attributions about content creators' commitment to the prosocial issue, and genuine interest in allyship. Although creating and

reposting content on low-persistent features such as Stories may be faster and more convenient to share, less positive viewer attributions may ensue.

Additionally, social media users should be wary about the use of comment restriction features and comment deletion in tandem with posting prosocial content. Utilization of these tools may signal to viewers that users may have some interest in maintaining their positive self-presentation, rather than facilitating feedback and engagement. To further bridge the gap between prospective allies and disadvantaged group members, applying frameworks such as Knowlton et al.'s (2023) relational perspective on allyship to digital contexts should be explored. By opening themselves to feedback from their audiences, digital allies may be able to ultimately increase how genuine disadvantaged individuals perceive them, increasing trust and perhaps the effectiveness of their efforts.

Limitations and Future Directions

Although my study provides useful findings for scholars and practitioners, several areas for future work should be considered. First, my experiment examined one specific prosocial issue. Much work on prosocial movements and allyship on social media have been about just that, a movement. My study examined digital allyship surrounding broad topics related to women's' empowerment, not tethered to a catalytic incident in time. Future work should consider examining prosocial messaging related to a distinct movement with more specific goals.

Related to affordances, my study compared the effects of just two perceived affordances—persistence and visibility. As highlighted by Fox and McEwan (2017), there are several commonly cited affordances that exist in computer-mediated environments. With the continued use of social media platforms to promote prosocial issues and causes, additional

experimental work is needed to examine the impact of other perceived affordances on viewer perceptions of digital allies. Moreover, as research has shown that individual goals and perceived social norms drive platform choice in younger users (Taber et al., 2023), additional work should attempt to contextualize these findings to other platforms where users tend to share prosocial content. Understanding how norms of use and affordances intersect should be examined to optimize opportunities for digital allies to effectively share prosocial content.

Another limitation of this work is related to the artificial, or controlled nature of the experiment. Although the dissemination control inductions were successful, the comment deletion condition was not natural to Instagram. In order to induce perceptions of comment deletion, I created stimuli that mimicked the "hidden comment" cue on Instagram. Future work should seek to understand what cues natural to social media environments signal viewers to suspect that the page proprietor removed third-party information. Relatedly, whereas the appearance of Posts in the high-persistence condition were natural to how they appear on Instagram, ephemeral Stories were not. As pointed out by McRoberts et al. (2017) in their study, Snapchat Stories and other ephemeral features of social media platforms, like Instagram Stories, are functional to users who self-select to view content. My experiment induction replicated design attributes of Instagram Stories, but not the interactive ability that the platform affords users, essentially forcing participants to *opt-in* to typically *optional* content. Additionally, Instagram Stories offer many more opportunities for engagement than what was facilitated by my survey experiment. Users can dialog one on one with page proprietors and can send a variety of reactions in low visibility, direct message chats. To further probe the social and psychological impact of ephemeral elements of social media platforms, perhaps a more diverse set of

methodologies should be employed. Ecological Momentary Assessment (EMA) may be a useful tool to understand users' engagement with prosocial messaging in their natural environments. EMA is a research methodology involving real-time repeated sampling of individuals' thoughts, feelings, behaviors, and experiences in their natural environments through their daily lives (Shiffman et al., 2008). As the proliferation of prosocial content within the context of a movement is vast and diverse, EMA methodology, if timed correctly, might be fruitful in understanding how in real time users sift between genuine prosocial content, and performative messaging. As social media technologies allow individuals access to a vast array of content anywhere, anytime, continuing to understand how patterns of use and exposure can impact individuals is necessary to understand the evolving impact of social media.

Conclusion

The goal of this work was to investigate how differences in the perceived persistence and visibility of two social media features native to Instagram—Stories vs. Posts—as well as perceived source control over audience feedback impacted viewer perceptions of digital allies' authenticity. Existing work has alluded to a growing problem related to the presence of self-promotion and prosocial content online, and the potential for it to negatively impact the efficacy and perceived authenticity of these movements (Wellman, 2022). Using an affordance approach, my experiment demonstrates how variability in the persistence and visibility of prosocial messaging can lead to content being perceived differently: greater persistence and visibility can positively contribute to viewer perceptions that allies are genuine, posting for prosocial reasons and less self-interested. Also, using features that let users restrict or delete comments can negatively impact attributions towards the source's genuineness. These findings may be

applicable to those interested in participating in digital allyship or considered when examining the impact of social media and prosocial movements.

References

Abidin, C. (2016). Visibility labour: Engaging with Influencers' fashion brands and# OOTD advertorial campaigns on Instagram. *Media International Australia*, *161*(1), 86-100. https://doi.org/10.1177/1329878X16665177

Abidin, C. (2022). Grief hypejacking: Influencers,# ThoughtsAndPrayers, and the commodification of grief on Instagram. *The Information Society*, *38*(3), 174-187. https://doi.org/10.1080/01972243.2022.2071212

Altman, I. (1975). The environment and social behavior: privacy, personal space, territory, and crowding.

Antheunis, M. L., & Schouten, A. P. (2011). The effects of other-generated and system-generated cues on adolescents' perceived attractiveness on social network sites. *Journal of Computer-Mediated Communication*, *16*(3), 391-406, https://doi.org/10.1111/j.1083-6101.2011.01545.x

Baker, N., Ferszt, G., & Breines, J. G. (2019). A qualitative study exploring female college students' Instagram use and body image. *Cyberpsychology, behavior, and social networking*, *22*(4), 277-282. https://doi.org/10.1089/cyber.2018.0420

Batson, C. D., & Powell, A. A. (2003). Handbook of psychology.

Bayer, J. B., Ellison, N. B., Schoenebeck, S. Y., & Falk, E. B. (2016). Sharing the small moments: ephemeral social interaction on Snapchat. *Information, communication & society*, *19*(7), 956-977. https://doi.org/10.1080/1369118X.2015.1084349

Bayer, J. B., Triệu, P., & Ellison, N. B. (2020). Social media elements, ecologies, and effects. *Annual review of psychology*, *71*, 471-497. ht https://doi.org/10.1146/annurev-psych-010419-050944tps://doi.org/10.1146/annurev-psych-010419-050944

Beer, J. S., & Brandler, S. (2021). Why don't we know more about the minds of authentic people?. *Social and Personality Psychology Compass*, *15*(10), e12637. https://doi.org/10.1111/spc3.12637

Berinsky, A. J., Margolis, M. F., & Sances, M. W. (2014). Separating the shirkers from the workers? Making sure respondents pay attention on self-administered surveys. *American journal of political science*, *58*(3), 739-753. https://doi.org/10.1111/ajps.12081

Berman, J. Z., & Silver, I. (2022). Prosocial behavior and reputation: When does doing good lead to looking good?. *Current opinion in psychology*, *43*, 102-107. https://doi.org/10.1016/j.copsyc.2021.06.021

Chen, J. M., Joel, S., & Castro Lingl, D. (2023). Antecedents and consequences of LGBT individuals' perceptions of straight allyship. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/pspi0000422

Choi, Y. H., & Bazarova, N. N. (2015). Self-disclosure characteristics and motivations in social media: Extending the functional model to multiple social network sites. *Human Communication Research*, *41*(4), 480-500. https://doi.org/10.1111/hcre.12053

Choi, J., & Seo, S. (2017). Goodwill intended for whom? Examining factors influencing conspicuous prosocial behavior on social media. *International journal of hospitality management*, *60*, 23-32. https://doi.org/10.1016/j.ijhm.2016.09.014

Choi, W., & Stvilia, B. (2015). Web credibility assessment: Conceptualization, operationalization, variability, and models. *Journal of the Association for Information Science and Technology*, *66*(12), 2399-2414. https://doi.org/10.1002/asi.23543

Choi, S., Williams, D., & Kim, H. (2020). A snap of your true self: How self-presentation and temporal affordance influence self-concept on social media. *New Media & Society*, https://doi.org/10.1177/1461444820977199

Chua, T. H. H., & Chang, L. (2016). Follow me and like my beautiful selfies: Singapore teenage girls' engagement in self-presentation and peer comparison on social media. *Computers in Human Behavior*, *55*, 190-197. https://doi.org/10.1016/j.chb.2015.09.011

Collier-Spruel, L. A., & Ryan, A. M. (2022). Are All Allyship Attempts Helpful? An Investigation of Effective and Ineffective Allyship. *Journal of Business and Psychology*, 1-26. https://doi.org/10.1007/s10869-022-09861-9

Cornet, V. P., Hall, N., Cafaro, F., & Brady, E. (2017). How image-based social media websites support social movements. *Proceedings of the 2017 CHI conference extended abstracts on Human Factors in Computing Systems* (pp. 2473–2479). https://doi.org/10.1145/3027063.3053257

Cox, J., Nguyen, T., Thorpe, A., Ishizaka, A., Chakhar, S., & Meech, L. (2018). Being seen to care: The relationship between self-presentation and contributions to online pro-social crowdfunding campaigns. *Computers in Human Behavior*, *83*, 45-55. https://doi.org/10.1016/j.chb.2018.01.014

Cranton, P., & Carusetta, E. (2004). Perspectives on authenticity in teaching. *Adult Education Quarterly*, *55*(1), 5-22. https://doi.org/10.1177/0741713604268894

DeAndrea, D. C. (2012). Participatory social media and the evaluation of online behavior. *Human Communication Research*, *38*(4), 510-528. https://doi.org/10.1111/j.1468-2958.2012.01435.x

DeAndrea, D. C. (2014). Advancing warranting theory. *Communication Theory*, *24*(2), 186-204. 10.1111/comt.12033

DeAndrea, D. C., & Carpenter, C. J. (2018). Measuring the construct of warranting value and testing warranting theory. *Communication Research*, *45*(8), 1193-1215. https://doi.org/10.1177/0093650216644022

DeAndrea, D. C., Van Der Heide, B., Vendemia, M. A., & Vang, M. H. (2018). How people evaluate online reviews. *Communication Research*, *45*(5), 719-736. 10.1177/0093650215573862

DeAndrea, D. C., & Vendemia, M. A. (2016). How affiliation disclosure and control over user-generated comments affects consumer health knowledge and behavior: a randomized controlled experiment of pharmaceutical direct-to-consumer advertising on social media. *Journal of Medical Internet Research*, *18*(7), e5972. 10.2196/jmir.5972

DeVito, M. A., Birnholtz, J., & Hancock, J. T. (2017, February). Platforms, people, and perception: Using affordances to understand self-presentation on social media. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing* (pp. 740-754). https://doi.org/10.1145/2998181.2998192

DiMicco, J. M., Geyer, W., Millen, D. R., Dugan, C., & Brownholtz, B. (2009, January). People sensemaking and relationship building on an enterprise social network site. In *2009 42nd*

Hawaii International Conference on System Sciences (pp. 1-10). IEEE. 10.1109/HICSS.2009.343

Duffy, B. E., & Hund, E. (2019). Gendered visibility on social media: Navigating Instagram's authenticity bind. *International Journal of Communication*, *13*, 20. 932–8036/20190005

Einwohner, R. L., & Rochford, E. (2019, December). After the March: Using Instagram to perform and sustain the Women's March. In *Sociological Forum* (Vol. 34, pp. 1090-1111). https://doi.org/10.1111/socf.12542

Ellison, N., Heino, R., & Gibbs, J. (2006). Managing impressions online: Self-presentation processes in the online dating environment. *Journal of computer-mediated communication*, *11*(2), 415-441. https://doi.org/10.1111/j.1083-6101.2006.00020.x

Enli, G. (2015). " Trust Me, I Am Authentic!": Authenticity Illusions in Social Media Politics. In *The Routledge companion to social media and politics* (pp. 121-136). Routledge.

Erickson, T., & Kellogg, W. A. (2000). Social translucence: an approach to designing systems that support social processes. *ACM transactions on computer-human interaction (TOCHI)*, *7*(1), 59-83. https://doi.org/10.1145/344949.345004

Erskine, S. E., & Bilimoria, D. (2019). White allyship of Afro-Diasporic women in the workplace: A transformative strategy for organizational change. *Journal of Leadership & Organizational Studies*, *26*(3), 319-338. https://doi.org/10.1177/1548051819848993

Evans, S. K., Pearce, K. E., Vitak, J., & Treem, J. W. (2017). Explicating affordances: A conceptual framework for understanding affordances in communication research. *Journal of computer-mediated communication*, *22*(1), 35-52. https://doi.org/10.1111/jcc4.12180

Flanagin, A. J., & Metzger, M. J. (2000). Perceptions of Internet information credibility. *Journalism & mass communication quarterly*, *77*(3), 515-540. https://doi.org/10.1177/107769900007700304

Flanagin, A. J., & Metzger, M. J. (2020). Source Credibility. *The International Encyclopedia of Media Psychology*, 1-5. https://doi.org/10.1002/9781119011071.iemp0289

Fogg, B. J., Soohoo, C., Danielson, D. R., Marable, L., Stanford, J., & Tauber, E. R. (2003, June). How do users evaluate the credibility of Web sites? A study with over 2,500 participants. In *Proceedings of the 2003 conference on Designing for user experiences* (pp. 1-15). https://dl.acm.org/doi/abs/10.1145/997078.997097?casa_token=XPrpDqhVHoIAAAAA:iUIXtuO3xb_SSyDNChQ0uDMYuOffunH8VeuhupXtFu1lPBK2bx6A7kWgtr1dFOkq6oDjbf2DfoQ

Fox, J., & McEwan, B. (2017). Distinguishing technologies for social interaction: The perceived social affordances of communication channels scale. *Communication Monographs*, *84*(3), 298-318. https://doi.org/10.1080/03637751.2017.1332418

Gaver, W. W. (1991, March). Technology affordances. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 79-84). https://dl.acm.org/doi/pdf/10.1145/108844.108856

Gibson, J. J. (2014). The theory of affordances:(1979). In *The people, place, and space reader* (pp. 56-60). Routledge.

Gibson, J. J. (1979/1986). *The ecological approach to perception.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Goffman, E. (2002). The presentation of self in everyday life. 1959. *Garden City, NY*, *259*.

Goodman, D. J. (2011). Promoting diversity and social justice: Educating people from privileged groups. Routledge.

Grinberg, N., Kalyanaraman, S., Adamic, L. A., & Naaman, M. (2017, February). Understanding feedback expectations on Facebook. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*(pp. 726-739). https://doi.org/10.1145/2998181.2998320

Grubbs, J. B., Warmke, B., Tosi, J., James, A. S., & Campbell, W. K. (2019). Moral grandstanding in public discourse: Status-seeking motives as a potential explanatory mechanism in predicting conflict. *PloS one*, *14*(10), e0223749. https://doi.org/10.1371/journal.pone.0223749

Hampton, K., & Wellman, B. (2003). Neighboring in Netville: How the Internet supports community and social capital in a wired suburb. *City & Community*, *2*(4), 277-311. https://doi.org/10.1046/j.1535-6841.2003.00057.x

Hancock, J.T., & Toma, C.L., (2009) Putting your best face forward: the accuracy of online dating photographs. Journal of Communication, 59(2), 367-386. https://doi.org/10.1111/j.1460-2466.2009.01420.x

Harter, S. (2002). Authenticity. In C. R. Snyder & S. J. Lopez (Eds.), *Handbook of positive psychology* (pp. 382–394). Oxford University Press.

Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford publications.

Hofstetter, R., Rüppell, R., & John, L. K. (2017). Temporary sharing prompts unrestrained disclosures that leave lasting negative impressions. *Proceedings of the National Academy of Sciences*, *114*(45), 11902-11907. https://doi.org/10.1073/pnas.1706913114

Hollenbaugh, E. E. (2021). Self-presentation in social media: Review and research opportunities. *Review of Communication Research*, *9*, 80-98. https://www.rcommunicationr.org/index.php/rcr/article/view/71

Hong, H., & Kim, Y. (2021). What makes people engage in civic activism on social media?. *Online Information Review*, *45*(3), 562-576. https://doi.org/10.1108/OIR-03-2020-0105

Hovland, C. I., Janis, I. L., & Kelley, H. H. (1953). Communication and persuasion.

Huang, X., & Vitak, J. (2022). " Finsta gets all my bad pictures": Instagram Users' Self-Presentation Across Finsta and Rinsta Accounts. *Proceedings of the ACM on Human-Computer Interaction*, *6*(CSCW1), 1-25. https://doi.org/10.1145/3512916

Instagram (2023a). Delete a comment on Instagram. *Instagram.* https://help.instagram.com/289098941190483

Instagram (2023b). Hide comments or message requests you don’t want to see on Instagram. *Instagram.* https://help.instagram.com/700284123459336/?cms_platform=iphone-app&helpref=platform_switcher

Instagram (2023c). Temporarily limit someone from commenting on my posts. *Instagram.* https://help.instagram.com/4106887762741654

Instagram (2023d). Turn comments on or off Instagram posts. *Instagram.* https://help.instagram.com/1766818986917552/?cms_platform=iphone-app&helpref=platform_switcher

Jackson, S. J., Bailey, M., & Welles, B. F. (2020). *#HashtagActivism: Networks of race and gender justice*. Mit Press.

Johnson, B. K., Vang, M. H., & Van Der Heide, B. (2015). Show me the goods: The warranting effect of user-generated photographs in online auctions. *Journal of Media Psychology: Theories, Methods, and Applications*, *27*(1), 3. https://doi.org/10.1027/1864-1105/a000126

Jordan, J. J., & Rand, D. G. (2020). Signaling when no one is watching: A reputation heuristics account of outrage and punishment in one-shot anonymous interactions. *Journal of personality and social psychology*, *118*(1), 57. https://doi.org/10.1037/pspi0000186

Kalina, P. (2020). Performative allyship. *Technium Soc. Sci. J.*, *11*, 478. https://heinonline.org/HOL/Page?handle=hein.journals/techssj11&div=45&g_sent=1&casa_token=w9C6KtKqiQ8AAAAA:WzdKjjju4Dlh4D0ujghuGRaB0mcUF9qXhkzQOHFvQq5k0VBNs_SzsL_cYhoME8iyOtmhZ0U&collection=journals

Kernis, M. H., & Goldman, B. M. (2006). A multicomponent conceptualization of authenticity: Theory and research. *Advances in experimental social psychology*, *38*, 283-357. https://doi.org/10.1016/S0065-2601(06)38006-9

Konrath, S., Fuhrel-Forbis, A., Lou, A., & Brown, S. (2012). Motives for volunteering are associated with mortality risk in older adults. *Health Psychology*, *31*(1), 87. https://doi.org/10.1037/a0025226

Knoll, M., Meyer, B., Kroemer, N. B., & Schröder-Abé, M. (2015). It takes two to be yourself: An integrated model of authenticity, its measurement, and its relationship to work-related variables. *Journal of Individual Differences*, *36*(1), 38. https://doi.org/10.1027/1614-0001/a000153

Knowlton, K., Carton, A. M., & Grant, A. M. (2023). Help (un) wanted: Why the most powerful allies are the most likely to stumble—and when they fulfill their potential. *Research in Organizational Behavior*, *42*, 100180. https://doi.org/10.1016/j.riob.2022.100180

Kreling, R., Meier, A., & Reinecke, L. (2022). Feeling authentic on social media: Subjective authenticity across Instagram Stories and Posts. *Social Media+ Society*, *8*(1), https://doi.org/10.1177/20563051221086235

Kutlaca, M., Radke, H. R., & Becker, J. C. (2022). The impact of including advantaged groups in collective action against social inequality on politicized identification of observers from disadvantaged and advantaged groups. *Political Psychology*, *43*(2), 297-315. https://doi.org/10.1111/pops.12755

Labroo, A. A., Khan, U., & Su, S. J. (2023). Reconsidering prosocial behavior as intersocial: A literature review and a new perspective. *Consumer psychology review*, *6*(1), 92-108. https://doi.org/10.1002/arcp.1088

Lane, D. S., & Dal Cin, S. (2018). Sharing beyond Slacktivism: the effect of socially observable prosocial media sharing on subsequent offline helping behavior. *Information, Communication & Society*, *21*(11), 1523-1540. https://doi.org/10.1080/1369118X.2017.1340496

Lavertu, L., Marder, B., Erz, A., & Angell, R. (2020). The extended warming effect of social media: Examining whether the cognition of online audiences' offline drives prosocial behavior in 'real life'. *Computers in Human Behavior*, *110*, 106389. https://doi.org/10.1016/j.chb.2020.106389

Leary, M. R., & Kowalski, R. M. (1990). Impression management: A literature review and two-component model. *Psychological bulletin*, *107*(1), 34. https://doi.org/10.1037/0033-2909.107.1.34

Lee, E. J. (2020). Authenticity model of (mass-oriented) computer-mediated communication: Conceptual explorations and testable propositions. *Journal of Computer-Mediated Communication*, *25*(1), 60-73. https://doi.org/10.1093/jcmc/zmz025

Lee, S. S., Liang, F., Hahn, L., Lane, D. S., Weeks, B. E., & Kwak, N. (2021). The impact of social endorsement cues and manipulability concerns on perceptions of news credibility. *Cyberpsychology, Behavior, and Social Networking*, *24*(6), 384-389. https://doi.org/10.1089/cyber.2020.0566

Lee, E.-J., & Jang, J. (2013). Not so imaginary interpersonal contact with public figures on social network sites: How affiliative tendency moderates its effects. Communication Research, 40, 27–51. https://doi.org/10.1177/0093650211431579

Lee, E.-J., & Shin, S. Y. (2014). When the medium is the message: How transportability moderates the effects of politicians' twitter communication. Communication Research, 41, 1088–1110. https://doi.org/10.1177/0093650212466407

Leonardi, P. M. (2011, January). When does technology enable organizational change? Convergent feature use and collective affordances. In *Academy of Management*

Proceedings (Vol. 2011, No. 1, pp. 1-6). Briarcliff Manor, NY 10510: Academy of Management. https://doi.org/10.5465/ambpp.2011.65869196

Leonardi, P. M. (2014). Social media, knowledge sharing, and innovation: Toward a theory of communication visibility. *Information systems research*, *25*(4), 796-816. https://doi.org/10.1287/isre.2014.0536

Li, B., Scott, O. K., Naraine, M. L., & Ruihley, B. J. (2021). Tell me a story: Exploring elite female athletes' self-presentation via an analysis of Instagram stories. *Journal of Interactive Advertising*, *21*(2), 108-120. https://doi.org/10.1080/15252019.2020.1837038

Li, M. (2022). Visual social media and black activism: Exploring how using Instagram influences Black activism orientation and racial identity ideology among Black Americans. *Journalism & Mass Communication Quarterly*, *99*(3), 718-741. https://doi.org/10.1177/10776990221108644

Lin-Healy, F., & Small, D. A. (2012). Cheapened altruism: Discounting personally affected prosocial actors. *Organizational Behavior and Human Decision Processes*, *117*(2), 269-274. https://doi.org/10.1016/j.obhdp.2011.11.006

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior research methods*, *47*, 1122-1135. https://doi.org/10.3758/s13428-014-0532-5

McAndrew, F. T., & Perilloux, C. (2012). The selfish hero: A study of the individual benefits of self-sacrificial prosocial behavior. *Psychological Reports*, *111*(1), 27-43. https://doi.org/10.2466/07.02.09.19.PR0.111.4.27-43

McRoberts, S., Ma, H., Hall, A., & Yarosh, S. (2017, May). Share first, save later: Performance of self through Snapchat stories. In *Proceedings of the 2017 CHI conference on human factors in computing systems* (pp. 6902-6911). https://doi.org/10.1145/3025453.3025771

Mendes, K., & Ringrose, J. (2019). Digital feminist activism:# MeToo and the everyday experiences of challenging rape culture. *# MeToo and the politics of social change*, 37-51. https://doi.org/10.1007/978-3-030-15213-0_3

Merriam Webster Dictionary. (2022). *Authenticity.* https://www.merriam-webster.com/dictionary/authenticity

Merriam Webster Dictionary. (2022). *Credibility.* https://www.merriam-webster.com/dictionary/credibility

Metzger, M. J. (2007). Making sense of credibility on the Web: Models for evaluating online information and recommendations for future research. *Journal of the American society for information science and technology*, *58*(13), 2078-2091. https://doi.org/10.1002/asi.20672

Metzger, M. J., & Flanagin, A. J. (2013). Credibility and trust of information in online environments: The use of cognitive heuristics. *Journal of pragmatics*, *59*, 210-220. https://doi.org/10.1016/j.pragma.2013.07.012

Metzger, M. J., Flanagin, A. J., & Medders, R. B. (2010). Social and heuristic approaches to credibility evaluation online. *Journal of communication*, *60*(3), 413-439. https://doi.org/10.1111/j.1460-2466.2010.01488.x

Monge, C., Matthews, N., & D.C. DeAndrea. (2024, May). The persistence of toxic online messages influences perceptions of harm and attributions of blame. Paper to be presented at the 74th annual International Communication Association, Gold Coast, Australia.

Morozov, E. (2009). Iran: Downside to the" Twitter revolution". *Dissent*, *56*(4), 10-14. https://doi.org/10.1353/dss.0.0092.

Mundt, M., Ross, K., & Burnett, C. M. (2018). Scaling social movements through social media: The case of Black Lives Matter. *Social Media+ Society*, *4*(4), 2056305118807911. https://doi.org/10.1177/2056305118807911

Mutz, D. C. (2006). *Hearing the other side: Deliberative versus participatory democracy*. Cambridge University Press.

Newman, G. E., & Cain, D. M. (2014). Tainted altruism: When doing some good is evaluated as worse than doing no good at all. *Psychological science*, *25*(3), 648-655. https://doi.org/10.1177/0956797613504785

Norman, D. (2013). *The design of everyday things: Revised and expanded edition*. Basic books.

Olivola, C. Y., & Shafir, E. (2018). Blood, sweat, and cheers: The martyrdom effect increases willingness to sponsor others' painful and effortful prosocial acts. http://dx.doi.org/10.2139/ssrn.3101447

Omoto, A. M., & Snyder, M. (1995). Sustained helping without obligation: motivation, longevity of service, and perceived attitude change among AIDS volunteers. *Journal of personality and social psychology*, *68*(4), 671. https://doi.org/10.1037/0022-3514.68.4.671

Owens, B. P., Johnson, M. D., & Mitchell, T. R. (2013). Expressed humility in organizations: Implications for performance, teams, and leadership. *Organization Science*, *24*(5), 1517-1538. tps://doi.org/10.1287/orsc.1120.0795

Oxford Learners Dictionary. (2022). *Authenticity.* https://www.oxfordlearnersdictionaries.com/us/definition/english/credibility?q=credibility

Oxford Learners Dictionary. (2022). *Credibility*. https://www.oxfordlearnersdictionaries.com/us/definition/english/credibility?q=credibility

Papacharissi, Z. (2002). The virtual sphere: The internet as a public sphere. *New media & society*, *4*(1), 9-27. https://doi.org/10.1177/14614440222226244

Park, J. W., Vani, P., Saint-Hilaire, S., & Kraus, M. W. (2022). Disadvantaged group activists' attitudes toward advantaged group allies in social movements. *Journal of Experimental Social Psychology*, *98*, 104226. https://doi.org/10.1016/j.jesp.2021.104226

Picone, I. (2015). Impression management in social media. *The international encyclopedia of digital communication and society*, *36*, 469-476. https://www.researchgate.net/profile/Ike-Picone/publication/314361839_Impression_Management_in_Social_Media/links/59ef8b9c0f7e9baeb26ac3f8/Impression-Management-in-Social-Media.pdf

Pittman, M., Oeldorf-Hirsch, A., & Brannan, A. (2022). Green advertising on social media: Brand authenticity mediates the effect of different appeals on purchase intent and digital

engagement. *Journal of Current Issues & Research in Advertising*, *43*(1), 106-121. https://doi.org/10.1080/10641734.2021.1964655

Pyram, R. H. (2023). *Allyship at Your Own Risk: The Role of Interpersonal Risk in Decisions to Engage in Allyship* (Master's thesis, Michigan State University).http://proxy.lib.ohio-state.edu/login?url=https://www.proquest.com/dissertations-theses/allyship-at-your-own-risk-role-interpersonal/docview/2897151683/se-2

Radke, H. R., Kutlaca, M., Siem, B., Wright, S. C., & Becker, J. C. (2020). Beyond allyship: Motivations for advantaged group members to engage in action for disadvantaged groups. *Personality and Social Psychology Review*, *24*(4), 291-315. https://doi.org/10.1177/108886832091869

Reinecke, L., & Trepte, S. (2014). Authenticity and well-being on social network sites: A two-wave longitudinal study on the effects of online authenticity and the positivity bias in SNS communication. *Computers in Human Behavior*, *30*, 95-102. https://doi.org/10.1016/j.chb.2013.07.030

Riedl, M., Schwemmer, C., Ziewiecki, S., & Ross, L. M. (2021). The rise of political influencers—Perspectives on a trend towards meaningful content. *Frontiers in Communication*, *6*, 752656. https://doi.org/10.3389/fcomm.2021.752656

Roden, J. (2023). *Hesitant Allyship? How Identities and Affordances Shape White People? s Perceptions of Racial Justice Activism on Social Media* (Doctoral dissertation). https://dx.doi.org/10.7302/22293

Salisbury, M., & Pooley, J. D. (2017). The# nofilter self: The contest for authenticity among social networking sites, 2002–2016. *Social sciences*, *6*(1), 10. https://doi.org/10.3390/socsci6010010

Salter, N.P. and Migliaccio, L. (2019), "Allyship as a Diversity and Inclusion Tool in the Workplace", *Diversity within Diversity Management* (*Advanced Series in Management, Vol. 22*), Emerald Publishing Limited, Leeds, pp. 131-152. https://doi.org/10.1108/S1877-636120190000022008

Schlenker, B. R., Britt, T. W., Pennington, J., Murphy, R., & Doherty, K. (1994). The triangle model of responsibility. *Psychological Review, 101*(4), 632–652. https://doi.org/10.1037/0033-295X.101.4.632

Shelby Rosette, A., & Thompson, L. (2005). The camouflage effect: Separating achieved status and unearned privilege in organizations. In *Status and groups* (pp. 259-281). Emerald Group Publishing Limited. https://doi.org/10.1016/S1534-0856(05)07011-8

Shin, S. Y., & Dai, Y. N. (2022). An Online Seller's Dilemma: How a User's Claim About Comment Deletion Affects Product Evaluation and Purchase Intention via Seller Disliking. *International Journal of Communication*, *16*, 24. https://ijoc.org/index.php/ijoc/article/view/16967

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annu. Rev. Clin. Psychol.*, *4*, 1-32. https://doi.org/10.1146/annurev.clinpsy.3.022806.091415

Silver, I., & Silverman, J. (2022). Doing good for (maybe) nothing: How reward uncertainty shapes observer responses to prosocial behavior. *Organizational Behavior and Human Decision Processes*, *168*, https://doi.org/10.1016/j.obhdp.2021.104113

Smith, B. G., Krishna, A., & Al-Sinan, R. (2019). Beyond slacktivism: Examining the entanglement between social media engagement, empowerment, and participation in activism. *International Journal of Strategic Communication*, *13*(3), 182-196. https://doi.org/10.1080/1553118X.2019.1621870

Solomon, J. F., Solomon, A., Joseph, N. L., & Norton, S. D. (2013). Impression management, myth creation and fabrication in private social and environmental reporting: Insights from Erving Goffman. *Accounting, organizations and society*, *38*(3), 195-213. https://doi.org/10.1016/j.aos.2013.01.001

Spielmann, N., Dobscha, S., & Shrum, L. J. (2023). Brands and social justice movements: The effects of true versus performative Allyship on brand evaluation. *Journal of the Association for Consumer Research*, *8*(1), 83-94.

Taber, L., Dominguez, S., & Whittaker, S. (2023). Ignore the Affordances; It's the Social Norms: How Millennials and Gen-Z Think About Where to Make a Post on Social Media. *Proceedings of the ACM on Human-Computer Interaction*, *7*(CSCW2), 1-26. https://doi.org/10.1145/3610102

Tisdell, E. J. (2003). *Exploring spirituality and culture in adult and higher education*. John Wiley & Sons.

Thai, M., & Nylund, J. L. (2024). What are they in it for? Marginalised group members' perceptions of allies differ depending on the costs and rewards associated with their allyship. *British Journal of Social Psychology*, *63*(1), 131-152. https://doi.org/10.1111/bjso.12670

Thorson, K. (2014). Facing an uncertain reception: Young citizens and political interaction on Facebook. *Information, Communication & Society*, *17*(2), 203-216. https://doi.org/10.1080/1369118X.2013.862563

Treem, J. W., & Leonardi, P. M. (2013). Social media use in organizations: Exploring the affordances of visibility, editability, persistence, and association. *Annals of the International Communication Association*, *36*(1), 143-189. https://doi.org/10.1080/23808985.2013.11679130

Treem, J. W., Leonardi, P. M., & Van den Hooff, B. (2020). Computer-mediated communication in the age of communication visibility. *Journal of Computer-Mediated Communication*, *25*(1), 44-59. https://doi.org/10.1093/jcmc/zmz024

Trepte, S., & Reinecke, L. (2011). The social web as a shelter for privacy and authentic living. In *Privacy online: Perspectives on privacy and self-disclosure in the social web* (pp. 61-73). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1016/j.copsyc.2020.04.002

Trieu, P., & Baym, N. K. (2020, April). Private responses for public sharing: understanding self-presentation and relational maintenance via stories in social media. In *proceedings of the 2020 CHI conference on human factors in computing systems*(pp. 1-13). https://doi.org/10.1145/3313831.3376549

Valkenburg, P. M. (2017). Understanding self-effects in social media. *Human communication research*, *43*(4), 477-490. https://doi.org/10.1111/hcre.12113

Vendemia, M. A., & DeAndrea, D. C. (2018). The effects of viewing thin, sexualized selfies on Instagram: Investigating the role of image source and awareness of photo editing practices. *Body image*, *27*, 118-127. https://doi.org/10.1016/j.bodyim.2018.08.013

Vendemia, M. A., DeAndrea, D. C., & Brathwaite, K. N. (2021). Objectifying the body positive movement: The effects of sexualizing and digitally modifying body-positive images on Instagram. *Body Image*, *38*, 137-147. https://doi.org/10.1016/j.bodyim.2021.03.017

Verduyn, P., Gugushvili, N., Massar, K., Täht, K., & Kross, E. (2020). Social comparison on social networking sites. *Current opinion in psychology*, *36*, 32-37. https://doi.org/10.1016/j.copsyc.2020.04.002

Vranken, S., Beullens, K., Geyskens, D., & Matthes, J. (2023). Under the influence of (alcohol) influencers? A qualitative study examining Belgian adolescents' evaluations of alcohol-related Instagram images from influencers. *Journal of Children and Media*, *17*(1), 134-153. https://doi.org/10.1080/17482798.2022.2157457

Waddell, T. F. (2020). The authentic (and angry) audience: How comment authenticity and sentiment impact news evaluation. *Digital Journalism*, *8*(2), 249-266. https://doi.org/10.1080/21670811.2018.1490656

Walther, J. B. (1996). Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction. *Communication research*, *23*(1), 3-43. https://doi.org/10.1177/009365096023001001

Walther, J. B., & Jang, J. W. (2012). Communication processes in participatory websites. *Journal of Computer-Mediated Communication*, *18*(1), 2-15. https://doi.org/10.1111/j.1083-6101.2012.01592.x

Walther, J. B., & Parks, M. R. (2002). Cues filtered out, cues filtered in: Computer-mediated communication and relationships. *Handbook of interpersonal communication*, *3*, 529-563.

Walther, J. B., Van Der Heide, B., Hamel, L. M., & Shulman, H. C. (2009). Self-generated versus other-generated statements and impressions in computer-mediated communication: A test of warranting theory using Facebook. *Communication research*, *36*(2), 229-253. https://doi.org/10.1177/0093650208330251

Wellman, M. L. (2022). Black squares for Black lives? Performative allyship as credibility maintenance for social media influencers on Instagram. *Social Media+ Society*, *8*(1), https://doi.org/10.1177/20563051221080473

Westra, E. (2021). Virtue signaling and moral progress. *Philosophy & Public Affairs*, *49*(2), 156-178. https://doi.org/10.1111/papa.12187

Westerman, D., Spence, P. R., & Van Der Heide, B. (2012). A social network as information: The effect of system generated reports of connectedness on credibility on Twitter. *Computers in Human Behavior*, *28*(1), 199-206. https://doi.org/10.1016/j.chb.2011.09.001

Willemsen, L. M., Neijens, P. C., & Bronner, F. (2012). The ironic effect of source identification on the perceived credibility of online product reviewers. *Journal of Computer-Mediated Communication*, *18*(1), 16-31. https://doi.org/10.1111/j.1083-6101.2012.01598.x

Willer, R., Youngreen, R., Troyer, L., & Lovaglia, M. J. (2012). How do the powerful attain status? The roots of legitimate power inequalities. *Managerial and Decision Economics*, *33*(5-6), 355-367. https://doi.org/10.1002/mde.2554

Wood, A. M., Linley, P. A., Maltby, J., Baliousis, M., & Joseph, S. (2008). The authentic personality: A theoretical and empirical conceptualization and the development of the Authenticity Scale. *Journal of counseling psychology*, *55*(3), 385. https://doi.org/10.1037/0022-0167.55.3.385

Xu, B., Chang, P., Welker, C. L., Bazarova, N. N., & Cosley, D. (2016, February). Automatic archiving versus default deletion: what Snapchat tells us about ephemerality in design. In *Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing* (pp. 1662-1675). https://doi.org/10.1145/2818048.2819948

Yau, J. C., & Reich, S. M. (2018). Are the qualities of adolescents' offline friendships present in digital interactions?. *Adolescent Research Review*, *3*, 339-355. https://doi.org/10.1007/s40894-017-0059-y

Zell, A. L., & Moeller, L. (2018). Are you happy for me … on Facebook? The potential importance of "likes" and comments. *Computers in Human Behavior, 78,* 26–33. https://doi.org/10.1016/j.chb.2017.08.05

Appendix A: Table results of analyses

Table 1:

Descriptive statistics for key dependent variables across all experimental conditions

Variable	Condition	N	Mean	Std. Deviation	Std. Error
Persistence	Low-Persistence	164	3.69	2.10	0.16
	High-Persistence, no comments	171	4.75	1.86	0.14
	High-Persistence, comments	170	4.81	1.70	0.13
	High-Persistence, comments limited	171	4.79	1.67	0.12
	High-Persistence, comments deleted	171	4.85	1.74	0.13
	Total	847	4.58	1.87	0.06
Visibility	Low-Persistence	163	3.91	1.55	0.12
	High-Persistence, no comments	171	3.43	1.68	0.12
	High-Persistence, comments	170	4.02	1.47	0.11
	High-Persistence, comments limited	170	3.91	1.47	0.11
	High-Persistence, comments deleted	171	3.90	1.46	0.11
	Total	846	3.88	1.25	0.04
Comment Deletion	Low-Persistence	164	3.61	1.36	0.10
	High-Persistence, no comments	171	3.39	1.36	0.10
	High-Persistence, comments	169	3.39	1.36	0.10
	High-Persistence, comments limited	171	4.98	1.31	0.10
	High-Persistence, comments deleted	171	4.80	1.49	0.11
	Total	846	4.04	1.54	0.05
Comment Restriction	Low-Persistence	164	3.69	1.33	0.10
	High-Persistence, no comments	171	3.44	1.38	0.10
	High-Persistence, comments	169	3.34	1.38	0.10
	High-Persistence, comments limited	171	4.81	1.31	0.10
	High-Persistence, comments deleted	171	4.29	1.43	0.10
	Total	846	3.92	1.47	0.05

Table 2:

Descriptive statistics for key dependent variables across all experimental conditions

Variable	Condition	N	Mean	Std. Deviation	Std. Error
Authentic Allyship	Low-Persistence	164	5.01	1.24	0.09
	High-Persistence, no comments	171	5.03	1.29	0.09
	High-Persistence, comments	169	5.30	1.18	0.09
	High-Persistence, comments limited	171	5.03	1.29	0.09
	High-Persistence, comments deleted	171	5.10	1.25	0.09
	Total	846	5.10	1.25	0.04
Self - Interest	Low-Persistence	164	1.24	0.09	4.33
	High-Persistence, no comments	171	1.45	0.11	4.29
	High-Persistence, comments	169	1.31	0.10	4.25
	High-Persistence, comments limited	171	1.32	0.10	4.32
	High-Persistence, comments deleted	170	1.26	0.09	4.38
	Total	845	1.32	0.04	4.43
Prosocial Reasons	Low-Persistence	164	5.16	0.90	0.07
	High-Persistence, no comments	171	5.24	0.98	0.07
	High-Persistence, comments	169	5.41	0.82	0.06
	High-Persistence, comments limited	171	5.26	0.96	0.07
	High-Persistence, comments deleted	171	5.40	0.90	0.06
	Total	846	5.29	0.92	0.03
Commitment	Low-Persistence	164	5.49	1.31	.102
	High-Persistence, no comments	171	5.54	1.37	.105
	High-Persistence, comments	169	5.79	1.29	.09
	High-Persistence, comments limited	171	5.46	1.35	.10
	High-Persistence, comments deleted	171	5.62	1.20	.09
	Total	846	5.58	1.31	.04
Spontaneity	Low-Persistence	164	4.57	1.50	0.11
	High-Persistence, no comments	171	4.78	1.59	0.12
	High-Persistence, comments	169	4.91	1.54	0.11
	High-Persistence, comments limited	171	4.73	1.49	0.11
	High-Persistence, comments deleted	171	4.81	1.44	0.11
	Total	846	4.76	1.51	0.05

Table 3:

Correlation matrix for key dependent variables

		1	2	3	4	5	6	7	8	9
Spontaneity	Pearson Correlation	1								
	Sig. (2-tailed)									
	N	846								
Comment Deletion	Pearson Correlation	-0.02	1							
	Sig. (2-tailed)	0.55								
	N	846	846							
Comment Restriction	Pearson Correlation	-0.03	0.86**	1						
	Sig. (2-tailed)	0.26	<.001							
	N	846	846	846						
Persistence	Pearson Correlation	0.45**	0.02	0.01	1					
	Sig. (2-tailed)	<.001	0.39	0.77						
	N	846	846	846	847					
Authentic Allyship	Pearson Correlation	0.61**	-0.08*	-0.11**	0.39**	1				
	Sig. (2-tailed)	<.001	0.01	<.001	<.001					
	N	846	846	846	846	846				
Prosocial Reasons	Pearson Correlation	0.53**	-0.03	-0.06	0.34**	0.68**	1			
	Sig. (2-tailed)	<.001	0.27	.062	<.001	<.001				
	N	846	846	846	846	846	846			
Self-Interest	Pearson Correlation	-0.32**	0.18**	0.18**	-0.21**	-0.49**	-0.34**	1		
	Sig. (2-tailed)	<.001	<.001	<.001	<.001	<.001	<.001			
	N	845	845	845	845	845	845	845		
Commitment	Pearson Correlation	0.64**	-0.08*	-0.09**	0.41**	0.73**	0.68**	-0.42**	1	
	Sig. (2-tailed)	<.001	0.01	0.004	<.001	<.001	<.001	<.001		
	N	846	846	846	846	846	846	845	846	
Visibility	Pearson Correlation	0.38**	-0.01	-0.02	0.35**	0.37**	0.32**	-0.11**	0.36**	1
	Sig. (2-tailed)	<.001	0.75	0.39	<.001	<.001	<.001	<.001	<.001	
	N	844	844	844	845	844	844	843	844	845

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Table 4:

PROCESS Model 4, M = Persistence

Path Estimates	Outcome Variables (Y)	
	Commitment	Spontaneity
	b (*SE*)	*b* (*SE*)
Constant	4.37(0.12)**	4.81 (0.14)**
High Persistence, no comments	-0.27(0.13)*	0.17 (0.15)
High Persistence, comments	-0.03(0.13)	0.07 (0.15)
High Persistence, comments limited	-0.37(0.13)*	0.25 (0.15)
High Persistence, comments deleted	-0.22(0.13)	0.10 (0.15)
Persistence	0.30(0.02)**	-0.37 (0.02)**
F	38.58	12.31
R^2	0.05	0.05
	Indirect Effects [95% CI]	
High persistence, no comments	0.32(0.07) [0.18,0.47]	-0.39 (0.08) [-0.57,-0.23]
High persistence, comments	0.33(0.06) [0.20,0.47]	-0.41 (0.08) [-0.58,-0.25]
High persistence, comments limited	0.33(0.06) 0.20,0.47]	-0.41 (0.08) [-0.58,-0.25]
High persistence, comments deleted	0.33(0.06) [0.21,0.49]	-0.43 (0.08) [-0.60,-0.27]

Note. *$p < .05$. **$p < .01$. Experimental conditions are coded as: 1 = low persistence, 2 = high persistence, no comments, 3 = high persistence, comments, 4 = high persistence, comments limited, 5 = high persistence, comments deleted.

Table 5:

PROCESS Model 4, post hoc analyses, M = Visibility

Path Estimates	Outcome Variables (Y)		
	Authentic Allyship	Prosocial Reasons for Posting	Self-Interested Reasons for Posting
	b (*SE*)	*b* (*SE*)	*b* (*SE*)
Constant	4.04(0.13)**	4.62(0.10)**	-0.02(0.15)**
High Persistence, comments	0.10(0.12)	0.06(0.09)	-0.01(0.14)
High Persistence, comments limited	-0.13(.12)	-0.06(0.09)	0.05(0.14)
High Persistence, comments deleted	-0.05(0.12)	0.07(0.09)	0.09(0.14)
Visibility	0.28(0.02)**	0.18(0.02)**	-0.07(0.03)*
F	24.90	17.86	1.35
R^2	0.12	0.09	0.008
		Indirect Effects [95% CI]	
High persistence, comments	0.17(0.05) [0.07,0.27]	0.10(0.03) [0.04,0.17]	-0.04(0.02) [-0.09,-0.002]
High persistence, comments limited	0.13(0.05) [0.04,0.24]	0.08(0.03) [0.02,0.15]	-0.03(0.02) [-0.08,-0.008]
High persistence, comments deleted	0.13(0.05) [0.04,0.24]	0.08(0.03) [0.02,0.15]	-0.03(0.02) [-0.08,-0.001]

Note. *$p < .05$. **$p < .01$. Experimental conditions are coded as: 1 = high persistence, no comments, 2= high persistence, comments, 3 = high persistence, comments limited, 4 = high persistence, comments deleted.

Table 6:

PROCESS Model 4, M = Comment Deletion

Path Estimates	Outcome Variables (Y)		
	Authentic Allyship	Prosocial Reasons for Posting	Self-Interested Reasons for Posting
	b (*SE*)	*b* (*SE*)	*b* (*SE*)
Constant	5.78(0.22)**	5.54(0.16)**	-3.82(0.23)**
Comment Factor	-0.001(0.07)	0.03(0.05)	-0.12(0.07)
Comment Deletion	-0.14(0.04)*	-0.05(0.03)	0.25(0.04)**
F	5.90	1.45	15.5
R^2	0.03	0.01	0.08
		Indirect Effects [95% CI]	
Comment Deletion	-0.09(0.03) [-0.17,-0.03]	-0.03(0.02) [-0.08,0.01]	-0.18(0.02) [0.01,0.26]

Note. *$p < .05$. **$p < .01$. Experimental conditions are coded as: 1 = high-persistence-no comments, 2= high-persistence-comments, 3 = high-persistence-comments limited, 4 = high-persistence-comments deleted.

Table 7:

PROCESS Model 4, M = Comment Restriction

Path Estimates	Outcome Variables (Y)		
	Authentic Allyship	Prosocial Reasons for Posting	Self-Interested Reasons for Posting
	b (*SE*)	*b* (*SE*)	*b* (*SE*)
Constant	5.88(0.34)**	5.74(0.25)**	4.24(0.36)**
Comment Factor	-0.12(0.15)	-0.11(0.11)	-0.18(0.16)
Comment Restriction	-0.09(0.05)	-0.02 (0.03)	0.17(0.05)*
F	3.84	1.55	5.69
R^2	0.02	0.01	0.03
		Indirect Effects [95% CI]	
Comment Restriction	-0.14(0.07) [-0.30,0.01]	-0.04(0.05) -0.15,0.06]	0.25(0.09) [0.07,0.46]

Note. *$p < .05$. **$p < .01$. Experimental conditions are coded as: 1 = high persistence, no comments, 2= high persistence, comments, 3 = high persistence, comments limited, 4 = high-persistence, comments deleted.

Table 8:

PROCESS Model 6, serial mediation, M = Persistence, Commitment

Path Estimates	Outcome Variables (Y)		
	Authentic Allyship	Prosocial Reasons for Posting	Self-Interested Reasons for Posting
	b (*SE*)	*b* (*SE*)	*b* (*SE*)
Constant	4.37(0.12)**	4.37(0.12)**	4.37(0.12)**
High Persistence, no comments	-0.27(0.13)*	0.27(0.13)*	0.27(0.13)*
High Persistence, comments	-0.03(0.13)	0.03(0.13)*	0.03(0.13)*
Hi Persistence, comments limited	-0.37(0.13)**	-0.37(0.13)**	-0.37(0.13)**
High Persistence, comments deleted	-0.22(0.13)	-0.22(0.13)	-0.22(0.13)
Persistence	0.30(0.02)**	0.30(0.02)**	0.30(0.02)**
F	38.58	38.58	38.58
R^2	0.18	0.18	0.18
Path Estimates			
Constant	1.13(0.13)**	2.53(0.11)**	6.87(0.19)**
High Persistence, no comments	-0.09(0.09)	0.02(0.07)	0.04(0.13)
High Persistence, comments	0.01(0.09)	0.08(0.07)	0.09(0.13)
High Persistence, comments limited	-0.03(0.09)	0.08(0.07)	0.02(0.13)
High Persistence, comments deleted	-0.07(0.09)	0.14(0.07)	0.14(0.13)
Persistence	0.07(0.01)**	0.03(0.01)	-0.03(0.02)
Commitment	0.65(0.02)**	0.45(0.01)**	-0.04(0.03)**
F	170.1	124.33	30.5
R^2	0.54	0.47	0.17
		Indirect Effects [95% CI]	
High persistence, no comments	0.21(0.04) [0.12,0.30]	0.14(0.03) [0.08,0.21]	-0.12(0.03) [-0.19,-0.07]
High persistence, comments	0.22(0.04) [0.13,0.31]	0.15(0.03) [0.08,0.22]	-0.13(0.02) [-0.19,-0.08]
High persistence, comments limited	0.21(0.04) [0.13,0.31]	0.15(0.03) [0.09,0.0.22]	-0.13(0.02) [-0.19,-0.08]]
High persistence, comments deleted	0.23(0.04) [0.14,0.33]	0.16(0.03) [0.09,0.23]	-0.14(0.03) [-0.20,-0.08]

Note. *$p < .05$. **$p < .01$. Experimental conditions are coded as: 1 = high persistence, no comments, 2= high-persistence, comments, 3 = high persistence, comments limited, 4 = high persistence, comments deleted.

Appendix B: Sample Stimuli

Table 9:

Prosocial messages and captions for all conditions

Post Message	Caption
(1) Reminder: Women's Rights are Human Rights. We are all allies.	(1) I'm always an ally for women. #humanrights #womensrights
(2) #Supportwomensrights. End the wage gap.	(2) Globally women are less paid than men, earning 77 percent of what men make. Personally, I stand for equal for women #womensequality
(3) Did you know? It may take us 131 years to close the gender gap. Let's make it sooner.	(3) #gendergap #women #empowerment
(4) 50% of women experience sexual violence in their lifetimes. Here's what you can do: Be an active bystander Be an ally, listen and learn Change the culture	(4) Here's what I'm doing to help prevent sexual violence against women. #humanrights #womensrights #metoo
(5) Join the fight for women's rights.	(5) I'm fighting, are you? #womensequality #fightforwomen

Figure 1: *Sample stimuli for low persistence (Stories) condition*

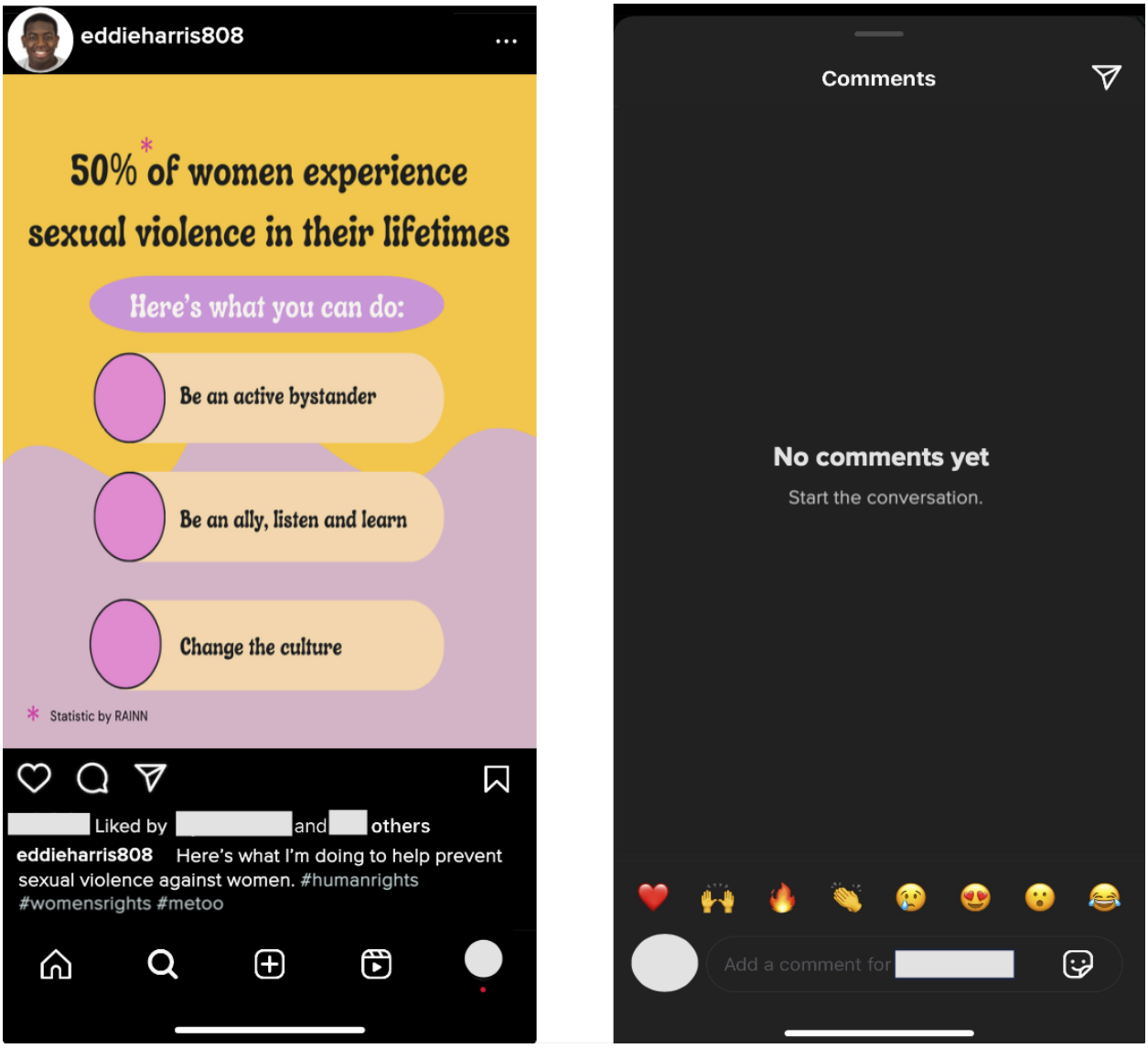

Figure 2: *Sample stimuli for high persistence (Posts), no comments condition*

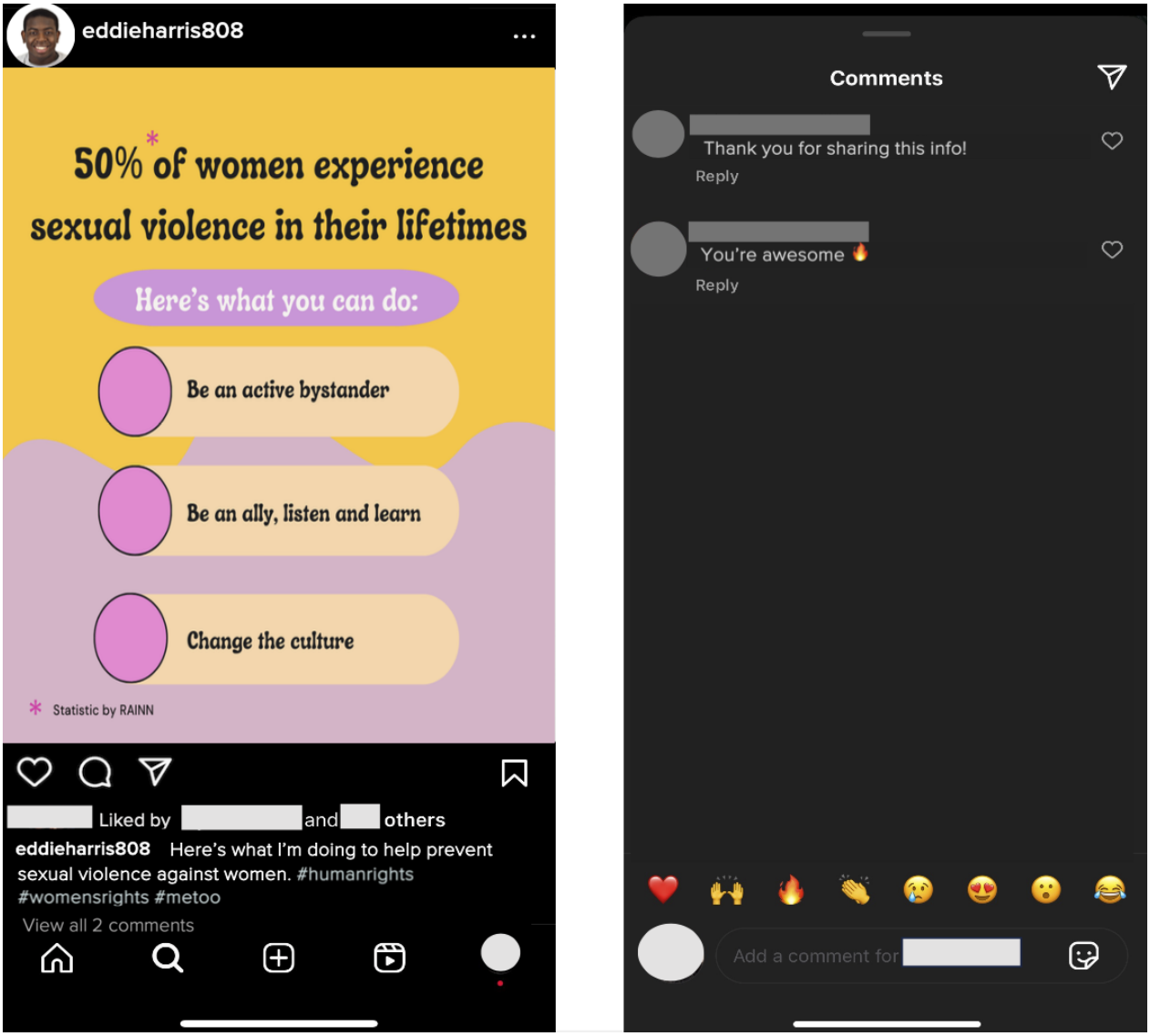

Figure 3*: Sample stimuli for high persistence (Posts), comments*

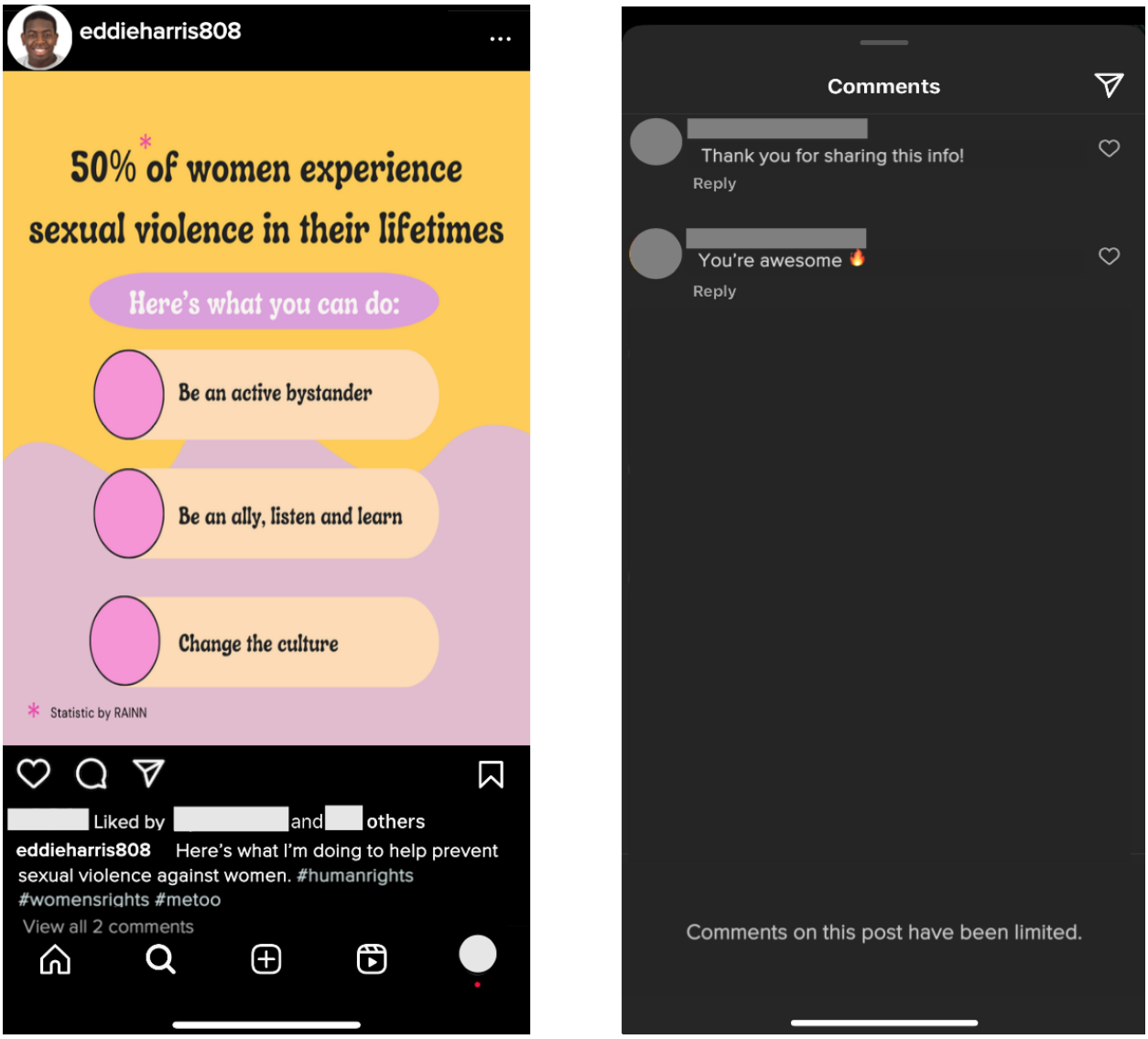

Figure 4: *Sample Stimuli for high persistence (Posts), comments limited*

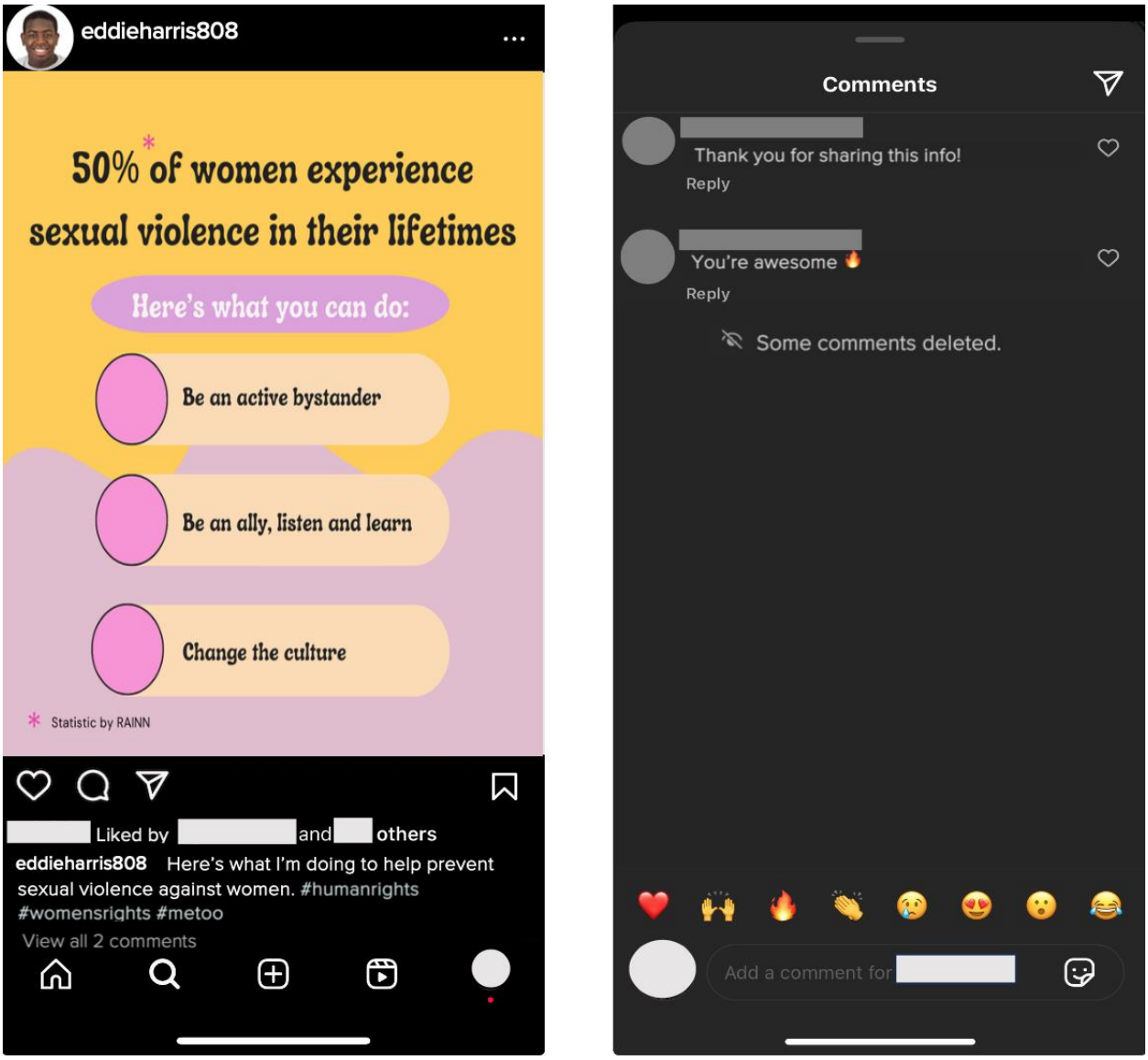

Figure 5: *Sample stimuli for high persistence (Posts), comments deleted*

Appendix C: Survey Instruments

Study Introduction

We are interested in understanding how you perceive images and comments on Instagram. Once you advance to the next screen you will see multiple images that were posted publicly on Instagram by various users. Please take a moment to view the user's profile, post as well as any accompanying comments and captions. You will be able to advance to the next screen after you read the message.

Induction check

Persistence– adapted from Monge et al. (2024).
Measurement: 7-point semantic differential response.

For each of these statements, please indicate your answer.

The users shared messages that support women in a way that will last…

1. A short duration of time – A long duration of time
2. A short amount of time – A long amount of time
3. Not a very long amount of time – A very long amount of time
4. An insubstantial amount of time – A substantial amount of time

Visibility adapted from Monge et al. (2024).
Measurement: 7-point semantic differential response.

For each of these statements, please indicate your answer.

How many people do you think have the posts you just viewed?

1. Not that many people – Many people
2. A small amount of people – A large amount of people
3. Relatively few people – Relatively many people
4. Hardly anyone – A lot of people

Outcome Measures

Dissemination Control – adapted from DeAndrea and Carpenter (2018).
Measurement: 1 = strongly disagree to 7 = strongly agree.

For each of these statements, please indicate your answer.

1. The Instagram users controlled what comments appeared on their posts.
2. Only comments approved by the Instagram users appeared on their posts.
3. The Instagram users picked what comments were presented on their post.
4. The Instagram users made sure viewers only saw comments that they wanted viewers to see.

5. The Instagram users controlled who could comment on their post.
6. Only viewers approved by the Instagram users could comment on their posts.
7. The Instagram users picked who could comment on their posts.
8. The Instagram users made sure viewers only saw comments from certain people.

Authenticity – adapted from Pittman et al. (2022)
Measurement: 1 = strongly disagree to 7 = strongly agree.

For each of these statements, please indicate your answer:

1. I think the Instagram users who posted are genuine
2. I think the Instagram users who posted are authentic
3. The Instagram users who posted seem to care about women's' empowerment issues
4. The Instagram users' posts seem to be aligned with their core values

Prosocial Reasons for Posting

Measurement: 1 = strongly disagree to 7 = strongly agree.

Instructions: Think about the reasons people might post images like the ones you saw on Instagram. Please indicate your agreement with the following reasons for why you think people posted these images on Instagram.

The men shared these images on Instagram to…

Information source
1. Share information about a social cause
2. Generate awareness about a social cause
3. Share the experiences of others regarding a social cause

Online mobilization
1. Mobilize online support for a social cause
2. Influence others to advocate for change regarding a social cause
3. Encourage others to share similar content

Opinion Expression
1. Show solidarity for a social cause

2. Influence opinion change regarding a social cause
3. Engage in conversation about a social cause with others

Self-Interested Reasons for Posting –adapted from Vendemia and DeAndrea (2018) and Vendemia et al., (2021)

Measurement: 1 = strongly disagree to 7 = strongly agree.

Instructions: Think about the reasons people might post images like the ones you saw on Instagram. Please indicate your agreement with the following reasons for why you think people posted these images on Instagram.

The men shared these images on Instagram to…

Social Feedback
1. gain "likes"/shares/endorsement
2. get favorable feedback
3. receive praise from others

Attention-Seeking
1. get attention
2. show off
3. brag

Audience Scale / Reach
1. gain popularity
2. grow a following
3. attract an audience

Spontaneity
Measurement: Semantic Differential; 1—7

The men shared the photos …

1. Without giving much thought – With careful deliberation
2. With little effort – With much effort
3. In the moment – After much thought
4. With little concern – With great care

Commitment

Measurement: Semantic Differential; 1—7

The men who posted the images are…

1. Not dedicated to supporting women– Dedicated to supporting women
2. Not committed to supporting women – Committed to supporting women
3. Not truly engaged with supporting women – Truly engaged with supporting women

Other

Screeners

Are you 18 years of age or older?
1. Yes
2. No

What is your gender identity?
1. Man
2. Woman
3. Non-binary
4. Other [fill in blank]

Do you use Instagram?
1. Yes
2. No

Demographics

Age: _____

Gender Identity:

1. Man
2. Woman
3. Non-binary
4. Other [fill in blank]

Race:

1. White /Caucasian
2. Black/ African American
3. Hispanic/ Latino
4. Asian/ Asian American
5. American/ Pacific Islander
6. American Indian/ Indigenous American
7. Multiracial
8. Other