ORIGINAL ARTICLE

Beyond Neutrality: Conceptualizing Platform Values

Blake Hallinan (ID) **, Rebecca Scharlach** (ID) **& Limor Shifman** (ID)

Department of Communication & Journalism, The Hebrew University of Jerusalem, Mount Scopus, Jerusalem 91905, Israel

Social media platforms are prominent sites where values are expressed, contested, and diffused. In this article, we present a conceptual framework for studying the communication of values on and through social media composed of two dimensions: scale (from individual users to global infrastructures) and explicitness (from the most explicit to the invisible). Utilizing the model, we compare the communication of two values—engagement and authenticity—in user-generated content and policy documents on Twitter and Instagram. We find a split between how users and platforms frame these concepts and discuss the strategic role of ambiguity in value discourse, where idealistic meanings invoked by users positively charge the instrumental applications stressed by platforms. We also show how implicit and explicit articulations of the same value can contradict each other. Finally, we reflect upon tensions within the model, as well as the power relations between the personal, cultural, and infrastructural levels of platform values.

Social media platforms have simultaneously been credited with bringing the world closer together and strengthening divisions, championing authentic self-expression and incentivizing fake interactions, mobilizing political engagement and reducing politics to empty posturing. These competing visions share the same premise: whatever else they may be, platforms are *not* neutral. From algorithmic "audits" (Sandvig, Hamilton, Karahalios, & Langbort, 2014) to accusations of suppressing speech through "shadow bans" (Myers West, 2018), politicians, researchers, and the broader public are increasingly grappling with the partisanship of platforms. The debate surrounding these issues undermines the foundational metaphor of social media as a mere conduit and foregrounds an understanding of platforms as sites where values are expressed, contested, and diffused.

Corresponding author: Blake Hallinan; e-mail: blake.hallinan@mail.huji.ac.il

 1

Despite growing interest in the normative dimensions of social media platforms (e.g., Leurs & Zimmer, 2017), there is no consensus on what "platform values" are or how they are communicated. The disagreement stems, in part, from the complexity of platforms. As Tarleton Gillespie explains, social media platforms such as Facebook, Twitter, Instagram, TikTok and YouTube are data infrastructures that "host, organize, and circulate user's shared content or social exchanges" (2017, p. 417). Such platforms feature "a convergence of different systems, protocols, and networks that connect people in different and particular ways and thus offer specific conditions of possibility" (Langlois, McKelvey, Elmer, & Werbin, 2009). Accordingly, academic research on platform values tends to take place in silos, clustered around different features like people or protocols, communities or content, which makes comparison difficult and yields seemingly incompatible conclusions.

To move beyond the recognition of non-neutrality and assess polarized claims about platforms, we need new ways of conceptualizing the construction of values on social media. We begin with an overview of platform studies, highlighting the theoretical and practical challenges of researching values in this context. Next, we examine existing value theories, showing how the concept has been formulated at different levels, from the individual (personal values) to the collective (cultural values) to those built into material systems (infrastructural values). While each domain is relevant to the study of social media, an integrative account of platform values needs to address multiple levels and the relationships between them. Following this observation, we present a conceptual framework for studying the communication of values on and through social media composed of two dimensions: scale (from individual users to global infrastructures) and explicitness (from the most explicit to the invisible). We demonstrate the utility of the model by comparing the communication of engagement and authenticity in user-generated content and policy documents on Twitter and Instagram, revealing tensions between the personal, cultural, and infrastructural levels of platform values. We conclude with a discussion of the theoretical, methodological, and political implications of this approach and identify directions for future research.

The problem of platform values

With billions of people around the globe using social media to express themselves, connect with others, and mobilize politically, platforms matter. Indeed, platforms have been described as fundamental infrastructures of contemporary societies that reflect and even "*produce* the social structures we live in" (Van Dijck, Poell, & Waal, 2018, p. 2, emphasis in original). The sub-field of platform studies investigates how information and communication technologies shape social life by connecting diverse and geographically dispersed people through data, programming, and design (Plantin, Lagoze, Edwards, & Sandvig, 2018). While companies initially used the term "platform" to frame social media as a neutral intermediary and avoid political pushback and regulations (Gillespie, 2010), mounting evidence demonstrates a multitude of ways that platforms exhibit or encourage particular biases and

values. From moderating content (Gillespie, 2018; Myers West, 2018) to promoting color-blind discourses about neutral platforms and respectable users (Brock, 2020), platforms set significant conditions for public life.

Even as people across the political spectrum reckon with the empty promise of platform neutrality, there is no shared understanding of what platform values are—to say nothing of what they should be. A special issue on the subject published in *Information, Communication, and Society* demonstrates diverse understandings of what platform values entail. The introductory essay argues that the concept refers to "top-down decisions such as the profit-oriented workings of algorithms that differentially value some users over others and bottom-up user practices that both sustain and subvert value-laden mechanisms" (Leurs & Zimmer, 2017, p. 805), while the individual papers address diverse issues including filter bubbles and affective communities, often without explicitly invoking values. Beyond the special issue, work on platform values clusters around specific topics such as user-generated content (Lewis, 2020; Shifman, 2019), community norms (Brock, 2020), algorithmic recommendations (Bucher, 2018; Rieder, 2020), and design (Fiesler, Morrison, & Bruckman, 2016). While this body of research effectively demonstrates the inherent normativity of social media, comparative work is rare and the relationships between different sites of value construction tend to go unspecified. Despite repeated calls to connect bottom-up and top-down mechanisms of valuation (Bechmann & Lomborg, 2013; Leurs & Zimmer, 2017), the theoretical tools for doing so are few and far between.

Somewhat counterintuitively, the tendency to organize values into coherent ideological structures further complicates the task of identifying and analyzing platform values. *The Platform Society*, for example, frames the phenomena of platformization as a clash between two sets of values associated with the competing ideologies of capitalism and democracy (Van Dijck et al., 2018). While this approach highlights important tensions for the global governance of platforms, we contend that the study of platform values would benefit from disarticulating the concepts of values and ideology. To put the distinction simply, ideology refers to a set of values and beliefs, entwined with social relations and power structures. However, the relationship between ideology and values is complicated by the fact that the same values frequently connote different meanings under different ideologies. For example, a capitalist notion of "freedom" might signify the "freedom of the market," while a feminist understanding might refer to "freedom from oppression" (Van Dijk, 2000, p. 69). When multiple ideologies operate in a given context, presuming that people use and interpret values in the same way obfuscates the analysis of valuation processes (Lamont, 2012). Furthermore, using values (rather than ideologies) as a starting point may open new paths to challenge the all-embracing dichotomies often invoked in discussions of ideology.

Values at scale

To further develop a conceptual framework for platform values, we turn to the different theorizations of values present in the literature. The concept has a long and

contentious history (Hitlin & Piliavin, 2004). Broadly defined as notions about the desirable that guide judgment and behavior, researchers have conceptualized the operation of values at various interconnected levels. A survey of the literature led us to define three prominent scales particularly relevant to social media platforms: personal, cultural, and infrastructural. Each of these formulations are associated with different disciplines and, by extension, draw upon distinctive (and at times contradicting) epistemological commitments and methodological approaches.

The study of *personal values* has been conducted mainly within the field of psychology and typically refers to an individual's guiding beliefs about desirable conduct. Shalom Schwartz's universal value model, which charts the relationship between distinct values found worldwide, is probably the most influential of these theories (Schwartz, 1992, 2012). Within this tradition, developed primarily through survey-based studies, personal values are considered relatively stable attributes that shape decision-making across domains (e.g., work, home, leisure) and over time (Sagiv, Schwartz, & Arieli, 2011).

The next scale of conceptualization, *cultural values*, is concerned with collective ideas about desirable conduct. Compared to personal values, cultural values can be trickier to identify given disagreements over the boundaries of human groupings. Within this level, a prominent trajectory of research focuses on national cultures, where large comparative studies have demonstrated strong cross-cultural differences (Hofstede, 2003; Inglehart, 1990). Studies of national values often privilege survey research based on statistically representative samples or strategic populations, such as the employees of a multinational corporation (Hofstede, 2003). A second trajectory focuses on guiding beliefs adopted or promoted by organizations, coming out of sociology, business, and journalism. While some of this work employs survey data, the values of organizations have also been analyzed through the statements of individuals within an organization, communicative genres, and official policies that reflect shared symbols and norms (Sagiv et al., 2011).

At the next level, *infrastructural values* refer to guiding beliefs about desirable conduct enacted through technological artifacts and systems, both intended and unanticipated. Foundational works in this trajectory, often drawing upon historical and interpretive analysis, show how technological systems are bound up in our social, ethical, and moral worlds (Friedman & Nissenbaum, 1996; Introna & Nissenbaum, 2000; Knobel & Bowker, 2011; Winner, 1986). Infrastructures embody values through the formalization of decision-making procedures and ideas about the good life, as well as the creation of incentive structures that encourage people to form, foster, and express particular values. Where personal and cultural values privilege the human as the locus of analysis, research on infrastructural values demonstrates that expression is not an exclusive property of humans—technology, architecture, and artifacts have expressive qualities and construct values.

While there are many points of overlap between these categories, personal, cultural, and infrastructural values are useful abstractions that reflect distinct approaches to conceptualizing values. Each level of value expression offers relevant

insights for the study of social media platforms but no single articulation of values offers a comprehensive view, prompting the need for an integrative model.

The construction of values on social media platforms: a conceptual framework

Our proposed framework is based on the notion that research on values and social media should address the fundamental question of *where* and *how* values are expressed on platforms. We use "expression" in the broad sense, referring to the manifestation of qualities, content, or feelings through signifying systems. Our model brings together two dimensions of value communication: the scale of expression (personal, cultural, infrastructural) and its explicitness (explicit, implicit, invisible). The different elements of the two dimensions are interwoven such that each site of value communication is always bound up in a range of scales and degrees of explicitness (see Figure 1). The analysis of platform values thus entails pulling apart the different elements, seeing how they relate to each other, and identifying which are most salient. After introducing the three main scales of value communication, we discuss how they intersect with the explicitness of communication.

The first scale of value expression on social media concerns personal values which manifest through user-generated content, evaluative practices, and usage patterns. Each post to social media is, as Tarleton Gillespie argues, "a tiny value assertion" (2018, p. 210). People communicate particular values both through producing content and interacting with the content of others by liking, commenting, and sharing.
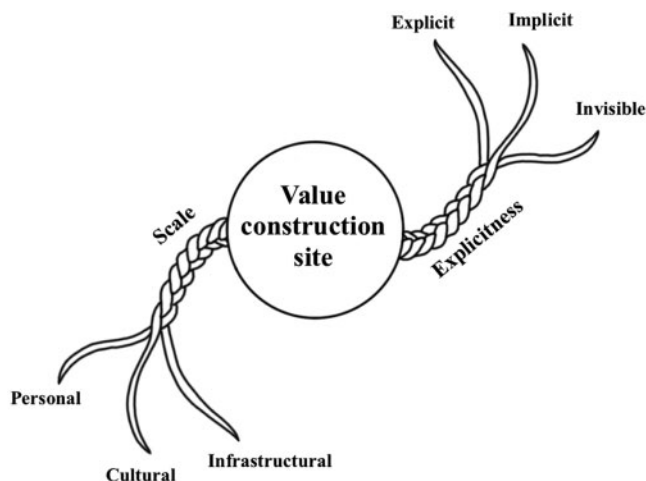

Figure 1 A conceptual framework for the study of platform values that locates potential sites of value construction as bound up in two dimensions: the scale and explicitness of communication.

Of course, no social media user is an island. Each person's network of relationships, demographic factors, and socioeconomic status plays a crucial role in the expression of values. Accordingly, it is also important to consider cultural values, which may be expressed through community norms such as the typical practices of authenticity and accountability associated with fashion bloggers (Abidin, 2018) or reactionary political commentators (Lewis, 2019). These informal modes of association exist alongside the various companies and formal organizations that operate within a given platform's ecosystem, including advertisers, small businesses, and, of course, social media corporations. Corporate value statements, policy documents, financial disclosures, and public statements contribute to the corporate construction of values such as privacy (Vaidhyanathan, 2018), sharing (John, 2016), and openness (Hoffmann, Proferes, & Zimmer, 2018).

While notions of personal and cultural values have been developed since the mid-20th century and apply to many realms of life, the study of infrastructural values is more closely tied to systems of communication. Infrastructural values may be expressed through interfaces, algorithms, APIs, engagement metrics, and reputation systems. Content moderation is perhaps the most obvious instance of "systemic and values-laden" design (Gillespie, 2018). At the immediate level, content moderation involves determinations of permissible content. However, the global scale of the decisions often places content moderation in the position of reconciling contradictory values, choosing, for example, to promote free speech or curtail harassment and bullying. The interface design of platforms is also bound up with the construction of values. For instance, tools of evaluation such as the like button simultaneously promote certain behaviors and flatten nuanced palettes of values into a single visual form and quantitative measure (Hallinan & Brubaker, 2021).

The second dimension that cuts across all three scales of expression is the degree of explicitness. While explicitness is, of course, a nuanced and continuous spectrum, we differentiate between three levels of value expression for the sake of analytical clarity which we dub "explicit," "implicit," and "invisible" (see Table 1). Broadly, we find an inverse correlation between explicitness and scale, such that the expression of personal values tends to be more explicit than the expression of infrastructural values. As discussed below, different levels of explicitness require the application of different methods.

As demonstrated in Table 1, each site of value communication contains varying degrees of explicitness. Consider the expression of personal values involved in sharing a New Year's resolution on Instagram. The caption of the image might clearly state a goal of being healthy while the accompanying gym selfie connotes a wide array of possible values, ranging from effort to normative beauty standards. This example demonstrates a broader principle regarding the difference between verbal and visual articulations of values. Whereas words can signify abstract value concepts directly, the inference of values from photos requires an additional level of interpretation, such as concluding that a smiling person represents "happiness." Moreover, as Machin (2007) argues, images contain a multitude of elements that

Table 1 Scale and Explicitness of Value Construction on Social Media Platforms: Definitions and Illustrative Examples

Scale	Sites	Example	Explicit	Implicit	Invisible
Personal	User-generated content, evaluation practices, usage patterns	Posting a New Year's resolution on Instagram	The goal mentioned in the caption	Qualities featured in the photo	Goals and concerns taboo to talk about
Cultural	Community norms, platform policies, PR statements	Terms of Service	Values named in the policy	How the platform puts the values into practice	Governing principles that are absent from the ToS (e.g., profit)
Infrastructural	Interfaces, algorithms, moderation systems, APIs	Instagram Filters	Aesthetics and beauty, creativity	Specific beauty standards, need for improvement	Biases in the algorithms behind filters (e.g., skin color and facial recognition)

cannot easily be broken down into discrete units and may thus invoke diverse interpretations.

Cultural value expression is crystalized in sites such as the community rules of a subreddit or Facebook group, along with official policy documents like Terms of Service. However, the enforcement of policies differs significantly between platforms and populations, which can reveal tensions between implicit and explicit values. Additionally, the values that are not mentioned in governing documents can be just as revealing as those that are highlighted. Historical analysis of platform governance, along with the comparison of different companies and regional policies, can help detect significant absences—that is, invisible values.

Infrastructural values are typically the most opaque level of analysis. While values may occasionally surface explicitly in design considerations—for example, the feminist design principles that guided the creation of a popular fan fiction platform (Fiesler et al., 2016)—this is often not the case. Consequently, attention to implicit and invisible values of infrastructure is particularly important. Consider the role of filters on Instagram: since its launch, the platform has featured built-in tools to modify the appearance of photographs and more recent developments have extended filters to other media formats (e.g., video, Boomerangs), as well as opened up the ability to create and share user-generated filters. At the explicit level, filters celebrate aesthetics and creativity, even as they implicitly support particular standards of beauty and the need for improvement. And, because digital filters are computational techniques, the biases of algorithms, especially facial recognition algorithms, can introduce a different set of values (Buolamwini & Gebru, 2018). Other ways of studying the implicit values of infrastructure include ethnographies of programmers and designers (Seaver, 2017), the interpretation of the constitutive techniques of algorithms (Rieder, 2020), and the analysis of a platform's incentive structures (Bucher, 2018). Invisible expressions of infrastructural values can be detected through the comparison of differential outcomes of search and recommendation systems (Noble, 2018) or the critical analysis of absences inapplication programming interfaces (APIs), such as the lack of available data on unfriending and disconnectivity on Facebook (John & Nissenbaum, 2018).

The different levels of explicitness associated with the articulation of personal, cultural, and infrastructural values resonate with the methodological approaches typical to each. While the analysis of explicit values embedded in user-generated content or policy documents could be done by quantitative methods such as content analysis, tracing the more implicit layers of value construction, particularly at the infrastructural level, requires the use of qualitative and inductive approaches such as interviews, ethnography, and walk-throughs. Furthermore, a comparative analysis of people, places, platforms, and periods can facilitate the extremely difficult process of tracing invisible or repressed values. A holistic understanding of the construction of values in digital spheres is thus a broad project that requires the integration of diverse data and methods.

Case study: the construction of engagement and authenticity as platform values

To test the utility of our suggested framework, we conducted an exploratory analysis of two prominent values associated with social media: engagement and authenticity. Both concepts have been studied extensively with conflicting and competing understandings. In what follows, we demonstrate how our model can shed light on the sources of different interpretations.

Social media engagement typically refers to interactivity and participation on and through social media (Jenkins, 2006), often reduced to the most visible and obvious signifiers: likes, comments, and articulated relationships (Bucher, 2018). Mark Zuckerberg, for example, defined Facebook's "natural engagement pattern" (2019) as aggregated interactions with social media posts. Understood this way, engagement is a defining characteristic of social media and a driving force behind its design and commercialization (Gehl, 2014; Lampe, 2011).

But why is engagement, understood as interactivity and participation, a value—that is, something that should be strived for? A cross-disciplinary book on the subject valorizes engagement as something that "will affect some kind of change in users for the better—be it affective, cognitive, or behavioural" (O'Brien & Cairns, 2016, p. 3). Through an integrative analysis of the literature, we identified three positive understandings of the term which we name strategic, social, and civic engagement. *Strategic engagement* refers to creating "engaging" content on social media, measured in terms of platform metrics. Strategic engagement acts as a pre-requisite to visibility on social media platforms (Cotter, 2019) and something that generates customer loyalty (Hollebeek, 2011). *Social engagement* refers to interactions with friends, family, and communities of interest. On social media, this type of engagement can be a form of social support (Scissors et al., 2016) and the enactment of participatory culture (Picone et al., 2019). *Civic engagement* refers to participation in activities and organizations dealing with topics of public concern (Putnam, 2001), including public debate, community organizing, and the news (Choi & Shin, 2017; Nelson, 2019).

If the first commandment of social media is "to engage," the second one might well be "authentically." Users are expected to engage with platforms, brands, and each other in a specific manner broadly defined as being authentic. Yet the meaning of authenticity in such environments is often contested (Lee, 2020). Shifman (2018) distinguishes between external and internal authenticity, with the former defined as communication that is truthful to an external reality and the latter reflecting an invisible "inner core." This notion of internal authenticity goes back to the philosophical movements of existentialism and idealism that continue to influence contemporary culture (Banet-Weiser, 2012; Taylor, 2018).

An integrative analysis of the literature on social media supports the general distinction between internal and external authenticity but also suggests the need for a third category. We propose three notions of authenticity as a value: informational,

individual, and cultural. *Informational authenticity* refers to the basic perception of authenticity as "external truth" and the importance of validating facts about people, products, and statements. In contrast, *individual authenticity* refers to the internal desire to express the "real self." Contemporary studies of social media note that being authentic is an important aspect of the relationship between artists and audiences (Baym, 2018), as well as influencers and followers (Abidin, 2018; Marwick & boyd, 2011). Finally, *cultural authenticity* is concerned with various facets of culture such as art and music (Dutton, 2003), as well as ethnic and national constructions of culture such as the longing for "authentic" experiences among tourists (MacCannell, 1973). Cultural authenticity is therefore related to internal notions of what the "essence" of a culture is and to external notions that reinforce the association between products or places and groups.

While social media platforms promote authenticity, with Instagram and Facebook's Community Standards even claiming it as the "cornerstone of our community," the realization of authenticity is highly contested. Critiques of social media in general, and Instagram in particular, argue that features such as photo filters, the public display of engagement metrics, and the prominence of influencers and sponsored content incentivize highly idealized representations of people and their lives. Instagram Stories, a feature for sharing ephemeral photos and videos introduced in 2016, provides a less formal way of posting that is not subject to the same pressures for idealized representations (Leaver, Highfield, & Abidin, 2020). The social salience of these pressures is exemplified by the rise of "finstas," private or pseudonymous Instagram accounts created for the purpose of posting "unfiltered" expression such as self-deprecation, edgy humor, and controversial opinions (Dewar, Islam, Resor, & Salehi, 2019). Even among the highly strategic forms of self-presentation associated with fitness influencers (Reade, 2020) and makeup artists (Hurley, 2019), audiences recognize specific aesthetic qualities and subject matters as indicators of authenticity, such that in some contexts, choosing not to use a filter can be a way of performing authenticity. Together, this research on user practices demonstrates that authenticity is also a matter of affordances and aesthetics.

Methods

To make sense of how the meanings of engagement and authenticity are constructed on social media platforms, we conducted a comparative content analysis involving several platforms and scales of communication. We focused on two sites where the expression of values was likely to be relatively explicit: user-generated content and platform policy documents. The two platforms selected for analysis—Twitter and Instagram—feature relatively public conversations. Furthermore, the platforms are associated with different audiences and types of communication that enable us to investigate differences between platforms and modes of value expression. Twitter favors text-based communication, although posts can contain images,

video, and audio messages. The platform also plays a major role in the circulation of news and political organizing (Burgess & Baym, 2020). Instagram favors photos and videos, and the platform's user base is significantly larger, younger, and more female compared to Twitter (Chen, 2020).

On Twitter, we used the program Mozdeh (http://mozdeh.wlv.ac.uk/) to collect Tweets containing the words "engage," "engagement," "authentic," and "authenticity" for two one-week periods, one month apart in February and March 2020. We then combined the tweets for each value and selected the most popular content units from each data set based on the total number of retweets (n = 300, 150 tweets per value). On Instagram, we adopted a similar approach using *Instagram Scraper* from the Digital Methods Initiative (University of Amsterdam), collecting posts that contained the search terms as hashtags. This tool used the API of Instagram to retrieve metadata of public posts for username or hashtag queries (without downloading the images—for more information, see https://wiki.digitalme thods.net/Dmi/ToolInstagramScraper). We then sorted the data according to the total number of likes, selected the top 150 units per data set, and manually generated screenshots for each post (n = 300, 150 posts per value). For both Twitter and Instagram posts, we excluded content that was not primarily written in English, as well as engagement posts related to weddings.

While the volume of user-generated content required us to narrow down the data set to enable human coding, the finite nature of platform policy documents allowed for an analysis of all relevant mentions of the two values. For Instagram, we analyzed the Terms of Use, Community Standards, Community Guidelines, Platform Policy, and Data Policy. For Twitter, we analyzed the Twitter Rules, Terms of Service, Developer Policy, and Privacy Policy. We also analyzed the help pages of both platforms. We excerpted any clause that included the value keywords (as well as their various declensions), resulting in a total of 54 clauses about engagement and 52 about authenticity (50 derived from Instagram and 56 from Twitter). We used these clauses as our units of analysis for exploring the platforms' articulation of values.

We then coded the entire data set (n = 706), operationalizing the aforementioned framings of the two values. In our codebook, *civic engagement* refers to participation in activities, issues, and organizations dealing with topics of public concern; *social engagement* refers to interactions with friends, family, and communities of interest; and *strategic engagement* relates to activities and interactions designed to achieve a specific goal (e.g., to influence others to buy a product). We defined *informational authenticity* as the external verification of data, news, or material goods as accurate; *individual authenticity* as relating to being "real" or "true" to yourself; and *cultural authenticity* as a quality of an individual, group, or material good that expresses the "essence" of a particular culture. After pilot testing of the codebook, we performed an inter-coder reliability test between two of the authors on a random sample of 50 units, resulting in scores of .90 for authenticity and .87 for engagement (Krippendorff's alpha). We used chi-square tests to identify

dependencies between the meanings associated with the two values and the scale/platform. When results were significant, we further used a post hoc pairwise test (Fisher's exact test) adjusted for multiple testing with the Benjamini and Hochberg false discovery rate (FDR) procedure.

We supplemented the quantitative analysis with three layers of qualitative analysis. First, we conducted a close reading of all policy documents and social media posts to understand the context in which the different meanings appeared. Second, given the central role of hashtags in the construction of shared meanings on Instagram (Leaver et al., 2020), we explored the most popular hashtags co-occurring with the posts in our sample. Finally, to study more implicit value articulations, we zoomed in on the visual construction of authenticity by Instagram users. We decided to delve deeper into this subset of our data given the wide interest in—and criticism of—the articulation of the authenticity on the platform, as discussed in works cited above.

Results

Our analysis of platform policy documents and user-generated content revealed that platforms and users do not invoke the same meanings when discussing authenticity and engagement. Across Instagram and Twitter, platforms employ the terms narrowly, while users express a range of associations. Regarding engagement, platform policies overwhelmingly focus on strategic considerations while users primarily frame the value in civic terms, yet also invoke the other two meanings (see Figure 2). A chi-square test comparing the platform and user framings of engagement (excluding the "other" category) revealed highly significant differences in the distribution of the three meanings ($\chi^2(2, N = 333) = 63.395$, p value < 0.00001). Post hoc tests indicated that while the differences between all pairs were statistically significant (adjusted p value < 0.05), the most significant difference was between strategic framings (higher in platforms) and civic framings (higher in users) (adjusted p value < 0.00001).

Both Instagram and Twitter's policy documents describe specific engagement metrics available through features like Instagram Insights or the Tweet Activity Dashboard. The policy documents also discuss the role of engagement metrics for advertisements, social media campaigns, and platform governance issues such as data protection and content moderation. The platforms simultaneously promote the importance of strategic engagement while placing limits upon it, distinguishing between desirable and undesirable forms. For example, promoting posts through Twitter or Instagram is considered legitimate but promoting posts through third-party services is not. Such qualifications, as discussed in the concluding section, hint at unarticulated values like wealth or power that do not appear in the explicit discussion of engagement. Although we found a few references to engagement as the expression of social bonds or civic participation in the policy documents, these were exceptions to the rule.

From discussions of electoral campaigns to the importance of community uplift, social media users typically invoked a civic understanding of engagement tied to
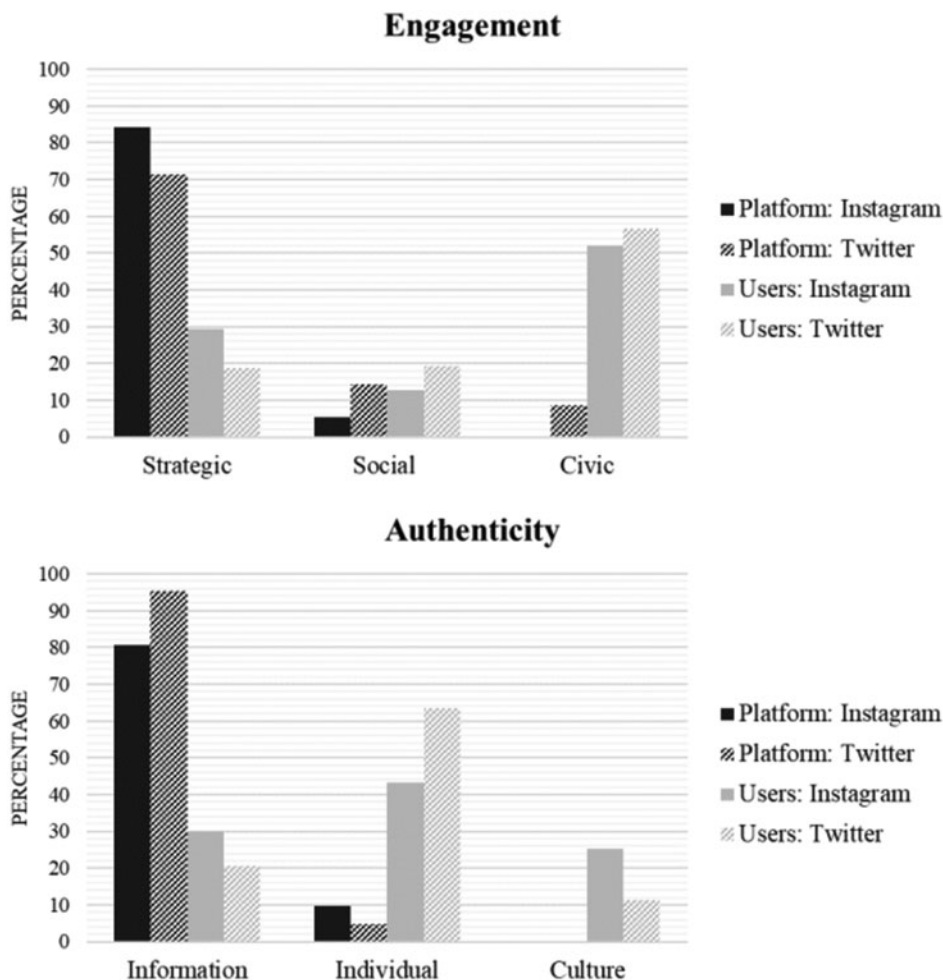
Engagement

Authenticity

Figure 2 A comparison of the meaning of "authenticity" and "engagement" between platform policies and user-generated content. (The "other" category is excluded from the graph and ranges between 0–10.5%).

ideas of democratic participation. Amidst the examples of user-generated content, there were frequent references to specific political candidates, especially Bernie Sanders and Donald Trump, along with voluntary organizations like minority business associations, veteran services, and school clubs. Popular hashtags associated with the civic engagement posts on Instagram included #empower, #community, and #educate, corroborating and inflecting the meaning of the value. While civic engagement was the most prevalent user articulation, we also found a strategic understanding of engagement in discussions of how to influence social media visibility, which ranged from general advice on growing an audience to deliberations over what kinds of content people should encourage. The emphasis on visibility and

transactional modes of social media interaction is also apparent in prominent co-occurring hashtags on Instagram such as #followforfollow and #likesforlikes. Although there were some references to engagement as constitutive of social relationships, including parenting and relationship advice, this use of the term was not as frequent as the civic or strategic inflections. Prominent co-occurring hashtags like #love, #motherhood, and #interact emphasize the interpersonal—and especially familial—nature of these discussions.

In addition to the primary distinction between the platform and user interpretations of engagement, we found minor differences between the two platforms (none of which were statistically significant). Such differences are particularly evident in the policy documentation, where Instagram's near-exclusive focus on strategic engagement contrasts with Twitter's occasional references to both social and civic engagement. The Twitter policy documents claim that specific products and features (e.g., Twitter Moments, Periscope) uniquely enable meaningful social interactions and invoke civic ideas of engagement in explanations of how content moderation policies serve the public interest. Despite similar results for Twitter and Instagram users, a close reading of the user-generated content reveals nuances in the articulation of civic and social engagement. Tweets tended to feature a more explicit and overtly political understanding of civic engagement, with references to elections and debates over how to best generate public participation, in line with prior research about the role of Twitter in political organizing (Burgess & Baym, 2020). On the other hand, Instagram posts about volunteering and community organizations treated civic engagement as something to be witnessed and indirectly encouraged by way of example. Similarly, social engagement posts on Twitter primarily featured opinions about how people should interact with others on the platform, while posts on Instagram typically involved community support for individual goals and creative endeavors such as making music or art. Despite these differences, the general meaning of engagement is surprisingly consistent across platform users and provides a strong contrast to the almost exclusively strategic orientation of the platform policy documents.

On the whole, platform policies articulated engagement as a *means to an end*, whether operating a platform or growing an audience. While some user-generated content also expressed an instrumental understanding of engagement, it was more commonly framed as an *end in itself*, constitutive of communities, political action, and social relationships.

We found a similar split between platforms and users regarding the conceptualization of authenticity. As Figure 2 shows, the platform policies overwhelmingly privilege an informational understanding of authenticity concerned with truth and validation, while users primarily frame authenticity as the expression of innate individual qualities (yet also invoke the other two meanings). A chi-square test comparing platform and user framings of authenticity (excluding "other") revealed highly significant differences in the distribution of the three meanings ($\chi^2(2, N = 340) = 79.218$, $p < 0.0001$). Post hoc tests indicated that this result stems from the

differences between informational framings (higher in platforms) compared to individual and cultural framings, both higher among users (adjusted $p < 0.0001$).

Instagram and Twitter policies discuss the level of authentication required for creating a new account, as well as the significance of verification badges for businesses and public figures (with no statistically significant differences between the two platforms). According to the policy documents, informational authenticity is a pre-requisite for safety, security, and trust for the platform and all of the interactions that take place upon it. References to individual authenticity are rare and primarily appear as a benefit of specific features. The Instagram policies also refer to the platform as an "authentic and safe place for inspiration and expression," a broad and ambiguous appeal that refers to unspecified qualities of the platform itself and does not fit neatly with established categories of authenticity. Twitter does not appeal to an imagined community and instead refers to authenticity solely as a factor of security and validation for user accounts and information.

The most prevalent meaning associated with authenticity in user-generated content focuses on authentic individual expression. On Twitter, ideas of individual authenticity feature personal affirmations (e.g., a mantra about being authentic and accepting yourself), assessments of the authenticity of celebrities and politicians, and political statements linking authenticity and LGBTQ rights. Instagram posts about individual authenticity featured hashtags such as #selflove, #mentalhealth, and #beyourself, suggesting an idea of authenticity as a combination of self-care and self-improvement in response to external pressures. Some of the captions elaborate on these concerns, narrating personal struggles with issues such as anxiety, eating disorders, doubt, drug use, and peer pressure.

The frank tone of the confessions on Instagram often contrasts sharply with the highly presentational and polished images paired with these captions. Individual authenticity is primarily represented through photographs of a person facing the camera or, less frequently, through infographics featuring inspirational quotes. More than half of images with individuals included evidence of careful composition, such as the use of filters, high-resolution images, and visually striking backgrounds (e.g., a city scene without people or a waterfall). Together, the images, captions, and hashtags suggest that being "internally" authentic is important, but not at the expense of the aesthetic standards and considerations typical to the platform (Leaver et al., 2020). As we discuss below, this reflects tensions between explicit and implicit articulations of authenticity, as well as the influence of platform infrastructure.

The other two framings of authenticity are present in user-generated content as well. Both Instagram and Twitter users connect informational authenticity to material goods (e.g., items available for purchase). Popular hashtags for Instagram posts about informational authenticity—#fashion, #sale, and #jordanfeatures—highlight the focus on commercial interactions and brands. The photos paired with these hashtags are overwhelmingly product-focused, depicting items for sale. While cultural authenticity is represented on both platforms, it varies in thematic focus. Twitter discussions of cultural authenticity often cluster around specific topics like the authenticity of

religious groups or the true representation of popular musicians such as the K-pop band BTS. Conversely, the framing of cultural authenticity often appears on Instagram alongside pictures of traditional food and handicrafts, reflected in the related hashtags of #food, #culture, and #travel. In this sense, cultural authenticity serves as both a marker of group affiliation and as an advertisement for businesses, restaurants, and resorts. While there were statistically significant differences between the two platforms regarding the variance between these three readings among users (adjusted $p < 0.001$), for both Twitter and Instagram, each articulation of authenticity was represented by at least 11% of the sample.

Overall, our analysis of engagement and authenticity demonstrates the patterned nature of the construction of platform values, as well as some of the contradictions underpinning this process. Despite differences in ownership, design, and user base, the policy documents of Twitter and Instagram consistently invoked values in highly specialized and delimited ways. In both cases, the broader understandings of the concepts employed by users may provide positive associations for social media activity, even when that activity is commercial in orientation. Thus, "engagement" is charged with notions of civic participation and social belonging, and "authenticity" with to the innate truthfulness of personal and cultural expression. This reading aligns with John's (2016) innovative investigation of sharing that shows how platforms use the terminology of sharing to camouflage commercial interests and exploitive relations. As a highly calculated form of communication produced under the auspices of large corporations, the consistency of value discourse in policy documents draws upon strategic ambiguity. By delimiting the boundaries of engagement and authenticity in platform policies, companies are able to call upon the broader, aspirational meanings of the values invoked by users while not assuming responsibility for their realization.

Although our analysis of user-generated content and platform policy documents focuses on the personal and cultural levels of value construction, a general reflection on the infrastructural level further complicates our understanding of engagement and authenticity. With regards to engagement, the infrastructures of Twitter and Instagram formalize a strategic understanding of the concept through quantified channels of interaction (e.g., like, comment, retweet/share), which in turn power recommendation and moderation systems. This case is different when it comes to authenticity. While the policy documents emphasize the importance of accurate representation and promote notions of informational authenticity, Instagram's infrastructure prominently features tools such as filters designed to alter and play with digital representations. As such, at least in the case of Instagram, there seems to be a tension between declared policies and infrastructural design.

Conclusion

The main aim of this article was to present a conceptual framework for studying the construction of values on social media platforms. First, we highlighted the significance of platforms for public life, along with the lack of consensus about what

platform values are and how they should be studied. Second, we reviewed different conceptualizations of values across disciplines and identified three formulations of the concept particularly relevant to platforms: personal values, cultural values, and infrastructural values. Next, we outlined the conceptual framework for studying platform values, composed of two interwoven dimensions: the scale of value communication and its degree of explicitness. With a comparative case study of the construction of authenticity and engagement across platforms and scales, we showed how the model can shed light on the complex construction of platform values. Building on the conceptual framework and case study, we now offer two overarching observations about the friction between the different components of the model and the power imbalances associated with them.

First, the construction of platform values is not a smooth, unified process; platform values are contested, with different ideas of the desirable playing out among actors through various modalities. Our framework identifies the scale and explicitness of communication as persistent sources of friction. As our case study demonstrates, while personal, cultural, and infrastructural scales of value expression may align—as in the consistent treatment of engagement at the policy and design levels—this need not be the case. Indeed, the different scales of communication can also constrict and contradict each other, as reflected in the competing senses of authenticity expressed in the organizational documents and platform infrastructure on Instagram. The degree of explicitness marks out another potential fault line in the construction of values, a gap between what is professed and what is practiced (e.g., in the case of verbal vs. visual representations of #authenticity). And while our case study focused on explicit and implicit expressions of values, an initial analysis of the policy clauses on engagement hints at the presence of invisible values. Although the platform policies do not speak about profit or power as values related to engagement, ample evidence suggests them as major motivations for corporate social media platforms (Van Dijck et al., 2018).

Secondly, an insistence on the complexity of platform values does not mean that all expressions of value are equally influential. On the contrary, commercial social media platforms are extremely imbalanced when it comes to the construction of values, with the interactions between social media users and corporations structurally constrained. An individual user cannot directly engage in dialogue with the platform about objections to the Terms of Service, for example, while content moderation systems can remove content and block users in service of the values outlined in the policy documents. These power dynamics were clearly evident in the case study where the dominant sense of engagement as a means, promoted by the platforms, acts as a prerequisite for engagement in the broader political or social sense. Without likes, comments, and articulated connections, alternative articulations of the value have very little visibility and thus influence. With respect to authenticity on Instagram, while each user is free to express their thoughts about this concept, the very existence of filters and other tools of image manipulation built into the platform constitute a powerful force that shapes the aesthetic manifestation of

authenticity in images. Together, the analysis of engagement and authenticity suggests a broader rule about the relationship between different scales of value expression: scale correlates with power. Accordingly, although platform values are not reducible to the commitments built into the infrastructures of social media, accounts which ignore the role of infrastructure are missing out on a foundational, and extremely powerful, component.

While the article presents a conceptual framework for a holistic understanding of the construction of values on social media platforms, the illustrative case study is but one step in that direction. Despite the prominence of engagement and authenticity in corporate and popular discourse, the two values represent a small corner within the greater universe of platform values. Future work can use our model to investigate how longstanding values like fairness, transparency, accountability, and justice play out in the context of social media platforms.

Of course, the identification of platform values, particularly implicit and invisible ones, poses a methodological challenge and more work is needed to develop strategies for tracing and analysing values. Sub-fields such as visual communication and aesthetics, along with techniques such as interviews and meta-analyses of communication norms represent promising directions. Single studies need not, and often cannot, encompass all facets of the platform values model; however, being able to locate studies within the model can help researchers identify points of connection to other work, limitations of existing studies, and directions for future research.

Platform values research also needs to grapple with the dynamics of globalization and cultural specificity. Our case study, with its focus on U.S.-based companies and English language user-generated content, participates in the Anglocentrism common to Internet research (Goggin & McLelland, 2009). Taking the cultural level of the platform values model seriously requires engagement with research from other contexts and vectors of difference, including language, demographics, and geography. The transnational nature and global aspirations of social media platforms increase the importance of studying the intersection between infrastructural and cultural values. Furthermore, cross-cultural comparisons offer a promising strategy for detecting the expression of implicit and invisible values, as the prominence of a value in one context raises questions about its seeming absence in another. While there is no guarantee that every question has an answer, let alone an interesting one, systematic comparison goes a long way towards narrowing the field of possibilities.

Social media platforms need not be monolithic forces of good or evil for there to be pressing social and political stakes. A detailed account of the multifaceted construction of values on platforms provides a way to make sense of existing conditions and advocate for new ones. As neutrality proves to be an empty, perhaps impossible, promise, what are the values that should guide platforms? Moving beyond neutrality requires charting the intensely value-laden ecosystem of social media, a process that requires many hands working together while moving in different ways, and which the conceptual framework put forth here aims to assist.

Acknowledgements

Funding

References

Abidin, C. (2018). *Internet celebrity: Understanding fame online* (1st ed.). Bingley, UK: Emerald Publishing.

Banet-Weiser, S. (2012). *Authentic TM: Politics and ambivalence in a brand culture*. New York: New York University Press.

Baym, N. K. (2018). *Playing to the crowd: Musicians, audiences, and the intimate work of connection*. New York: New York University Press.

Bechmann, A., & Lomborg, S. (2013). Mapping actor roles in social media: Different perspectives on value creation in theories of user participation. *New Media & Society*, *15*(5), 765–781. https://doi.org/10.1177/1461444812462853

Brock, A. (2020). *Distributed Blackness: African American cybercultures*. New York: New York University Press.

Bucher, T. (2018). *If...then: Algorithmic power and politics*. Oxford, England: Oxford University Press.

Buolamwini, J. & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. Proceedings of the 1st Conference on Fairness, Accountability and Transparency. *PMLR*, **81**: 77–91.

Burgess, J. & Baym, N. (2020). *Twitter: a biography*. New York: New York University Press.

Chen, J. (2020, August 4). *Social media demographics to inform your brand's strategy in 2020*. Sprout Social. Retrieved from https://sproutsocial.com/insights/new-social-media-demographics/

Choi, D.-H., & Shin, D.-H. (2017). A dialectic perspective on the interactive relationship between social media and civic participation: The moderating role of social capital. *Information, Communication & Society*, *20*(2), 151–166. doi:10.1080/1369118X.2016.1154586

Cotter, K. (2019). Playing the visibility game: How digital influencers and algorithms negotiate influence on Instagram. *New Media & Society*, *21*(4), 895–913. doi:10.1177/1461444818815684

Dewar, S., Islam, S., Resor, E., & Salehi, N. (2019). Finsta: Creating "Fake" Spaces for Authentic Performance. Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems, 1–6. Retrieved from 10.1145/3290607.3313033

Dutton, D. (2003). Authenticity in art. In J. Levison (Ed.), *The Oxford handbook of aesthetics* (pp. 258–274). Oxford, England: Oxford University Press.

Fiesler, C., Morrison, S., & Bruckman, A. S. (2016). An archive of their own: A case study of feminist HCI and values in design. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16, 2574–2585. doi:10.1145/2858036.2858409

Friedman, B., & Nissenbaum, H. (1996). Bias in computer systems. *ACM Transactions on Information Systems*, *14*(3), 330–347. doi:10.1145/230538.230561

Gehl, R. W. (2014). *Reverse engineering social media: Software, culture, and political economy in new media capitalism*. Philadelphia, PA: Temple University Press.

Gillespie, T. (2010). The politics of "platforms". *New Media & Society*, *12*(3), 347–364. doi:10.1177/1461444809342738

Gillespie, T. (2017). Regulation of and by platforms. In J. Burgess, A. E. Marwick, & T. Poell (Ed.), *The SAGE handbook of social media* (pp. 417–456). Thousand Oaks, CA: Sage.

Gillespie, T. (2018). *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media*. New Haven, CT: Yale University Press.

Goggin, G., & McLelland, M. (2009). Internationalizing Internet studies: Beyond Anglophone paradigms. In G. Goggin & M. McLelland (Eds.), *Internationalizing Internet studies: Beyond Anglophone paradigms* (pp. 3–17). Abingdon, England: Routledge.

Hallinan, B., & Brubaker, J. (2021). Living with everyday evaluations on social media platforms. *International Journal of Communication*, *15*, 1551–1569.

Hitlin, S., & Piliavin, J. A. (2004). Values: Reviving a dormant concept. *Annual Review of Sociology*, *30*(1), 359–393. doi:10.1146/annurev.soc.30.012703.110640

Hoffmann, A. L., Proferes, N., & Zimmer, M. (2018). "Making the world more open and connected": Mark Zuckerberg and the discursive construction of Facebook and its users. *New Media & Society*, *20*(1), 199–218. doi:10.1177/1461444816660784

Hofstede, G. (2003). *Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Hollebeek, L. (2011). Exploring customer brand engagement: Definition and themes. *Journal of Strategic Marketing*, *19*(7), 555–573. doi:10.1080/0965254X.2011.599493

Hurley, Z. (2019). Imagined affordances of Instagram and the fantastical authenticity of female Gulf-Arab social media influencers. *Social Media + Society*, *5*(1). doi:10.1177/2056305118819241

Inglehart, R. (1990). *Culture shift in advanced industrial society*. Princeton, NJ: Princeton University Press.

Introna, L. D., & Nissenbaum, H. (2000). Shaping the web: Why the politics of search engines matters. *The Information Society*, *16*(3), 169–185. doi:10.1080/01972240050133634

Jenkins, H. (2006). *Fans, bloggers, and gamers: Exploring participatory culture*. New York: New York University Press.

John, N. A. (2016). *The age of sharing*. Malden, MA: Polity.

John, N. A., & Nissenbaum, A. (2019). An agnotological analysis of APIs: Or, disconnectivity and the ideological limits of our knowledge of social media. *The Information Society*, *35*(1), 1–12. doi:10.1080/01972243.2018.1542647

Knobel, C., & Bowker, G. C. (2011). Values in design. *Communications of the ACM*, *54*(7), 26. doi:10.1145/1965724.1965735

Lamont, M. (2012). Toward a comparative sociology of valuation and evaluation. *Annual Review of Sociology*, *38*(1), 201–221. doi:10.1146/annurev-soc-070308-120022

Lampe, C. (2011). The role of reputation systems in managing online communities. In H. Masum & M. Tovey (Eds.), *The reputation society: How online opinions are reshaping the offline world*. Cambridge, MA: MIT Press.

Langlois, G., McKelvey, F., Elmer, G., & Werbin, K. (2009). Mapping commercial web 2.0 worlds: Towards a new critical ontogenesis. *The Fibreculture Journal*, *14*.

Leaver, T., Highfield, T., & Abidin, C. (2020). *Instagram: Visual social media cultures*. Medford, MA: Polity Press.

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. doi:10.1093/jcmc/zmz025

Leurs, K., & Zimmer, M. (2017). Platform values: An introduction to the #AoIR16 special issue. *Information, Communication & Society*, 20(6), 803–808. doi:10.1080/1369118X.2017.1295464

Lewis, R. (2020). "This is what the news won't show you": YouTube creators and the reactionary politics of micro-celebrity. *Television & New Media*, 21(2), 201–217. doi: 10.1177/1527476419879919

MacCannell, D. (1973). Staged authenticity: Arrangements of social space in tourist settings. *American Journal of Sociology*, 79(3), 589–603.

Machin, D. (2007). *Introduction to multimodal analysis*. London, England: Hodder Arnold.

Marwick, A. E., & boyd, d. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, 13(1), 114–133. doi: 10.1177/1461444810365313

Myers West, S. (2018). Censored, suspended, shadowbanned: User interpretations of content moderation on social media platforms. *New Media & Society*, 20(11), 4366–4383. doi: 10.1177/1461444818773059

Nelson, J. L. (2019). The next media regime: The pursuit of "audience engagement" in journalism. *Journalism*, 1–18. doi:10.1177/1464884919862375

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York: New York University Press.

O'Brien, H., & Cairns, P. (Eds.). (2016). *Why engagement matters*. New York: Springer International Publishing. doi:10.1007/978-3-319-27446-1

Picone, I., Kleut, J., Pavlíčková, T., Romic, B., Møller Hartley, J., & De Ridder, S. (2019). Small acts of engagement: Reconnecting productive audience practices with everyday agency. *New Media & Society*, 21(9), 2010–2028. doi:10.1177/1461444819837569

Plantin, J.-C., Lagoze, C., Edwards, P. N., & Sandvig, C. (2018). Infrastructure studies meet platform studies in the age of Google and Facebook. *New Media & Society*, 20(1), 293–310. doi:10.1177/1461444816661553

Putnam, R. D. (2001). *Bowling alone: The collapse and revival of American community* (1st ed.). New York: Touchstone Books by Simon & Schuster.

Reade, J. (2020). Keeping it raw on the 'gram: Authenticity, relatability and digital intimacy in fitness cultures on Instagram. *New Media & Society*. doi:10.1177/1461444819891699

Rieder, B. (2020). *Engines of order: A mechanology of algorithmic techniques*. Amsterdam University Press.

Sagiv, L., Schwartz, S. H., & Arieli, S. (2011). Personal values, national culture and organizations: Insights applying the schwartz value framework. In N. M. Ashkanasy, C. P. M. Wilderom, & M. F. Peterson (Eds.), *The handbook of organizational culture and climate*, 2nd ed. (pp. 515–537). Thousand Oaks, CA: Sage Publications.

Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2014, May 22). *Auditing algorithms: Research methods for detecting discrimination on Internet platforms* [Preconference paper]. Seattle, WA: International Communication Association. Retrieved from https://pdfs.semanticscholar.org/b722/7cbd34766655dea10d0437ab10df3a127396.pdf

Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. *Advances in Experimental Social Psychology*, *25*, 1–65. doi:10.1016/s0065-2601(08)60281-6.

Schwartz, S. H. (2012). An overview of the Schwartz theory of basic values. *Online readings in psychology and culture*, *2*(1), 2–20. doi:10.9707/2307-0919.1116

Scissors, L., Burke, M., & Wengrovitz, S. (2016). What's in a Like? Attitudes and behaviors around receiving Likes on Facebook. Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing - CSCW '16, 1499–1508. doi: 10.1145/2818048.2820066

Seaver, N. (2017). Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big Data & Society*, *4*(2), 1–12. doi:10.1177/2053951717738104

Shifman, L. (2018). Testimonial rallies and the construction of memetic authenticity. *European Journal of Communication*, *33*(2), 172–184. doi:10.1177/0267323118760320

Shifman, L. (2019). Internet memes and the twofold articulation of values. In M. Graham & W. H. Dutton (Eds.), *Society and the Internet: How networks of information and communication are changing our lives*, 2nd ed. (pp. 43–57). Oxford, England Oxford University Press.

Taylor, C. (2018). *The ethics of authenticity*. Cambridge, MA: Harvard University Press.

Vaidhyanathan, S. (2018). *Antisocial Media: How Facebook disconnects us and undermines democracy*. Oxford, England: Oxford University Press.

Van Dijck, J., Poell T., & Waal, M. d. (2018). *The platform society: Public values in a connective world*. Oxford, England: Oxford University Press.

Van Dijk, T. A. (2000). *Ideology: A multidisciplinary approach*. London, England: Sage Publications.

Winner, L. (1986). *The whale and the reactor: A search for limits in an age of high technology*. Chicago, IL: University of Chicago Press.

Zuckerberg, M. (2019, March 6). *A privacy-focused vision for social networking*. Facebook. Retrieved from https://www.facebook.com/notes/mark-zuckerberg/a-privacy-focused-vision-for-social-networking/10156700570096634/