



A critical appraisal of diversity in digital knowledge production: Segregated inclusion on YouTube

new media & society

1–22

© The Author(s) 2021

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/14614448211034846

journals.sagepub.com/home/nms**Kaiping Chen** 

University of Wisconsin–Madison, USA

June Jeon 

Chungnam National University, South Korea

Yanxi Zhou

CVS Health, USA

Abstract

Diversity in knowledge production is a core challenge facing science communication. Despite extensive works showing how diversity has been undermined in science communication, little is known about to what extent social media augments or hinders diversity for science communication. This article addresses this gap by examining the profile and network diversities of knowledge producers on a popular social media platform—YouTube. We revealed the pattern of the juxtaposition of inclusiveness and segregation in this digital platform, which we define as “segregated inclusion.” We found that diverse profiles are presented in digital knowledge production. However, the network among these knowledge producers reveals the rich-get-richer effect. At the intersection of profile and network diversities, we found a decrease in the overall profile diversity when we moved toward the center of the core producers. This segregated inclusion phenomenon questions how inequalities in science communication are replicated and amplified in relation to digital platforms.

Corresponding author:

Kaiping Chen, University of Wisconsin–Madison, Madison, WI 53706, USA.

Email: kchen67@wisc.edu

Keywords

Digital knowledge production, diversity, network, profile, segregated inclusion, YouTube

Social media has become a major platform where people encounter, seek, and share science information (Brossard and Scheufele, 2013). As Funk et al. (2017) noted in the Pew report, a third of the Americans who actively seek and consume science news consider social media to be an important source. Extensive works have debated the role of social media in diversifying participation and perspectives (Burgess and Green, 2018; Choi and Cho, 2017). With the growth of social media, some science communication scholars hoped it would diversify the knowledge production process of science, in contrast to traditional sites of science communication, which is often dominated by groups of certain profiles.

Despite debates on how social media might facilitate or hinder diversity regarding civic participation, more scholarship is needed to investigate how social media can diversify science communication and what diversity means in the digital public sphere. What does diversity mean in digital platforms? To what extent diverse voices are produced and shared on such platforms? To address these questions, this article explicates the meaning of diversity on digital platforms. Specifically, we define and focus on two aspects of diversity—*profile diversity* and *network diversity*. Empirically, we analyzed content creation on YouTube to examine the issue of diversity in science communication. Compared to Facebook and Twitter, YouTube is much less studied despite its large audience size. YouTube has 2 billion monthly active users, is localized to 91 countries, and is accessible in 80 languages (Iqbal, 2020). YouTube is particularly suitable for content creation (Cunningham and Craig, 2019), beyond the mere circulation of knowledge (Dubovi and Tabak, 2020; Shifman, 2012); therefore, it provides a useful case study to understand online knowledge producers who actively create various types of knowledge, ranging from textbook scientific knowledge to public opinions based on life experiences. We collected and analyzed a novel dataset on digital knowledge production from YouTube starting from four controversial/politicized science topics: climate change, genetically modified organisms (GMOs), vaccination, and human genome editing. Through analyzing the citation network among YouTubers and the profile of a group of frequently cited YouTubers, we found that the YouTube community is still dominated by a few YouTubers who talk among themselves rather than to outsiders. At the intersection of profile and network diversities, we found that the most frequently cited “super connectors” tend to be YouTube channels connected to mainstream media; these channels are from profiles that are less likely to present gender and race, and explicitly claim to focus on teaching science topics to publics.

We conceptualize the above findings as *segregated inclusion in digital platforms*, the juxtaposition of (1) knowledge production by diverse populations, (2) the segregation of super connectors from the average content producers, and (3) the lack of profile diversity among those super connectors. This pattern of seemingly open but ultimately unequal

social space is not unique in digital platforms. In offline social settings, cultural sociologists also found “a form of inclusion in which new types of boundaries emerge between previously separate groups” (Accominotti et al., 2018). This form of inclusion happens when previously inaccessible social activities become accessible to the masses (e.g. classical music concerts). While there is a quantitative increase in newcomers (e.g. increase in subscribers), a new level of qualitative distinction occurs within the group (e.g. classification of seats) in tandem. The concept of “segregated inclusion” is useful to explain how the privilege of a powerful social group is maintained in an increasingly democratizing and inclusive society. As manifested in our examination of digital knowledge production, we demonstrated that inequality of the knowledge production process happens as well in digital platforms such as YouTube.

Our article provides two implications. First, it contributes to the study of civic participation on social media and science communication through revealing the affordances and challenges social media platforms pose to knowledge production. We provide an empirical examination of who the knowledge production communities are on social media. These producers heavily promote their own videos while sometimes interacting with each other through social capital such as business ties and shared life experiences. This examination of the knowledge producer community and their interaction behavior will advance our understanding of the information environment that heterogeneous publics encounter in the digital space. This new understanding will further inform practical implications on how to effectively communicate knowledge on social media platforms.

Second, this article offers a critical evaluation of diversity in digital platforms. Sociologists and science communication scholars have discussed how social inequality is correlated with a lack of diversity among legitimate knowledge producers. As concerned with both profile and network diversity, we offer a critical appraisal of diversity on YouTube. While there is an emphasis on the democratic production of knowledge to mitigate political, economic, and health inequalities, further questions on how the inclusive knowledge production process might look remain unanswered (Davies, 2011). We reveal that inequality can persist even in inclusive groups because the included people might be classified, stratified, and even excluded from a certain group of accredited people. In this way, inclusivity without a critical reflection upon *the network structure* of inclusiveness can create new forms of exclusion that are hidden and unrecognized by participants (Bourdieu, 1991; Butler, 2015). The pattern of segregated inclusion among YouTubers poses new questions on how inequality is organized in the field of knowledge—a social space where knowledge producers and consumers interact, compete, and evolve.

Studying diversity in knowledge production on social media

Theorizing diversity in knowledge production

Diversity in knowledge production is a core topic among science communication and social media scholars (Canfield et al., 2020; Dawson, 2018). We engage with science communication literature to examine knowledge production on social media for two

reasons. First, the lens of science communication helps us understand how various publics consume politicized science issues on social media. For instance, the COVID-19 pandemic has revealed how people's beliefs on science/health topics, such as wearing masks, vaccination, and even the existence of the disease itself, are intertwined with their social positions, political views, and identities (Ballew et al., 2020; Dunwoody, 2020; Scheufele et al., 2020). Second, the field of science communication stresses knowledge production by laypeople, beyond institutionally accredited experts. The meaning of knowledge is broader than scientific experts' textbook knowledge—it also includes narratives from people's life experiences, tacit knowledge, and local wisdom by community members (Kimura and Kinchy, 2019). In other words, knowledge is by no means uniform, and we need to understand who are producing knowledge and how they are curated in social media to enhance our understanding of what information the public encounter.

While science communication scholars use the term “knowledge producer” to describe those entities that produce knowledge or information in general, digital media scholars use “content creators” or “cultural intermediaries” to situate knowledge production in the context of social media affordances and culture. As Cunningham and Craig (2019) noted, social media entertainment is a new screen ecology. Platforms like YouTube intersect Hollywood and Silicon Valley where amateur content creators brand their contents through authenticity and community to compete for visibility and to express activism to challenge the power of platform (Cunningham and Craig, 2019). This content creator community has the potential to create a more diversified and multicultural production on digital platforms (Craig and Cunningham, 2019). We align with digital media scholars that knowledge production conflates with creator culture on social media, which foregrounds visibility and personality. The populations we study in this article include content creators, cultural intermediaries, and traditional media outlets. Furthermore, social media scholars are increasingly examining content creation in digital platforms run by diverse social groups (Bärtl, 2018; Bishop, 2019; Matias, 2019). Those new producers are aware of algorithmic rules and often strategically deploy themselves in social media for higher visibility and civic engagement. Given that, we think it would be useful to explicitly build a criterion of diversity to better evaluate such new producers. In this section, we theorize what diversity means, based on extensive literature from both social media studies and science communication.

We identify two distinctive ways to discuss the meaning of diversity in knowledge production: *profile diversity* and *network diversity*. *Profile diversity* denotes diversities of social identities and positions of knowledge producers. Early influences on this topic were feminist scholars (Haraway, 1988; Kloppenburg, 1991). Sociologists have argued that the lack of diversity in knowledge production is connected to the broader issue of civic participation in democracy (Kimura and Kinchy, 2019). Without inclusive science communication, knowledge claims from a margin, such as communities of people of color, women, or lay publics without formal degrees, are systematically ignored and muted (Mott and Cockayne, 2017). Authoritative knowledge by powerful social groups, the “weak objectivity,” can risk epistemic violence that “questions about who gets to be a credible witness, and to knowers' embodied situations” (Sweet, 2018: 223). In turn, scholars have experimented with various models to make science communication more inclusive, such as consensus conference between scientific experts and the lay publics

(Powell and Kleinman, 2008), a consortium of professional scientists and field practitioners to resolve ecological problems (Suryanarayanan and Kleinman, 2016), and community-based projects to facilitate “grounded knowledge” (Ashwood et al., 2014). These cases show how epistemic authorities by dominant knowledge producers can be renegotiated by broader contributors, and they further question how different platforms, such as social media, might be utilized toward this goal.

In relation, scholars have also discussed the network structure among accredited scientific experts, noting a lack of network diversity among knowledge producers who collectively achieve a scientific consensus. *Network diversity*, in this regard, means the degree to which the citation network among knowledge producers is dispersed. Robert Merton (1968) developed the concept Matthew effect to explain that accredited scientists receive more credits than others. Latour (1987) further argued that knowledge claims are legitimated via the cycle of citations among legitimated experts, who give credentials to each other. Recent network analyses further revealed that a scientific consensus can be portrayed as a status of the network structure, rather than contents of the debate (Shwed and Bearman, 2010). In other words, lack of network diversity is a defining condition of real-world scientific communities. Critical network analysts further elaborate that network structure in academic knowledge production is embedded within social power structures (Chakravarty et al., 2018).

When it comes to digital platforms, it is debatable whether such prevailing inequalities in knowledge production are replicated. On the one hand, social media have facilitated counter-publics forming social movements, thus stimulating marginalized voices in socio-political decision-making (Choi and Cho, 2017). One example is the Black Lives Matter movement on Twitter, within which counter-public narratives about racial profiling are expressed (Jackson and Foucault Welles, 2015). Furthermore, some scholars showed that social media can challenge the traditional hierarchy of knowledge production by stimulating engagement from amateur scientists such as on YouTube (Burgess and Green, 2018). Scholars also found that social media provides a platform to encounter diverse perspectives, offering people more opportunities to connect to strangers (Bakshy et al., 2015). When people become more involved with online discussions, they are more likely to be inadvertently exposed to political differences (Brundidge, 2010).

On the other hand, scholars have warned that social media may actually limit the diversity of participants and perspectives. The echo chamber phenomenon has been well documented; like-minded groups convene on Facebook, Twitter, and political blogs and only speak to people with a similar political ideology (Sunstein, 2001) or similar racial profile (Wimmer and Lewis, 2010). In addition to homophily, power-law distribution (i.e. rich-get-richer phenomenon) hinders social media’s ability to enhance diversity, presenting an extreme imbalance regarding who is heard or cited on the Internet. While some websites experience extreme popularity, the majority websites sink in the big web (Easley and Kleinberg, 2010: 18). This power-law problem has been found in social media platforms such as Reddit (Olson and Neal, 2015). Communication technologies are not immune to social inequalities (Habermas, 1991). Digital platforms are certainly more accessible to diverse populations, but the pattern of how knowledges are produced and shared is embedded within evermore opaque algorithmic processes, exemplifying

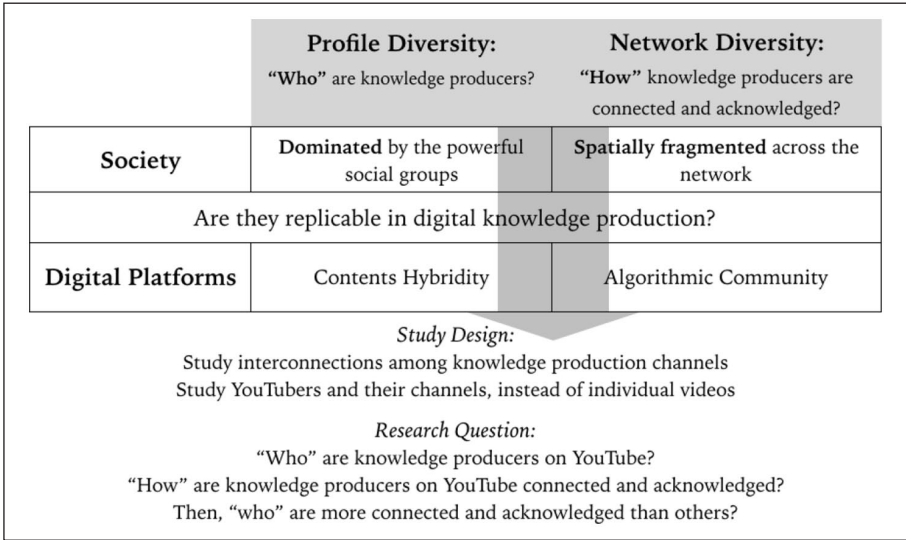


Figure 1. Conceptualize diversity and its application to digital platforms.

that power inequalities are inherently choreographed by existing social institutions (Stewart and Hartmann, 2020).

Connecting science communication research on unequal production of knowledge and social media research on diversity in digital platforms, this article aims to answer to what extent knowledge producers are diverse in a digital platform through investigating profiles and connectivity of knowledge producers. We contribute to science communication literature by revealing social media’s affordability for diversity in knowledge production. Our article also advances understanding about the participatory structure of social media by proposing and analyzing both profile and network diversity and the relationship between the two, thereby enhancing our examination of diversity beyond limited social categories.

Specificities of digital knowledge production—community and hybridity

To analyze the diversity of knowledge producers in a digital platform, how should we decide the unit of analysis (e.g. videos or YouTubers) and social characteristics (e.g. race, gender, or connectivity) of such unit? This is not merely a methodological question, but a theoretical question, as it requires a critical reflection upon distinctive characteristics of the digital knowledge production processes. Among others, we highlight two crucial characteristics, *algorithmic community* and *content hybridity* (Figure 1). Due to these two characteristics, we focus on profiles and connectivity of YouTube *channels*, not *individual videos*.

First, the notion of *algorithmic community* means that YouTube videos are inherently relational entities—they are connected, based on keywords, citations, and channel

owners (Burgess and Green, 2018). As stressed by Yoganarasimhan (2012), YouTube is a social arena for networking; therefore, it is a community of algorithmically interlinked videos, not a random collection of individual videos. Certain videos are more cited, circulated, and viewed because of reasons not reducible to the quality of individual videos. YouTubers know this well and strategically react in response to the algorithm—their approach to other YouTubers, based on personal relationships, similar life experiences, social capitals (e.g. fans), product promotion, or weak ties (Lange, 2019). Therefore, channels, not individual videos themselves, are the unit of analysis in our article.

Second, the notion of *content hybridity* means that it is difficult to specify a single genre in both channel owners and their videos. The boundaries between science videos and non-science videos, as well as science channels and non-science channels are vague. The videos we found with search terms such as climate change, GMOs, vaccination, and human genome editing were hybrid contents and even more so when we examined what links these videos cited—some were political campaigns, some were cosmetic advertisements, and some were religious speeches. The very fact that video contents are hybrids of multiple topics means that only focusing on “scientific knowledge-making” videos is, in fact, misleading. As channel owners produce various contents that could mention science-related content, audiences are cumulatively exposed to a certain degree of “knowledge” that channel owners provide to them. YouTube videos should be understood as not only a community but also a hybrid of contents.

Based on these two specificities, we studied both profile and network diversity among YouTube channels—who they are and how they are connected. To answer “*who*” *the knowledge producers are on YouTube*, we analyzed the profiles of YouTube channels that produced these knowledge-related hybrid contents. To answer “*how*” *knowledge producers in YouTube are connected and acknowledged*, we conducted network analysis to construct the network structure among knowledge producers on YouTube, based on their citation relationship. This examination uncovers the pattern of who are connected with one another and who are more acknowledged than others in YouTube’s algorithmic community. We found that knowledge production on YouTube might look inclusive at glance; however, it is ultimately segregated and stratified. YouTubers are segregated across the citation network. Among highly cited YouTubers, profiles with deracialized pictures, as well as mainstream media, tend to establish higher citations. We will discuss the theoretical and practical implications of the segregated inclusion pattern to social media and science communication studies in the “Discussion” section.

Data and method

Sampling design to study how knowledge producers are connected

We purposely chose topics to include historically politicized science topics (climate change, GMO, vaccination) and emerging technologies (human genome editing) that have been recently debated (Dunlop and Veneu, 2019). We started by collecting a random sample of highly viewed videos of these politicized and emerging science topics. These videos were published between January 2015 and November 2019. We developed the search terms for each science topic by drawing from existing literature and using

Table 1. Summary statistics of datasets.

Topic	No. of parent videos collected	No. of videos collected (10 depth)	No. of channels that produced these videos	% of videos that cited other videos in their video description	Average no. of videos cited by each video (median, max)
Climate change	1059	3650	857	63.32	(4, 43)
GMOs	640	2840	566	65.46	(3, 42)
Vaccination	300	1339	469	54.00	(1, 41)
Gene-editing	923	3159	872	63.32	(2, 18)

GMOs: genetically modified organisms.

computational methods (e.g. BERT) to generate new phrases to enrich our seed lexicon.¹ We used the YouTube Application Programming Interface to collect our dataset (ordered by view count), which provides video-level data (e.g. title, description, transcript, user clicks) and channel-level data (e.g. channel name, number of channel subscriptions, and channel profile). We also extracted all the YouTube links cited in a video’s description, crucial for building the citation network among the YouTubers. In this article, we use citation network to describe how YouTubers cited other YouTubers and/or themselves in their video description.

These initial videos are the parent videos, with contents closely related to the four science topics. Since we are interested in studying the network of knowledge producers, we then used snowball sampling to collect the children videos (cited in the video descriptions of the parent videos).² We iterated the snowball sampling process 10 times to collect all the descendants of the initial parent videos. The deeper we went, the less related a video’s content was to the four science topics we initiated; however, these descendant videos allowed us to go beyond the initial four science topics to include other topics related to the broad knowledge domains (such as videos that teach cooking, skincare, and music). Snowball sampling was crucial for us to identify how knowledge producers on YouTube are connected (Table 1).

Identifying frequently cited channels: “core producers”

To identify the frequently cited channels, we applied network analysis to the sampled videos. Network structures have two constituents: nodes, individual points in a network, and edges, the relations between points (Otte and Rousseau, 2002). For the edges, we used citations of videos in video descriptions to measure the connections between channels. We define a parent node as the one that cites other videos in its video description and a child node as the one that is cited. This gave us two types of networks, a channel-channel network, and a video-channel network. We developed four citation metrics to calculate the number of times a channel was cited: *channel-to-channel citation (self-citation included)*, *channel-to-channel citation (self-citation excluded)*, *video-to-channel citation (self-citation included)*, and *video-to-channel citation (self-citation excluded)*.³ Using these four metrics, we obtained the top 10 most-cited channels from each metric

and for each science topic. Because some channels were in the top list across topics, we had 77 channels left after removing duplicates. Among these 77 frequently cited channels (i.e. core producers), the mean number of subscriptions is 1,828,239 and the median is 115,000.

Content analysis of core producers' profiles

Our content analysis zooms into these 77 core producers to examine their profile, banner, and channel descriptions. They do not necessarily reflect their real identities, but rather how these knowledge producers present their identities to the audience.

Gender display. Among these 77 core producers, 43 are individual channels that allow us to examine their gender and racial display. We studied YouTuber's gender identification and display borrowing literature about gender studies on social media. Two interrelated dimensions were developed for content analysis. First, we align with scholars of gender and queerness that argue for treating gender as a non-binary and fluid concept to acknowledge gender variances and diversity (Bivens, 2017; Cromwell et al., 2006; Halberstam, 2019). For some digital platforms, such as MySpace, scholars can identify non-binary gender users when they explicitly state their gender (Boyd, 2008); however, it is not the case in YouTube. In cases where users use pronouns in their profile description, we use pronouns as an indicator of gender identification, as did in other social media studies that examine gender identity (Bivens and Haimson, 2016; Oakley, 2016). Therefore, the use of pronouns is the first dimension to examine how YouTubers identify their genders.

Second, we drew from the literature on impression management and gender performance on social media, which informs how researchers can use visual cues from profiles to examine gender display (Butkowski et al., 2020; Goffman, 1978; Goffman, 1979). As Szulc (2019) pointed out, a profile is a way to display one's identities and constitute the self. Similarly, Boyd (2008) argued that the choice of people's photos signals meaningful cues about themselves. For instance, the body people use to portray themselves on social media platforms is one angle to examine how the user identifies a particular gender category (Döring et al., 2016). Drawing on Goffman's theory of gender advertisements—the codes to reflect the variance of visual presentation in posts—scholars have developed coding categories to study the nuanced degree of femininity from people's gestures (Baker and Walsh, 2018; Butkowski et al., 2020). We followed these authors' operationalization such as coding the existence of self-touch, asymmetric posture, withdraw gaze, excessive emotional expression, and body exposure as our second dimension.⁴

Race display. To study racial display, we drew from the literature that examines race on social media through skin color. For instance, Kapidzic and Herring (2015) studied self-representations of race through teens' profile pictures from chat sites. In their study, they inferred race based on physical features such as skin color and coded race into Black, White, and other. Similarly, skin color has also been used by researchers to study race in magazine advertisements (Frith et al., 2005) and on television news (Dixon and Linz, 2000). Sociologists who study inequality (Lagos, 2019) emphasized the importance to focus on these "embodied characteristics" instead of self-reported identity to understand

race, which is critical for understanding inequality issues. She suggested that skin tone could be a robust predictor of racial and ethnic social stratification compared to self-reported categories.

Limitations. This project examines the gender and racial presentation of YouTubers. Our method does not identify the true gender and race of YouTubers. Rather, we studied how YouTubers presented themselves in their profiles, banners, and channel descriptions. It would be methodologically incorrect and ethically inappropriate to equate one's presentation with reality. Although we are interested in understanding the roles of multiple hierarchies in society including gender and race, this is not an intersectional project. According to the analytical framework of intersectionality, one's life experience is irreducible to a single or combinations of variables, such as race and gender (Crenshaw, 1989). Our content analysis on race and gender presentation does not represent a matrix of domination or inequality. It will be fruitful for future studies to utilize an intersectional framework on diverse groups' experience with digital science communication, supplemented by interview or survey materials. Finally, we acknowledge our personal biases in coding the gender and racial presentations of channels. We actively worked to ensure inter-coder reliability for a fair understanding of the channels.

Channel type and focus. While race and gender are the two most prominent indicators of individuals' social position, feminist scholars argue that social oppressions are also mediated by multiple factors such as class and education, which operate together in defining one's "particular social locations marked as outside the mainstream" (Choo and Ferree, 2010: 137). Social media scholars agree that competition between "amateur-made" contents and professionally generated contents (Burgess and Green, 2018) is as relevant as racial and gender competition in digital platforms. Therefore, we included YouTubers' channel type as additional variables that locate digital knowledge producers' social positions. We classified channels into four categories to see how many producers come from groups that traditionally hold the power of spreading information—academic, mainstream media, organization, and individuals. Channels were coded as academic if it has direct affiliations with research institutions. Channels operated by mainstream media companies were categorized as mainstream media. We also identified organization channels that are named after or run by established organizations or companies. In addition, we specified the channel focus as science-related or not from video descriptions. Two researchers coded an overlap random sample of core producers, and the average Cohen's kappa score is .83, indicating a high inter-coder agreement.⁵

Results

Profile diversity: feminine presence, imbalanced skin color distribution, and hybridity in channel type and focus

We have mixed findings about profile diversity among these frequently cited 43 individual producers. In terms of gender identification, we first examined pronouns and gender-specific names in their channel descriptions. Five channels used pronouns and 11

Table 2. Display of gender identification on YouTube.

Gender display	Frequency	Percentage of sample among presented
Pronouns (she/her)	2	40
Pronouns (he/his)	3	60
Gender-specific names (female)	7	50
Gender-specific names (male)	4	29
Gender-specific names (both)	3	21
Self-touch	8	24
Asymmetrical posture	17	49
Withdraw gaze	13	38
Excessive emotional expression	7	22
Body exposure	8	26

channels used gender-specific names. Among those that used pronouns, 40% used she/her and 60% used he/his. Similarly, 50% used gender-specific female names, and 29% used male names (Table 2.). While gender is a non-binary concept, this finding shows a strong feminine presence among the individual knowledge producers. Second, we examined gender presentation from visual cues in profile pictures and banners. Among the channels that display pictures with human subjects, we observed that between 20% and 50% of channels present feminine elements in their pictures. Asymmetrical posture is the most frequent (49%) among these feminine presentation categories.

Race imbalance presented by skin color in profile pictures and banners still exists among core producers. Among those with skin color presented, only 10% of the profile pictures display people of color, while a large segment displays White people (39%). This pattern can also be seen in banners where most producers are shown to be White (22%), while only a few are presented as people of other skin colors (6%) (Figure 2).

While profile differences are often measured by gender and racial compositions, examining other components of diversity can provide a more comprehensive understanding of digital knowledge producers on YouTube. We found hybridity in channel type and focus. A significant portion of producers is classified as individual channels (56%). Only 10% of these core producers are academics associated with research institutions, and both mainstream media and organization core producers are 17%. This widening participation from non-academic and non-legacy media actors brings a new way of communicating knowledge. For topics that these producers stated as their channel’s focus, 31% of channels presented their focus on science topics, while 69% on other topics that are not science-focused (Figure 3).

Segregation of knowledge production on YouTube

In terms of network diversity, we found that the majority of channels across our datasets were cited very few times. For example, regarding the topic of human genome editing, 71% of the channels were only cited by one video, and 89% of the channels were cited

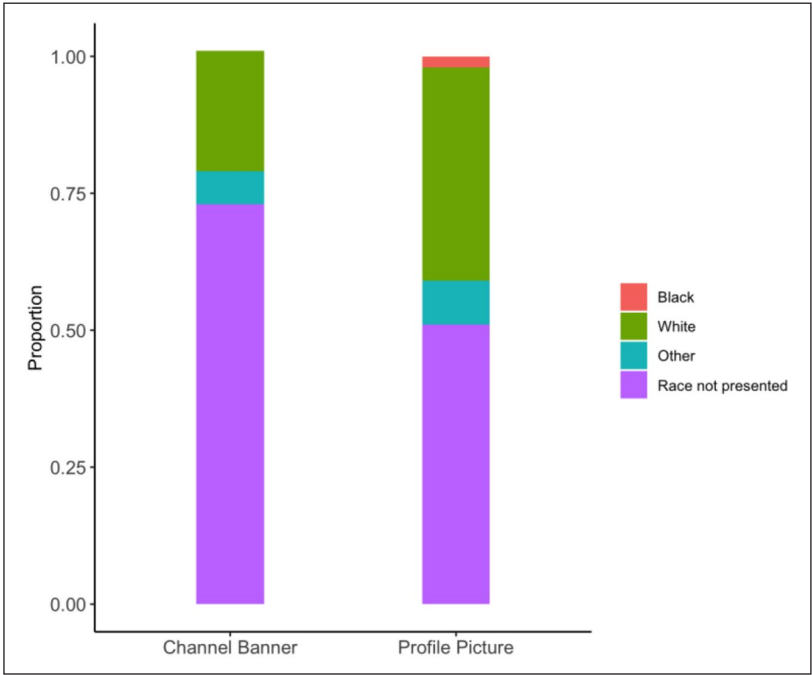


Figure 2. Display of race by skin color in channel banners and profile pictures.

by one channel (Figures 4 and 5). In contrast, the channel cited by most videos in the topic, “Cullen and Katie,” was cited by 171 videos about genome editing.⁶ This heavily skewed distribution of citation suggests that the network of knowledge production on YouTube is largely dominated by a few channels, while the majority remains unnoticed. As shown in the channel network graph of all four science topics (Figure 6) where a node’s size represents the number of citations a channel received, only a small group of producers is at the core of the network, being cited among each other more often than the rest. Examining who the core producers are, many of them are initiated by the PBS Digital Studio to build a digital producer community that serves to educate audiences on different topics such as history, culture, and science (PBS Digital Studios, 2020). These results suggest evidence about the segregation of knowledge production on this digital platform.

Profile diversity at the core of the network: a segregated inclusion

As described above, profile diversity and network diversity on YouTube demonstrate a pattern that diverse YouTubers, in terms of feminine presence, channel type, and focus, are in fact clustered throughout the network, overshadowing thousands of other YouTubers. To examine the relationship between network and profile diversity, we analyzed how these highly cited 77 core producers can be differentiated regarding their

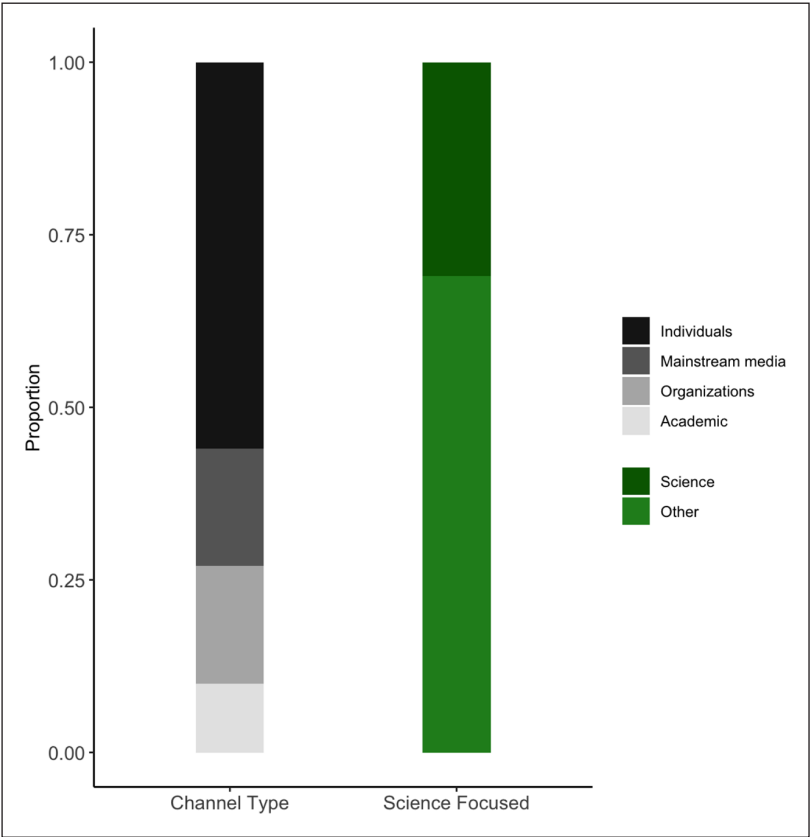


Figure 3. Channel type and focus.

connections *within* the core producers and their association with profile diversity. We found that 36 out of 77 YouTubers are cited less than twice by other YouTubers (group 1) and 16 “super connectors” are cited by more than five other channels (group 3).

We also noticed an overall decrease in the profile diversity when we moved toward the center of these core producers. Specifically, the super connectors (group 3) have a much higher proportion of channels that are mainstream media and stated the channel focus on science, compared to the other two groups (Table 3). This indicates a structural inequality between independent producers and corporate media organizations—the channels cited frequently are more likely to be mainstream media rather than channels with user-generated content. In terms of gender and race presentation, we compared group 1 and group 2 as there are very few individual channels in group 3 (Table 4). We found that fewer channels used pronouns to identify their gender in group 2 (14%) compared to group 1 (44%). Among the channels that used pronouns to describe themselves, group 2 has a much lower proportion of use of female pronouns compared to group 1. This suggests that channels that have a larger network size tend to use fewer pronouns and female pronouns.

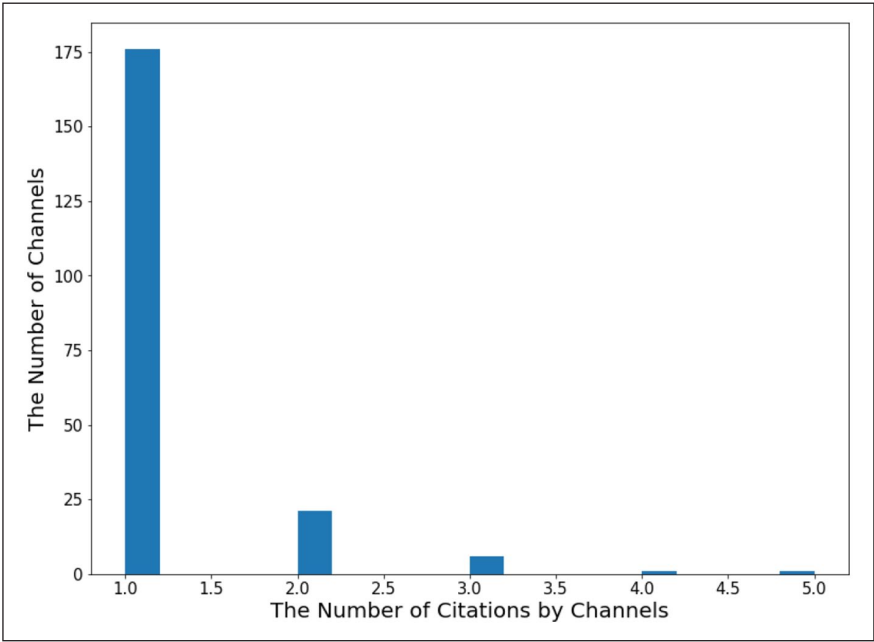


Figure 4. The number of citations by channels excluding self-citations (Gene Editing).

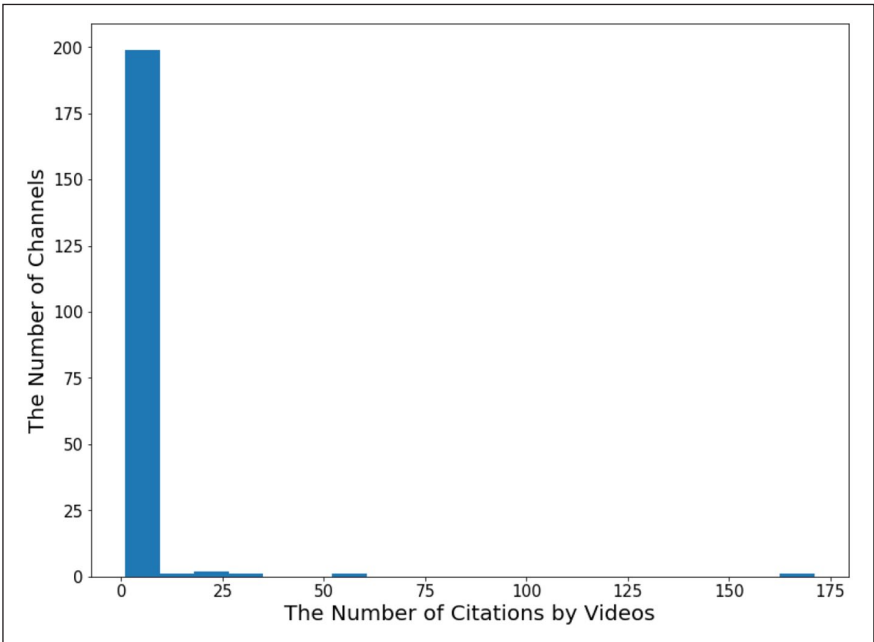


Figure 5. The number of citations by videos excluding self-citations (Gene Editing).

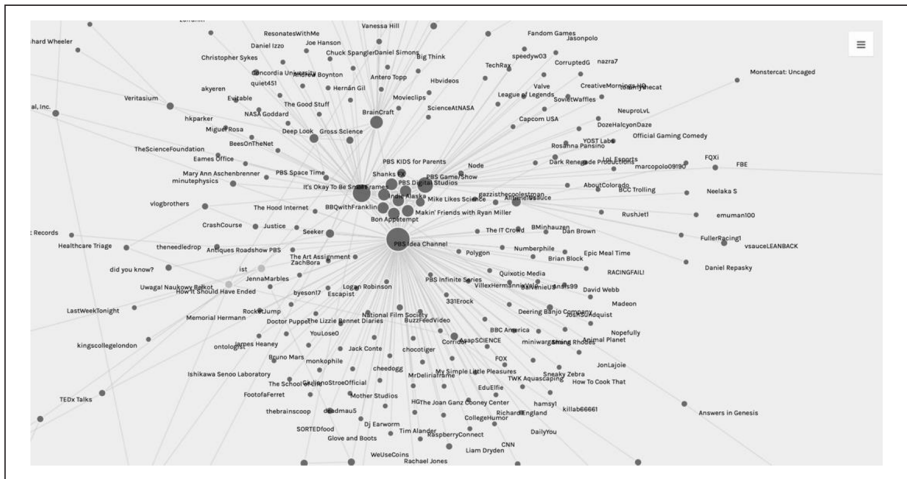


Figure 6. Channel-channel network graph among knowledge producers. Figure 6 is the central portion of the complete network, generated with Palladio.

Table 3. Relationship between profile diversity and network diversity: channel type and focus.

	$N \leq 1$ (36 channels)	$2 \leq N \leq 5$ (25 channels)	$N > 5$ (16 channels)
Channel type	Individual: 21 Academic: 4 Mainstream media: 1 (3%) Organization: 10	Individual: 18 Academic: 4 Mainstream media: 2 (8%) Organization: 1	Individual: 4 Academic: 0 Mainstream media: 10 (63%) Organization: 2
Channel focus	Science-related: 10 (28%)	Science-related: 8 (32%)	Science-related: 6 (38%)

In terms of racial display, there is a much smaller proportion of channels that display profiles as another skin color (not White, not Black) in group 2 (5%) compared to group 1 (16%). We also observed that group 2 has a higher proportion of channels that do not present skin color (i.e. color-blind) in their profiles and banners (59%) compared to group 1 (39%), indicating that “color-blind” presentation consistently increased as these core producers are more connected to others. This is a meaningful pattern to report because sociologists of race and ethnicity have argued that racism can be explained by the increasing trend of colorblindness, when powerful social groups increasingly adopt deracialized rhetoric to justify inequality (Bonilla-Silva, 2006; Mueller, 2017). The fact that super connectors’ race is predominantly not presented implies that they might be predominantly White YouTubers who strategically deracialized their profiles, portraying their brand logo as a profile and banner pictures instead of personalized profiles.

Table 4. Relationship between profile diversity and network diversity: gender and race presentation.

	<i>N</i> ≤ 1 (36 channels)	2 ≤ <i>N</i> ≤ 5 (25 channels)	<i>N</i> > 5 (16 channels)
Gender identification	<p>There are 25 individuals and academics for whom we were able to code the gender identification:</p> <p>(1) Pronouns/gendered name</p> <ul style="list-style-type: none">• 11 (44%) people use pronouns or gender-specific names to identify themselves• 7 (28%) people use female pronouns/ names	<p>There are 22 individuals and academics for whom we were able to code the gender identification:</p> <p>(1) Pronouns/gendered names</p> <ul style="list-style-type: none">• 3 (14%) people use pronouns/gender-specific names to identify themselves• 1 (5%) person uses female pronouns/ names	<p>There are 4 individuals and academics for whom we were able to code the gender identification:</p> <p>(1) Pronouns/ gendered names</p> <ul style="list-style-type: none">• 2 (50%) people use pronouns/gender-specific names to identify themselves• 1 (25%) person uses female pronouns/ names
Skin color	<p>There are 25 individuals and academics for whom we were able to code the skin color to study race display:</p> <ul style="list-style-type: none">• 4 (16%) are other skin colors• 9 (36%) are NAs (skin color not presented)• 12 (48%) are White• No person is of Black tone	<p>There are 22 individuals and academics for whom we were able to code the skin color to study race display:</p> <ul style="list-style-type: none">• 1 (5%) is other skin colors• 13 (59%) are NAs (skin color not presented)• 8 (36%) are White• No person is of Black tone	<p>There are 4 individuals and academics for whom we were able to code the skin color to study race display:</p> <ul style="list-style-type: none">• 3 (75%) are White• 1 person is of Black tone (25%)

Discussion and conclusion

This article analyzed who the knowledge producers are and the connectivity among them on the YouTube platform to critically assess diversity in digital knowledge production. For this purpose, we defined profile diversity and network diversity as two different but related aspects of diversity. Profile diversity denotes diversities of racial and gender identities and social positions of knowledge producers. Network diversity refers to the degree to which the citation network among knowledge producers is dispersed. Using random and snowball sampling methods, we collected YouTube channels starting from science-related topics and measured profile and network diversities among those channels and their cited channels. We revealed the phenomena of segregated inclusion, which means that (1) diverse profiles, regarding channel type and focus, are presented in digital knowledge production among 77 frequently cited knowledge producers; (2) except these 77 YouTubers, thousands of other YouTubers are less connected in the network; and (3) among these 77 YouTubers, the more the channels are cited by other channels, the lower the profile diversity is observed. These findings demonstrate that by focusing on those at

the core, the illusion of diversity is dispelled—the critical network positions of super-strong knowledge producers are unequally taken by those who already are overrepresented in knowledge production. These findings shed new light on the emerging pattern of juxtaposed inclusiveness and segregation in digital platforms—*segregated inclusion*.

Our findings provide several implications for studies of science communication, social inequality, and digital platforms. First, this study flags that scholars on the democratization of knowledge production should study digital platforms, more than ever. This is because the current social patterns can be amplified via digital platforms, operated by capitalistic firms' profit-making algorithm. Social media scholars have shown that social media sites are not simply open public spheres, but platforms where interpersonal interactions are organized and mediated via particular algorithmic rules of the service (Gorwa, 2019; Jackson and Foucault Welles, 2015). In other words, digital platforms can structure enduring patterns of inequality, embedded within the system shaping how users and contents of the platform are displayed and organized. Particularly, YouTube is a profit-making platform where users' algorithm literacy (i.e. awareness of the algorithms' functions and ability to consciously interact with them) critically affects the channel's success and failure (Gran et al., 2020). Our findings on segregated inclusion suggest that winners and losers in such platform economy should be further scrutinized to find a connection in how social inequality is compounded with inequality in digital platforms. We would also like to acknowledge that as the COVID-19 pandemic evolves, there has been increasing participation in the discussion of immunization strategies and vaccines. Future research can further examine how the patterns of segregated inclusion change as digital content creators react to critical social events such as the pandemic. It will also be fruitful to investigate the patterns of digital knowledge production in other social media platforms and country contexts.

Second, our findings on segregated inclusion imply that YouTube users might be exposed to an invisible form of unequal knowledge production pattern. By “invisible,” we mean that YouTube users are not in a position to comprehend a network structure of inequality in the platform; therefore, they become a part of this trend. In other words, if YouTube users search science topics and watch every video, then they will have a chance to be exposed to various voices from diverse channels; however, in reality, YouTube users rely on YouTube's algorithm that prioritizes super connectors. As a result, YouTube users would likely be exposed to videos from less diverse video producers who are more likely to be aligned with our society's power holders. To be clear, digital platforms' racial and gender biases are not new, thus not surprising. Scholars have reported how a video's racial and gender characteristics can affect the popularity it receives and how such characteristics are connected to the amplification of racial and gender prejudices (Amarasekara and Grant, 2018; Matamoros-Fernández, 2017). As shown in the research by Amarasekara and Grant (2018), female creators on YouTube received more comments that are hostile, critical, sexual, and appearance-related compared to male hosts. Other studies also call attention to the prevalence of “digital sexism” manifested by humiliation and shaming targeted toward women (Sobieraj, 2018). Although we are increasingly aware of such explicit prejudice in digital platforms, less is known about how the larger network among users, content creators, and their channels represents a structure of unequal relationships. Given that the network structure on YouTube represents the segregated inclusion of

knowledge producers, YouTube might be a stratified social world where participants can hardly recognize their unequal playground. Our findings echo with other scholarship that notes YouTube as a tiered and classified social space where rules of the game are disproportionately applied to its users (Caplan and Gillespie, 2020). When such inequality is only visible on a macro scale but invisible for users, YouTube might be a platform that generates the illusion of equality—a false belief that digital platforms provide equality to all.

Finally, this article also provides policy implications regarding inequality in digital platforms. As mentioned above, selective groups with higher algorithm literacy are strategically citing other channels (and themselves) to elevate their exposure rates in the platform, given that YouTube is an algorithmic community. Those channels not only display their knowledge on YouTube more effectively but also make enormous profits. The inequality issue in content creation might be further affected with the introduction of the Digital Single Market Directive in 2019 (Reynolds, 2019). As Susan Wojcicki, CEO of YouTube, is concerned, Article 13 could have a long-term impact on the platform as it is hard to design an optimal upload filter to decide copyright violation by content creators, which can have downstream effects on who has the power to share content (Chmielewski, 2018). Thus, policymakers and algorithm programmers should consider designing platform policies and algorithms that can prevent a cycle of rich-get-richer. According to our findings, digital platforms have yet to fully realize the democratic potentials of the Internet space. In the face of structural and policy challenges, the goal for a more just and democratic algorithm might not be unachievable—we are just at the beginning.

Acknowledgements

We are grateful for the valuable suggestions we received at different stages of this manuscript from Eunsil Oh, Amy Gaeta, Myria Georgiou, Ellen Helsper, and Nick Couldry. We are also thankful for the discussion from participants at the Life Sciences Communication Department Colloquium (LSC700) and the Robert F. and Jean E. Holz Center at the University of Wisconsin–Madison, and the Virtual Symposium Platformed Creation organized by the Stanford University Ethnography Lab.

Authors Contribution

All authors contributed equally to this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: K.C. would like to thank for the funding support received from the University of Wisconsin–Madison Vice Chancellor for Research and Graduate Education.

ORCID iDs

Kaiping Chen  <https://orcid.org/0000-0002-5864-5333>

June Jeon  <https://orcid.org/0000-0002-5440-1227>

Supplemental material

Supplemental material for this article is available online.

Notes

1. For details on our search word development, see Supplemental Material I.
2. For more explanation on why we used snowball sampling, see Supplemental Material II.
3. For details on how we constructed our network metrics among producers, see Supplemental Material III.
4. For detailed operationalization, see Supplemental Material V.
5. See Supplemental Material VI for inter-coder agreement score for each content analysis variable.
6. See Supplemental Material IV for similar citation patterns across all of our science topics.

References

- Accominotti F, Khan SR and Storer A (2018) How cultural capital emerged in Gilded Age America: musical purification and cross-class inclusion at the New York Philharmonic. *American Journal of Sociology* 123(6): 1743–1783.
- Amarasekara I and Grant WJ (2018) Exploring the YouTube science communication gender gap: a sentiment analysis. *Public Understanding of Science* 28(1): 68–84.
- Ashwood L, Harden N, Bell MM, et al. (2014) Linked and situated: grounded knowledge. *Rural Sociology* 79(4): 427–452.
- Baker SA and Walsh MJ (2018) “Good Morning Fitfam”: top posts, hashtags and gender display on Instagram. *New Media & Society* 20(12): 4553–4570.
- Bakshy E, Messing S and Adamic LA (2015) Exposure to ideologically diverse news and opinion on Facebook. *Science* 348(6239): 1130–1132.
- Ballew M, Bergquist P, Goldberg M, et al. (2020) *American Public Responses to COVID-19*. Yale University and George Mason University, Yale Program on Climate Change Communication. Available at: <https://climatecommunication.yale.edu/wp-content/uploads/2020/04/american-public-responses-covid19-april-2020b.pdf>
- Bärtl M (2018) YouTube channels, uploads and views: a statistical analysis of the past 10 years. *Convergence* 24(1): 16–32.
- Bishop S (2019) Managing visibility on YouTube through algorithmic gossip. *New Media & Society* 21(11–12): 2589–2606.
- Bivens R (2017) The gender binary will not be deprogrammed: ten years of coding gender on Facebook. *New Media & Society* 19(6): 880–898.
- Bivens R and Haimson OL (2016) Baking gender into social media design: how platforms shape categories for users and advertisers. *Social Media+ Society* 2(4): 1–12.
- Bonilla-Silva E (2006) *Racism without Racists: Color-Blind Racism and the Persistence of Racial Inequality in the United States*. Lanham, MD: Rowman & Littlefield.
- Bourdieu P (1991) *Language and Symbolic Power*. Cambridge, MA: Harvard University Press.
- Boyd D (2008) Why youth (heart) social network sites: the role of networked publics in teenage social life. In: Buckingham D (ed.) *Youth, Identity, and Digital Media*. Cambridge, MA: Harvard University Press, pp. 2007–2016.
- Brossard D and Scheufele DA (2013) Science, new media, and the public. *Science* 339(6115): 40–41.
- Brundidge J (2010) Encountering “difference” in the contemporary public sphere: the contribution of the Internet to the heterogeneity of political discussion networks. *Journal of Communication* 60(4): 680–700.
- Burgess J and Green J (2018) *YouTube: Online Video and Participatory Culture*. New York: John Wiley & Sons.

- Butkowski CP, Dixon TL, Weeks KR, et al. (2020) Quantifying the feminine self(ie): gender display and social media feedback in young women's Instagram selfies. *New Media & Society* 22(5): 817–837.
- Butler J (2015) *Notes toward a Performative Theory of Assembly*. Cambridge, MA: Harvard University Press.
- Canfield KN, Menezes S, Matsuda SB, et al. (2020) Science communication demands a critical approach that centers inclusion, equity, and intersectionality. *Frontiers in Communication* 5: 2.
- Caplan R and Gillespie T (2020) Tiered governance and demonetization: the shifting terms of labor and compensation in the platform economy. *Social Media + Society* 6(2): 1–13.
- Chakravartty P, Kuo R, Grubbs V, et al. (2018) #CommunicationSoWhite. *Journal of Communication* 68(2): 254–266.
- Chmielewski D (2018) YouTube CEO Susan Wojcicki warns against “unintended consequences” of EU's Article 13. Available at: <https://deadline.com/2018/11/youtube-ceo-susan-wojcicki-warns-unintended-consequences-article-13-1202500444/> (accessed 21 March 2021)
- Choi SY and Cho Y (2017) Generating counter-public spheres through social media: two social movements in neoliberalised South Korea. *Javnost-The Public* 24(1): 15–33.
- Choo HY and Ferree MM (2010) Practicing intersectionality in sociological research: a critical analysis of inclusions, interactions, and institutions in the study of inequalities. *Sociological Theory* 28(2): 129–149.
- Craig D and Cunningham S (2019) *Social Media Entertainment: The New Intersection of Hollywood and Silicon Valley*. New York: NYU Press.
- Crenshaw K (1989) Demarginalizing the intersection of race and sex: a black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. *University of Chicago Legal Forum* 1989(1): 8.
- Cromwell J, Stryker S and Whittle S (2006) Queering the binaries. In: Stryker S and Whittle S (eds) *The Transgender Studies Reader*. New York: Routledge, pp. 509–520.
- Cunningham S and Craig D (2019) Creator governance in social media entertainment. *Social Media + Society* 5(4): 1–11.
- Davies SR (2011) The rules of engagement: power and interaction in dialogue events. *Public Understanding of Science* 22(1): 65–79.
- Dawson E (2018) Reimagining publics and (non) participation: exploring exclusion from science communication through the experiences of low-income, minority ethnic groups. *Public Understanding of Science* 27(7): 772–786.
- Dixon TL and Linz D (2000) Overrepresentation and underrepresentation of African Americans and Latinos as lawbreakers on television news. *Journal of Communication* 50(2): 131–154.
- Döring N, Reif A and Poeschl S (2016) How gender-stereotypical are selfies? A content analysis and comparison with magazine adverts. *Computers in Human Behavior* 55: 955–962.
- Dubovi I and Tabak I (2020) An empirical analysis of knowledge co-construction in YouTube comments. *Computers & Education* 156: 103939.
- Dunlop L and Veneu F (2019) Controversies in science. *Science & Education* 28(6): 689–710.
- Dunwoody S (2020) Science journalism and pandemic uncertainty. *Media and Communication* 8(2): 471–474.
- Easley D and Kleinberg J (2010) Power laws and rich-get-richer phenomena. In: Easley D and Kleinberg J (eds) *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*. Cambridge: Cambridge University Press, pp. 543–560.
- Frith K, Shaw P and Cheng H (2005) The construction of beauty: a cross-cultural analysis of women's magazine advertising. *Journal of Communication* 55(1): 56–70.
- Funk C, Jeffrey G and Mitchell A (2017) *Science News and Information Today*. Washington, DC: Pew Research Center.

- Goffman E (1978) *The Presentation of Self in Everyday Life*. Harmondsworth: Penguin Books.
- Goffman E (1979) *Gender Advertisements*. New York: Harper & Row.
- Gorwa R (2019) What is platform governance? *Information, Communication & Society* 22(6): 854–871.
- Gran AB, Booth P and Bucher T (2020) To be or not to be algorithm aware: a question of a new digital divide? *Information, Communication & Society*. Epub ahead of print 20 August. DOI: 10.1080/1369118X.2020.1736124.
- Habermas J (1991) *The Structural Transformation of the Public Sphere: A Inquiry into a Category of Bourgeois Society*. Cambridge, MA: MIT Press.
- Halberstam J (2019) *Female Masculinity*. Durham, NC: Duke University Press.
- Haraway D (1988) Situated knowledges: the science question in feminism and the privilege of partial perspective. *Feminist Studies* 14(3): 575–599.
- Iqbal M (2020) *YouTube Revenue and Usage Statistics (2020)*. Available at: <https://www.businessofapps.com/data/youtube-statistics/> (accessed 12 January 2021).
- Jackson SJ and Foucault Welles B (2015) Hijacking #myNYPD: social media dissent and networked counterpublics. *Journal of Communication* 65(6): 932–952.
- Kapidzic S and Herring SC (2015) Race, gender, and self-presentation in teen profile photographs. *New Media & Society* 17(6): 958–976.
- Kimura A and Kinchy A (2019) *Science by the People*. New Brunswick, NJ: Rutgers University Press.
- Kloppenborg J (1991) Social theory and the de/reconstruction of agricultural science: local knowledge for an alternative agriculture. *Rural Sociology* 56(4): 519–548.
- Lagos D (2019) Hearing gender: voice-based gender classification processes and transgender health inequality. *American Sociological Review* 84(5): 801–827.
- Lange P (2019) *Thanks for Watching: An Anthropological Study of Video Sharing on YouTube*. Boulder, CO: University Press of Colorado.
- Latour B (1987) *Science in Action : How to Follow Scientists and Engineers through Society*. Cambridge, MA: Harvard University Press.
- Matamoros-Fernández A (2017) Platformed racism: the mediation and circulation of an Australian race-based controversy on Twitter, Facebook and YouTube. *Information, Communication & Society* 20(6): 930–946.
- Matias JN (2019) The civic labor of volunteer moderators online. *Social Media + Society* 5(2): 1–12.
- Merton RK (1968) The Matthew effect in science: the reward and communication systems of science are considered. *Science* 159(3810): 56–63.
- Mott D and Cockayne D (2017) Citation matters: mobilizing the politics of citation toward a practice of conscientious engagement. *Gender, Place & Culture* 24(7): 954–973.
- Mueller JC (2017) Producing colorblindness: everyday mechanisms of white ignorance. *Social Problems* 64(2): 219–238.
- Oakley A (2016) Disturbing hegemonic discourse: nonbinary gender and sexual orientation labeling on Tumblr. *Social Media+ Society* 2(3): 1–12.
- Olson RS and Neal ZP (2015) Navigating the massive world of Reddit: using backbone networks to map user interests in social media. *PeerJ Computer Science* 1: e4.
- Otte E and Rousseau R (2002) Social network analysis: a powerful strategy, also for the information sciences. *Journal of Information Science* 28(6): 441–453.
- PBS Digital Studios (2020) *PBS Digital Studios*. Available at: <https://www.pbs.org/digital-studios/> (accessed 29 June 2021).

- Powell M and Kleinman DL (2008) Building citizen capacities for participation in nanotechnology decision-making: the democratic virtues of the consensus conference model. *Public Understanding of Science* 17(3): 329–348.
- Reynolds M (2019) What is Article 13? The EU's divisive new copyright plan explained. *WIRED*. Available at: <https://www.wired.co.uk/article/what-is-article-13-article-11-european-directive-on-copyright-explained-meme-ban> (accessed 6 April 2021).
- Scheufele D, Krause N, Freiling I, et al. (2020) How not to lose the COVID-19 communication war. *Issues in Science and Technology* 17. Available at: <https://slate.com/technology/2020/04/covid19-misinformation-science-communication.html>
- Shifman L (2012) An anatomy of a YouTube meme. *New Media & Society* 14(2): 187–203.
- Shwed U and Bearman PS (2010) The temporal structure of scientific consensus formation. *American Sociological Review* 75(6): 817–840.
- Sobieraj S (2018) Bitch, slut, skank, cunt: patterned resistance to women's visibility in digital publics. *Information, Communication & Society* 21(11): 1700–1714.
- Stewart E and Hartmann D (2020) The new structural transformation of the public sphere. *Sociological Theory* 38(2): 170–191.
- Sunstein CR (2001) *Republic.com*. Princeton, NJ: Princeton University Press.
- Suryanarayanan S and Kleinman D (2016) *Vanishing Bees: Science, Politics, and Honeybee Health*. New Brunswick, NJ: Rutgers University Press.
- Sweet PL (2018) The feminist question in realism. *Sociological Theory* 36(3): 221–243.
- Szulc L (2019) Profiles, identities, data: making abundant and anchored selves in a platform society. *Communication Theory* 29(3): 257–276.
- Wimmer A and Lewis K (2010) Beyond and below racial homophily: ERG models of a friendship network documented on Facebook. *American Journal of Sociology* 116(2): 583–642.
- Yoganarasimhan H (2012) Impact of social network structure on content propagation: a study using YouTube data. *Quantitative Marketing and Economics* 10(1): 111–150.

Author biographies

Kaiping Chen is an Assistant Professor in Computational Communication at the Department of Life Sciences Communication at the University of Wisconsin–Madison. Kaiping's research applies data science methods and interviews to study public deliberation, science communication, and information campaign in the United States, China, and Ghana. Kaiping's works have been supported by the National Science Foundation and were published or forthcoming in flagship journals across disciplines including *American Political Science Review*, *Journal of Communication*, *New Media & Society*, *Public Opinion Quarterly*, *Public Understanding of Science*, *International Public Management Journal*, *Harvard Kennedy School Misinformation Review*, and *PNAS*, and among other peer-reviewed journals. For more information, see her website: <https://lsc.wisc.edu/facstaff/chen-kaiping/>.

June Jeon received his Ph.D. in sociology at the University of Wisconsin–Madison. June will join the Department of Sociology at the Chungnam National University as an Assistant Professor in summer 2021. His research intersects science, inequality, and social theory. His works have been published in journals such as *Social Studies of Science*, *Engaging Science, Technology, and Society*, and *Agriculture and Human Values*.

Yanxi Zhou is a Data Scientist at CVS Health. She received her bachelor's degree in communication from the University of Wisconsin–Madison and her master's degree in statistical practice from Carnegie Mellon University.