The view from somewhere: Critically reflexive recruitment heuristics for big data
by Corinne Jones

Abstract
Recent scholarship has raised important critical questions about the ethical uses of big data scraped from social media platforms. Critical and feminist researchers have argued that big data can naturalize a “view from nowhere,” which ultimately reinscribes the status quo. In response, researchers have sought to incorporate critical reflexivity into their big data research and to include participants in their research through surveys and interviews. With surveys and interviews, researchers must engage in critically reflexive recruitment practices though, as reaching out to participants always reveals a “view from somewhere” that can (but does not always) function in the milieu of the dominant gaze. This article builds on scholarship that calls for more ethical data practices by offering three heuristics for critical reflexive recruitment practices. First, researchers should consider whether to recruit participants based on shifting boundaries of publicity and privacy. They should evaluate if posts were meant for their particular “view from somewhere.” Second, researchers should assess the names of research accounts, and how the names might be experienced as extractive. Finally, researchers should critically reflect on channels of recruitment and how those channels function in public/private binaries.

Contents
Introduction
Internet research and feminism
Case study
Discussion: Recruitment heuristics
Conclusions

Introduction
As digital tools to collect large datasets from social media platforms become more widely available and easier to use, scholars have argued that digital humanities researchers should contextualize and humanize big data, rather than rely on positivistic paradigms that imagine those data as neutral. Scholars who situate their work in feminist approaches specifically argue that critical and ongoing self-reflection is paramount (Cooky, et al., 2018; D'Ignazio and Klein, 2020; Leurs, 2017; Luka and Millette, 2018; Markham, 2006; Markham, et al., 2018), and many have sought to include participants in their research through surveys and interviews.
Following these trends, this article takes up questions about publicity and privacy in the recruitment phase of big data research. I focus on how I reflexively reoriented to data collected in the recruitment stage of my dissertation research, especially as it related to gazing in at people’s experiences for whom I may not have been the intended audience. Specifically, I found myself continuously reflecting on my “view from somewhere.” I navigated questions about publicity, the flow of information, and how the act of reaching out might be experienced by different participants. I understand these considerations to be integral to critically reflexive recruitment practices, and I hope that future researchers will draw from this work and critically consider their own “view from somewhere” and how it might be experienced by participants. Overall, this article echoes Abidin’s (2020) point that researchers must negotiate complicated issues around visibility along multiple spectrums, and the article picks up Markham, et al.’s (2018) call to develop practical, less abstract ethical heuristics for big data research.

Ultimately, (1) researchers might usefully consider if to reach out to certain participants as privacy shifts, and they should critically consider how some participants may perform in ways not meant for the researcher. (2) They should also critically reflect on appropriate account names, and (3) the appropriate channels through which to reach out to participants. These considerations are important for critically reflexive recruitment, in which researchers must account for different positions of power and the researcher’s particular “view from somewhere.”

Internet research and feminism

In this section, I outline some key definitions of big data and a few of the methodological and ethical challenges big data pose. Next, I outline some of the ways that feminist researchers have sought to address those pitfalls through critical reflection, surveys, and interviews. In the final section, I focus on the recruitment stage of the research process and situate recruitment within wider Internet research trends that call for the development of flexible heuristics.

Big data

As digital tools become more widely available and easy to use, researchers have started using “big data,” which often involves the use of data mining tools to scrape posts and metadata about people’s locations, usernames, timestamps, etc. with publicly available tools and through Application Programming Interfaces (APIs) offered by companies. Examples of these tools include TAGS (Hawksey, 2022), GetOldTweets3 (Giroux, 2020), and Twitter Developer accounts offered to researchers through Twitter. These tools generally scrape only “public” posts, though social media challenge traditional definitions of what data is public and what data is private. While researchers have previously considered the ethical implications of using public posts in qualitative analysis (Ess and Association of Internet Researchers (AoIR) Ethics Working Committee, 2002; franzke, et al., 2020; Markham, 2012; Markham and Buchanan, 2012; McKee and Porter, 2009; Navar-Gill and Stanfill, 2018), big data are often imagined as de-personalized. However, big data are not without ethical complications, including issues around privacy, illusions of objectivity, de-contextualized analysis of data, and extractive research practices that may further harm and silence affected communities while benefitting researchers (boyd and Crawford, 2012; Cooky, et al., 2018; Leurs, 2017; Luka and Millette, 2018; Markham, et al., 2018).

Big data are not defined simply by the size of the dataset though. Humans have stored vast amounts of information in libraries for centuries, and the threshold at which data become “big” changes with technological capacities. Rather than largeness, one of the things that boyd and Crawford (2012) argue is unprecedented about the current “era of big data” [1] is the technological capacities that alter the kinds of meaning people can make from “big data.” Technological capacities and infrastructures fundamentally affect the collection, access, storage, and analysis of data. In terms of collection and access, data mining is an opaque process, as large companies with financial interests in the data have control over what data
The view from somewhere: Critically reflexive recruitment heuristics for big data analysts can access. Overall, with little knowledge of their sources, these data sets are apocryphal, often incomplete, and not necessarily representative (Andrejevic and Gates 2014). Partial data sets are particularly problematic for feminist researchers, as they can inadvertently silence some people and leave out subjugated knowledges (Cooky, et al., 2018). Stated more simply, the access and collection stage of Internet research calls big data sets themselves into question.

Scholars also argue that big data are defined by unprecedented analytic processes carried out by algorithms that bring to light patterns and correlations that human analysis alone may miss (Andrejevic and Gates 2014; boyd and Crawford, 2012). This is especially true when cross-referencing datasets. These analytic processes are largely unknown to analysts, who may simply accept that “data science knows best” (Andrejevic and Gates 2014). The problem with this assumption is that analytic processes cannot be disentangled from the knowledge that they create. Ultimately, the effects of relying on specious sources from companies with financial interests in the data, and the unknown processes to analyze big data are multiple, complex, and often threaten to reinscribe power dynamics. Big data present epistemological questions as they challenge what counts as knowledge, for whom, and with what effects (Andrejevic and Gates, 2014; boyd and Crawford, 2012).

Despite the uncertainty of big data, vast data sets are often accepted as objective sources of knowledge. With a focus on quantitative analysis, big data can suggest a positivistic paradigm (Cooky, et al., 2018; Leurs, 2017; Markham, et al., 2018) in which data are understood to exist independently as “raw” data. Positivist paradigms work on false assumptions of objective, pre-existing “truths” that researchers simply uncover. In effect, positivistic paradigms overlook how researchers interact with data through processes of categorization and interpretation (Bowker, 2006; Gitelman, 2016). Gitelman (2016) argues that data are always “cooked” in the sense that they are the product of the circumstances and classification schemes in which they are produced. Classification systems maintain scientific paradigms that define what questions are legitimate to ask, what counts as knowledge, and therefore, what kinds of knowledge is produced (Foucault, 1973; Kuhn, 2012). The types of knowledge that are produced under these paradigms and classification schemes often reinscribe the dominant status quo. Specifically, the insistence on normative classification systems can perpetuate racist, gendered, and colonial domination by imposing dominant categories on people and groups that construct those groups as an imagined Other (Smith, 2012). The effect is to overlook cultural context and to deny people and cultures the fundamental right to articulate their own cultural values and experiences. The stakes for these people and groups are high; dominant classification schemes can be used to justify criteria for ranking cultures, which in turn inform policy decisions that maintain power dynamics and the inequitable distribution of material resources (Smith, 2012). Since scholars collect “cooked” data and actively classify and interpret big data, the knowledge created from those big data sets are not neutral.

**Feminist approaches**

The power dynamics embedded in big data are particularly important to feminist researchers. Specifically, Leurs (2017), who situates her research in a feminist stance, argues big data research scraped from social media threaten to reinscribe power dynamics around gender, race, and nationality, by naturalizing a “view from nowhere” [3]. This “view from nowhere” gestures to the idealized equitable public sphere, which maintains power dynamics. Put another way, the “view from nowhere” assumes Western, white, male, heteronormative positions as the default. The effect is to obscure power dynamics and leave them uncontested, thereby reinscribing the positivistic paradigms and classification schemes that falsely present big data as neutral. Echoing Leurs’ arguments, D’Ignazio and Klein (2020) argue that both power and binary classification schemes are areas of concern for feminist scholars who work with data. Essentially, the assumption of a “view from nowhere” in big data naturalizes power dynamics.

To account for the power dynamics embedded in big data, some scholars have argued a feminist approach to big data research must account for the lived-experiences of participants that high-level quantitative analysis of big data overlooks. Consequently, big data research must thoughtfully contextualize the differences in data (Cooky, et al., 2018; D’Ignazio and Klein, 2020; Leurs, 2017; Luka and Millette, 2018).
Echoing existing feminist scholarship that centers the voices of participants to disrupt power dynamics (Reinharz, 1992), feminist big data researchers seek to include the voices of participants to add context. Moreover, as a part of engaging with participants, feminist Internet researchers argue big data require ongoing critical self-reflexivity (Cooky, et al., 2018; D'Ignazio and Klein, 2020; Leurs, 2017; Luka and Millette, 2018; Markham, et al., 2018). Broadly, reflexivity means continually considering how research always implicates a “view from somewhere.” Essentially, to account for how data are always “cooked” (Gitelman 2016), researchers reflect on how their lived-experiences affect how they collect, interpret, and present data; they must therefore continually interrogate taken-for-granted assumptions. Reflexivity must account for more than the researcher’s own positions though. Leurs (2017) argues that feminist researchers need to attend to power embedded in “dependencies and relationalities” [4]. Attending to these power dynamics is especially important for intersectional feminist research, which broadly refers to how different parts of a person’s identity will influence other parts in complex and nuanced ways in different contexts and situations (Crenshaw, 1991).

**Recruitment and initial contact: The view from somewhere**

To reflexively account for their limited standpoints and to represent participants’ voices safely and accurately, some big data scholars have reached out to participants to include them in their research through surveys and interviews (Cooky, et al., 2018; Leurs, 2017; Luka and Millette, 2018). Picking up these arguments, I focus on recruitment and articulate some critically reflexive recruitment practices. Critically reflexive recruitment is not necessarily the “next” step in the research process — it is a reciprocal and ongoing part of Internet research.

Recruitment initiates contact with participants and sets the stage for the relationship, though it is not simply building “rapport” with participants. “Rapport” can simply validate the researcher’s sense of themselves as a feminist, becoming an exploitative relationship between researchers and participants (Reinharz, 1992). Therefore, feminist researchers strive for reciprocal relationships with participants that center on “relations of respect, shared information, openness, and clarity of communication” [5]. While reaching out to participants may resist “the view from nowhere,” reaching out also tells those participants that there was a “view from somewhere.” When researchers reach out to people as potential participants, the unstated implication is, “I have seen you, and I found ways to contact you.” Being “found” may feel exposing and frightening for some people, because, despite illusions to the contrary, the public sphere is far from equitable; people with less power are more vulnerable in the public sphere (Fraser, 1990; Warner, 2002). With big data sets where it is infeasible for researchers to reach out to every participant, recruitment is always an act of selection by the researcher that tells people that they have been “found,” and therefore calls for critical reflexivity about the researchers relative position of power in relation to each person to whom they reach out, and how those people may experience being seen differently based on their shifting relative positions of power and vulnerability. This is not to say that people are fundamentally disempowered in the digital public sphere. Scholars have shown digital activism can be effective (Jackson, et al., 2020; Vie, 2014). It is also not to say groups are homogenous, nor is it to say that people in activist groups are always “outside” dominant public spheres, as people always occupy multiple positions. The point here is that in fundamentally inequitable public spheres, people may experience being seen differently based on their relative positions of power and vulnerability. Abidin’s (2020) insights are useful here. Though researchers often assume maximum visibility is always ideal, Abidin challenges this assumption, arguing visibility is labor implicating relational care. Following Abidin, I argue that when reaching out to potential participants in big data research, researchers should critically consider how visibility and being “found” can be experienced as a type of labor.

Picking up the Association of Internet Researchers’ (AoIR) Internet Research Ethics (IRE) 3.0’s emphasis on ethical approaches to big data and different stages of the research process (franzke, et al., 2020), in the remainder of this article, I follow previous scholarship on Internet research and develop heuristics to respond to emerging issues (Ess and Association of Internet Researchers (AoIR) Ethics Working Committee, 2002; franzke, et al., 2020; Markham and Buchanan, 2012; McKee and Porter, 2009). Specifically, I build from Markham, et al.’s (2018) call for researchers to develop less abstract heuristics for
big data. These heuristics focus on the recruitment stage of the research process, though researchers should alter them for particular contexts and projects.

---

**Case study**

Briefly, my dissertation research focused on the transformation of a particular hashtag, #NotAgainSU, as different people re-produced the hashtag across platforms and contexts. The hashtag originated from a Black student-led organization, NotAgainSU. The organization formed in response to a white supremacist harassment campaign at Syracuse University that began in November 2019. I used a mixed-methods approach that combined quantitative data, qualitative content analysis, and survey and interview data. Specifically, for my social media data, this entailed using digital tools to collect data on the public social media posts using #NotAgainSU from 6 November 2019 to 28 June 2020 on Twitter, YouTube, and Instagram. For Twitter, I used GetOldTweets3 (Giroux, 2020). For YouTube, I used free tools available from the Digital Methods Initiative (Digital Methods Initiative, 2020). Finally, I hand-collected public Instagram posts. In total, I collected 6,132 public Tweets, 1,326 public Instagram posts, and 499 YouTube videos. I created data visualizations following Salter, et al. (2020) to explore things like the number or retweets, word counts, co-occurring hashtags, and circulation networks. I included the collection of social media posts in my research protocol, I used ethical fabrication to present composites of social media posts with under 1K followers (Markham, 2012), and I corresponded with an organizer at NotAgainSU about using one of their posts.

Though my project was not designed as participatory research, as a person who benefits from systemic white supremacy, it was important for me to reflect on the limits of my own lived experiences. So, I wanted to included participants in the research, and I reached out to the organization, NotAgainSU, through their official e-mail with an IRB-approved survey. I offered NotAgainSU a deliverable as defined by the group if they were willing to take and circulate the survey. The organizer with whom I spoke did not indicate they wanted a deliverable at that time.

I also reached out to people who created supportive and oppositional content using #NotAgainSU, though I also had to consider the ethics of amplification (Phillips, 2018), or the possibility of amplifying and giving legitimacy to people who espouse hateful rhetoric, in this process. Following Florini (2019), I opted for e-mail recruitment when possible [6], and direct messages on Twitter and Instagram where e-mail was not possible. The survey asked people what kind of #NotAgainSU content participants created, their decisions to create content, and their perceptions of platform environments. It ended with the potential for a follow-up interview.

I conducted IRB-approved reflexive e-mail interviews (Linabary and Hamel, 2017) and teleconferencing interviews (Deakin and Wakefield, 2014). Using both types of interviews gave participants more control over the interview process. I also interviewed an organizer at NotAgainSU, who posted the survey link through their Instagram account. This snowball sampling recruitment was not meant to be representative; rather, it constituted purposeful sampling and was the only way to reach potential participants [7].

Twenty-one people completed the survey. I interviewed three participants. One person was an organizer. Two other participants supported NotAgainSU. Following Florini (2019), I sent interview participants drafts of chapters with their quotes highlighted to lessen participants’ labor when reviewing their quotes. I compensated interview participants with US$20 funded by the College of Graduate Studies from my U.S.-based institution.
Discussion: Recruitment heuristics

To recruit or not to recruit?

One of the first things researchers may want to think about is if to recruit certain participants. These decisions are tricky because publicity puts some people at more risk than others, and because privacy shifts over time. Researchers may tactically choose to not recruit some participants to respect privacy boundaries some participants set up for themselves, or because they understand that some online performances may not be for them. Simultaneously, these decisions may conflict with other feminist values of including participant voices in the research, and these decisions should not rest on patriarchal formulations of care.

As Baym and boyd (2012) argue, publicity is not static, and I add that researchers may need to adapt their recruitment practices to respect shifting publicness. For instance, while the posts in my dataset were “public” when I collected them, some of the accounts and posts had since become private when I began recruitment. I suspect some people made this choice because they were harassed, again pointing to how people experience being in public differently. I recognized that after being harassed online and offline, people may have set their accounts to private to protect themselves from further harassment. Consequently, their current positions of power and the timing were important for me to consider; publicity involves different levels of risk and exposure for different people in different relative positions of power at different times.

If the people posting had been harassed, having a researcher reach out to them to ask them about their experience may have felt threatening. As established above, it sends the message someone is not only still looking, but that the person who has been harassed can still be found by others. Reaching out to potential survey and interview participants and establishing a “view from somewhere” could be a threatening experience. Therefore, these peoples choices to switch their accounts to “private” gesture to the important relationalities and power dynamics implicated in recruitment practices.

While people may have made their accounts private to protect themselves from overt harassment, they may have also simply wanted privacy or been exhausted. While different than directed harassment, people may experience exhaustion from visibility, even when the people and researchers who are watching are supportive. Briefly, I am situating this watching in scholarship about “the gaze.” Broadly, gazing can be an act of power that limits subjectivity. Feminists often discuss it in relation to the male gaze. Mulvey (2012) develops the concept through the metaphor of a camera lens to argue that people who are gazing in on others are positioned as active subjects. In contrast, their gaze can position people socialized as women as passive objects to be possessed and controlled by those gazing in (Mulvey, 2012). However, scholarship on the gaze often overlooks important intersectional power relations; it is integral to recognize that race also intersects with the male gaze. For instance, hooks (1999) points out that gazing, or the denial of the right to gaze, can function within racialized power dynamics too. Of course, gazing can also be an act of recognition and power; hooks (1999) goes on to define the oppositional gaze, which is employed by those with less power as a way to interrogate dominant racialized gazes and reclaim agency; her point is that the position of the person looking is important. My point builds from this work; gazing, and who is doing the looking, is always situated within inequitable power dynamics. Therefore, when I was looking at posts that had since been made private to identify potential participants, I had to consider my own power and position in this looking.

For my purposes, the position of the researcher is important, especially when researchers are interested in counterpublics, which are broadly defined as publics with less power who are in tension with dominant publics (Warner, 2002). Though counterpublics are not fundamentally disempowered, when researchers with relative positions of power gaze in on counterpublic groups with less power, the relationship can become inequitable, even if researchers are supportive of counterpublic causes. Through Collected Authors (2014), Cooky, et al. (2018) point out that when relatively privileged researchers look in on counterpublic groups, the relationship can become exploitative. Similarly, other scholars argue that researchers may not be the intended audience for social media data (Leurs, 2017), and social media participants may not even
know that researchers can scrape their data for published research (Fiesler and Proferes, 2018). Likewise, I understood that even when these posts were public, they may not have been for my gaze. Since I recognized that the tweets and posts may not have been for me to look at in the first place, reaching out to those people after they had chosen to make their accounts private felt voyeuristic, even when they had publicly available e-mail addresses.

If the posts were not meant for me, then it was integral for me to reflect on the power dynamics embedded in my relationships with these potential participants, and how reaching out to them after they had made their posts private may have affected them. Specifically, I am returning to Abidin’s (2020) argument that visibility, or being gazed at, on social media is a kind of labor. Though Abidin’s research focuses on influencers, the arguments are still salient here. As labor, visibility can be exhausting. Ultimately, I recognized that even if my research was supportive, the act of reaching out to potential participants, especially those who had made their accounts private, could still be experienced as exhausting and exploitative if I was not their intended audience. I understood their choice to make their accounts private to say, “I am done doing the labor of visibility right now.”

Nevertheless, discerning whether these posts were meant for me was an ongoing and uncertain process of continual reflection in ambiguous contexts. If privacy online is not just about strict binaries between public and private spheres, but also about the appropriate flow of information (Nissenbaum, 2010), it was unclear for whose gaze these posts were meant. This was especially true on Instagram, where people can cross-post to other platforms, infinitely complicating issues of privacy and power relations (Highfield and Leaver, 2015). If feminist researchers reflect on their relative positions of power, as other people’s intended audiences become ambiguous, so too do relationalities and how potential participants may experience being viewed from the researcher’s particular “view from somewhere.”

These ambiguous contexts were especially difficult for my feminist approach because I also recognized that some people who set their accounts to private may have still wanted to be consulted. Thus, wanting to respect these people’s decisions to make their accounts private and to not further drain their energies came into conflict with my commitment to reflect on the limits of my own position and to center the voices of those affected. Not reaching out to these people meant they would be excluded from my research. However, since groups are not homogenous, I could not know how people would experience my e-mails. On the other hand, I considered how not reaching out to these participants could enact a patriarchal notion of care, in which researchers appoint themselves to make decisions for others. People in similar intersectional positions may rightly have different experiences of being seen; while I could not know how participants might feel about having a researcher reach out to them about posts that they had made private, I also did not want to decide for them how they would be represented simply because they made their posts private.

Ultimately, I decided to respect the potential participants’ decisions to have private accounts in the present through practicing attunement. Amrute (2019) argues feminist social media researchers need to be affectively attuned to their digital environments and the labors of those environments, which, I argue include the labors of visibility. Like Amrute (2019), Dadas (2016) takes a queer approach and argues researchers need to be “attuned” to how privacy shifts. For Amrute, attunements can be understood as a practice that proceeds from an awareness of all the factors that go into a situation, including affect. By affect, I am broadly drawing from Ahmed’s (2014) understanding of affect as a kind connective force and structure of feeling that intimately shapes and impacts bodies. Among other scholars, Amrute (2019) argues that affect is intertwined with technology. Following suit, I recognized the affective sensitivity of the posts, and the emotional and bodily impacts that those posts could have on potential participants. However, to avoid patriarchal care and making decisions for participants about how they were represented, since I did not include these people’s voices in my research, I decided to avoid direct references to these people in the dissertation. When necessary, I included references to general events, but I maintained my decision not to include direct references to these people even when publicly available news media did report on white supremacist harassment against them. Much like AoIR argues researchers will need to think about what to do with data that has become private, researchers will also want to think about if they will recruit participants whose posts become private.
In another case, a participant used #NotAgainSU in a way that Phillips and Milner (2017) might characterize as “odd.” They used the hashtag to interact with high-profile right-wing actors, and they echoed white supremacist talking points in an over-the-top and affected manner. The account identified itself as a parody account. Because of my immersion in studying the case, I understood this account’s tweets to be ironic, performative, and disidentificatory. Broadly, Muñoz (1999) argues that rather than identifying within oppressive power regimes or counteridentifying against oppressive power regimes, disidentification refers to a survival strategy for people with less power to re-appropriate toxic stereotypes. Disidentification becomes a site of self-creation. Yet, as I argued through above, these performances may not have been meant for my gaze.

I was also concerned about to whom this account might circulate the survey. They had previously interacted with accounts that had histories of doxing people, which refers to publishing private or identifying information about people, such as home addresses. I did not want them to performatively circulate the survey into what Massanari (2018) calls the “alt-right gaze.” Broadly, Massanari defines the alt-right gaze as the means by which “white, ethnonationalist, fascist, misogynistic, and anti-intellectual communities” that surveille, intimidate, and dox those with less power online. While Massanari’s arguments are about the danger to researchers, these arguments extend to other people endangered by the alt-right gaze. Since my reflexive process involved critical considerations of relationalities, if this person had chosen to circulate the survey to people in the alt-right, this could have caused dangerous backlash for both myself and the organization. Accordingly, while this account remained “public,” I did not reach out to this participant either. Again, this decision clashed with my commitment to center the voices of people affected and to not make decisions for participants. As with the participants who made their accounts private, I avoided references to this participant in my dissertation.

In both of these cases, despite my commitment to center the voices of people affected, different relative positions of power in the public sphere meant I chose to not reach out to some potential participants because I recognized their posts may not have been meant for my particular “view from somewhere.” My point here is not that researchers should abandon reaching out to participants altogether, but reaching out may be an ongoing and messy process of attunement and reorientation. Reaching out always concerns power because it relies on data the researcher could “publicly” collect about participants with different levels of vulnerability. First, I recognized that publicity and privacy shifts over time, and people may choose to make their posts private at a later date. Reaching out to them could demand the labor associated with visibility from them and violate their expectations of privacy in the present. Second, I recognized that though some posts and accounts remained “public,” their (seemingly) disidentificatory performances may not have been meant for me, and I could not predict where they might circulate the survey. So, I chose to omit direct references to these people despite media coverage.

Researchers could ask:

- If an account has since been set to private after I have collected my big data, will I still reach out to that account to include their voice in my research, and how will I avoid making decisions for them?
- Could reaching out to this particular participant at this particular time feel frightening or exhausting by demonstrating to the participant others can still find and contact them? How will I tell? Again, how will I avoid making decisions for participants?
- If I choose to not reach out to particular participants, should I omit direct references to them in any publications about the events, since I did not seek their input?
- Even if posts remain public, is this person enacting a performance that may not be meant for me? How will I tell?
- Based on past online interactions, might a participant circulate a survey to people who are dangerous to the researchers or communities?

The view from “Research Accounts”

Second, in critically reflexive recruitment, researchers will want to think about how they inform potential
participants that they are researchers. This is especially important since feminist researchers strive to engage with participants equitably and on transparent terms that extends beyond simply building rapport with participants.

Though I recognized transparency would need to shift for different contexts (Dadas, 2016), I followed Porter and McKee’s (2009) suggestion to cultivate a culturally appropriate ethos [10] and for transparency, I labelled accounts dedicated to research as “research accounts.” My intention behind labelling these accounts was to have my initial contact with potential participants be on public, and equitable terms about my own position as a researcher. Researchers may choose to create research accounts for clarity and transparency, but they may also wish to add a layer between themselves and potentially hostile people. I e-mailed potential participants and used the private messaging systems of Twitter and Instagram to reach participants through these “research accounts.”

Overall, I received few responses to the survey, effectively omitting some voices and context from the big data. In terms of e-mails, I suspect two things may have made potential participants less likely to see or respond to e-mails. As third-party companies collect people’s e-mails for advertising purposes, people must sort through advertising e-mails. This situation was exacerbated during the COVID-19 pandemic, in which many people expressed feeling e-mail fatigue. When overwhelmed with e-mails, many people skim through the preview of the message and look at the sender to decide whether to send the e-mail directly to the trash folder. These practices are necessary to allow people to set their own public/private boundaries, although it is also important to acknowledge that positioning individual people as solely responsible for protecting themselves in fundamentally inequitable systems obscures power structures. Pragmatically though, for researchers seeking to exercise transparency, names like “Research Account” could easily blend into the deluge of other e-mails that people receive, and they may have either missed or deleted it.

More importantly, since people experience publicity differently, “Research Account” would also impact different audiences differently based on their relative positions of power. While “Research Account” is clear, some potential participants may be rightfully wary. For some communities, “research” often functions in an extractive paradigm that centers on the findings of the research, rather than on the effects on communities. This is especially true for participants who are a part of communities affected by oppression. Many of these communities are already well-aware of the way research is often “about” them, rather than for them (Smith, 2012). Again, not all online content is for researchers.

Potential participants may have felt like the e-mail had infiltrated their inbox to demand more emotional work, labor, and time from them; in effect, the name of the account could set up an inequitable and extractive relationship between a researcher and potential participants, especially if the media was not meant for researchers. Moreover, though no group is hegemonic and general “research” e-mails may not feel extractive (even if they are), when that research is about power dynamics implicated in systemic white supremacy combined with the threat of the “alt-right gaze” (Massanari, 2018), group affiliation matters. When e-mails showed up in potential participants’ inboxes, they would know a “Research Account” had found their e-mail. Even if the researcher is supportive, “Research Account” still alerts participants to the fact that a researcher found their contact information, and receiving an e-mail from “Research Account” could feel not only voyeuristic but also threatening in this context.

When I reached out to potential participants on social media, I had similar concerns. Though the context of direct messages on social media are different than e-mails, as I discuss below, it is not uncommon for people to receive messages or message requests from accounts seeking to sell them different products and services. This is especially true on Instagram, as it is deeply embedded with an influencer economy (Leaver, et al., 2020). In this context, “Research Account” could suggest I was doing market research. The name could also suggest extraction, and it therefore raises the same potential for inequitable initial contact with potential participants.

Moreover, researchers should also critically reflect on the means by which they do establish a view from “Research Accounts.” Although I circulated the survey through these accounts, in my recruitment e-mail, I
The view from somewhere: Critically reflexive recruitment heuristics for big data

included standard language for snowball sampling asking participants to further circulate the survey. The organizer offered to post the survey link to their Story [11] and as a link in their Instagram bio. I suspect their posts helped my recruitment efforts. Though the survey had some responses before they posted the survey and some people to whom I reached out on social media responded to me directly and took the survey, some responses appeared directly after the organizer posted the survey to Instagram.

This unsurprisingly suggests the recruiting account and their position matters. Through their efforts, NotAgainSU’s account had developed an important ethos with their audiences, and they had built and maintained networks. Importantly, Fisher (2015) argues that building and maintaining networks is a kind of labor from which marketers profit. Though I was not a marketer, and I asked NotAgainSU for their help and they had the autonomy to decline, since NotAgainSU had taken the time to build those relationships in the first place, it could be argued that some of the survey responses that benefitted me came from their unpaid labor to build networks with their audiences.

Broadly then, future researchers who use e-mail and social media as a recruitment method may want to critically reflect on the names that they select for their research accounts. Researchers may reflect on how extractive research has impacted some communities, and about how having a person’s e-mail “found” by a “Research Account” may be experienced as voyeuristic, threatening, or exhausting; “Research Account” can be understood as demanding more time and labor from participants. While researchers may ask accounts with more credibility to help recruit other potential participants, researchers might also consider the labor on which those networks were built.

Researchers could ask:

- How can I transparently and equitably announce my presence as a researcher to potential participants?
- What relationship do the people to whom I am reaching out have with research? Though no group is homogenous or experiences an event in the same way, could they understand or experience this research as extractive? If I choose to create accounts, what is a culturally appropriate name for the research account that is both transparent, but also not potentially draining and threatening to potential participants given their shifting positions of power?
- If I am using snowball sampling, which accounts will hold the most ethos with potential participants, and whose labor built the networks through which a survey would circulate?

Channels of recruitment

Finally, since privacy also implicates the appropriate flow of information (Nissenbaum, 2010), researchers will want to think about which channels are the most appropriate for initial contact. However, they will also want to think about how their selection of those channels can re-inscribe power dynamics through adherence to strict boundaries of publicity and privacy.

Obviously, researchers will want to think about how e-mail and social media private messaging systems are different rhetorical contexts. This broadly reiterates the IRE 1.0’s (Ess and Association of Internet Researchers (AoIR) Ethics Working Committee, 2002) emphasis on considering people’s expectations for privacy, thinking about the most appropriate medium through which to ask for informed consent and to recruit participants, and thinking about the expectations of potential participants [12]. For my purposes here, I argue that researchers will need to make thoughtful decisions about which channels are the most appropriate for particular participants.

Selecting channels is important because it implicates the public/private binary, as some channels are socially and culturally perceived as more or less public. If a researcher reaches out on a channel the participant perceives to be more private, the recruitment message/e-mail could again function in a milieu of a normative gaze, though it is difficult to ascertain particular participants’ beliefs and expectations as publicity and privacy shift (Baym and boyd, 2012). In addition, as outlined above, the risks of publicity are different for different groups of people in particular contexts, though people who are in similar relative
positions of power may understand those same channels differently.

In general, e-mail is traditionally considered more “professional” than social media. When reaching out to people to potentially use their online content in her research, Florini (2019) writes about how she opted for publicly available e-mail addresses over social media direct messages. Some potential participants who were central to my networks had publicly available e-mail addresses, both through their employers and posted to their Twitter bios. So, I followed Florini, and I opted for e-mail when possible. I did not reach out to people on both social media and e-mail; reaching out to them on multiple channels could be draining for participants; it is demanding on the part of the researcher.

I chose e-mail for two reasons. First, as people often use social media to connect with people in their social lives, social media accounts may be perceived by participants as more intimate spaces. Indeed, Leaver, et al. (2020) argued Instagram positions itself as cultivating a “networked intimacy,” though people may experience the platform differently. Twitter colloquially occupies a more public position. Although many journalists use Twitter as a part of their professional lives (Phillips, 2018), Twitter can also be used as a space of communication between friends. In contrast, people commonly use e-mail for their work and professional lives, so e-mail culturally serves a relatively more public function.

The fact that social media socially and culturally may function as relatively more intimate spaces is important because Hurtado (1989) argues that the public/private distinction was only ever relevant for Western, middle-class women who are privileged by systems of white supremacy. The state has historically intervened in the “private” spaces of groups affected by systems of oppression. As Hurtado succinctly writes, “There is no such thing as a private sphere for people of color except that which they manage to create and protect in an otherwise hostile environment.” While communications between people on social media occupy an ambiguous public space, when they are communications between in-group members, they function with more relative privacy, speaking to the colloquial understandings of social media as a more private space. These networks and communications between users, though accessible to me, may have been relatively more private spaces that potential participants “managed to create and protect in an otherwise hostile environment,” in Hurtado’s words. Consequently, I wanted to respect the relatively private communications and support networks that people affected by the white supremacist harassment campaign had set up for themselves.

The second reason that I selected e-mail was because I wanted to respect the authority of potential participants for my initial contact to be as equitable as possible. Potential participants included public journalists and people who were well-respected in their fields. Feminist researchers do not position themselves as authorities with participants, and instead they see participants as co-creators of knowledge with specialized expertise in their own communities and practices (Reinharz, 1992). I understood reaching out to official e-mail accounts associated with people’s positions of authority as well-regarded academics or journalists as a way for me to respect this expertise and authority.

At the same time, opting for e-mail in every instance does not account for how some people use social media as a public space and how they may expect for others to reach out to them there. Most obviously, though I relied on e-mail as a public/respectful method of communication, some people use social media as a public, professional, and activist space. If potential participants used social media as a public space on which they expected people to reach out to them, it is possible my choice to opt for e-mail excluded them. E-mails may have seemed too formal and stuffy since they were active and public on social media.

More directly related to my project, though social media is often colloquially viewed as frivolous, activists frequently use social media to get around traditional authoritative gatekeeping mechanisms to control their own narratives (Jackson, et al., 2020). Activists use social media as a public space through which to garner coalitional support. If activists use social media as a public space, then they may expect for people to reach out to them there. As privacy is about the appropriate flow of information (Nissenbaum, 2010), then the appropriate flow of information may be on social media in this case.
The view from somewhere: Critically reflexive recruitment heuristics for big data

The fact that some people use social media as a public space is important because feminist and queer researchers have raised legitimate questions about traditional authoritative structures. They argue those authoritative regimes maintain their power through strict divisions between public and private spaces (Fraser, 1990). Briefly, the strict division overlooks the lived experiences of some people in favor of those naturalized as the default, which Leurs (2017) and D’Ignazio and Klein (2020) identified as issues of particular concern for feminist data researchers. Ultimately, issues defined as “private” matters are not considered “common” concerns, thus relegating those matters outside the realm of politics and the “public” debate, ultimately maintaining the status quo (Fraser, 1990). Thus, for feminist researchers to include participants in their research and to interact with potential participants on equitable terms, not meeting potential participants where they are (social media) due to traditional notions about e-mail being more “public” or “respectful,” could reinscribe this problematic binary that relegates some issues to the “private” sphere, and thus, outside the realm of politics. The decision presumes social media is a private space and that traditional e-mails are more “professional” and more public.

Thus, I ran into another tension between my commitment to center participant voices while also respecting their privacy. In choosing to opt for e-mails over social media, I could have unintentionally excluded potential participants by relying on colloquial and hierarchical boundaries between publicity and privacy. Reaching out solely via e-mail could be understood as an act that re-invests in existing hierarchies through reinforcing existing boundaries between public and private spaces. By the same token, respecting these distinctions and reaching out via “professional” channels rather than relatively more “private” social media channels could also be understood as a way to honor the private networks that systemically oppressed groups have set up for themselves. This is especially important since these groups are often denied private spaces through state interventions. Furthermore, I wanted to respect participant’s positions of authority by using e-mail. The trouble for researchers is that it is difficult to ascertain how individual people use social media based solely on (semi) public interactions. These tensions broadly reiterate Nissenbaum’s (2010) point that privacy is about the appropriate flow of information on the appropriate channels. Likewise, researchers will need to critically reflect on which channels of contact are the most appropriate for particular participants, given their relative positions of power at a particular time on a given issue.

Researchers may want to think about how to adapt to individual participants and how those participants seem to view their own social media interactions. If participants openly use social media as a public and professional space, researchers could think about opting for social media direct messaging instead. However, researchers should think critically about what markers they will use to clarify if people use social media as a public or private space, and how reaching out on social media and other channels colloquially considered more private could also reinscribe power dynamics.

Researchers could ask:

- Does the participant have e-mail or another preferred contact method in their social media bio? Based on their particular position, my own position, and the sensitivity of the topic, how might a participant experience having a researcher reach out to them on social media? Via e-mail?
- How does this particular participant seem to interact on social media; do they seem to use it as a public or a private space? What makes me believe they use it in one way or another? Are those assumptions based on my own position and background?
- Am I potentially excluding participants by relying on my own traditional assumptions about the public/private divide? Am I potentially violating the tenuous private communications that people affected by systemic oppression manage to set up for themselves online?

Conclusions

This manuscript adds to existing scholarship about ethical research practices with big data collected from
social media by offering heuristics that can help researchers’ reach out to include participants in their research. I focus on feminist reflexivity in the recruitment stage of research as it implicates important issues around publicity and privacy.

While being seen can be experienced as empowering, queer and feminist scholars, as well as scholars of race and ethnicity, argue that not all information is meant for all people to see. This is important because the public sphere is not experienced by everyone in the same way. Some people may experience being publicly seen as threatening. Reaching out to potential participants to inform them of the researcher’s “view from somewhere” can function in the milieu of the dominant gaze.

When reaching out to potential participants to include them in their big data research, scholars should reflectively consider how being seen will impact different participants based on their relative and shifting positions of power, but they should also consider how strict adherence to traditional public/private binaries can exclude participants. I contend researchers should critically consider if to include certain participants based on how those participants’ public/private boundaries have shifted over time. Researchers should also consider if public online performances were meant for them, and if some performances may further endanger communities. Of course, researchers should still reach out to participants as to not exclude participants from the research. Researchers might also critically consider the name of the accounts through which they initiate contact with participants and how those names can feel extractive. They should also think about whose work and labor built the networks through which they recruit participants. Finally, researchers can critically consider the channels by which they reach out to participants, especially as it relates to upholding strict divisions between publicity and privacy.

While these heuristics are meant to take up Markham, et al.’s (2018) call to develop less abstract heuristics, other researchers can add to these heuristics. For instance, these heuristics are based around researching with communities affected by systems of oppression; researchers would need to consider a different set of questions when recruiting from groups with more power. Ultimately, while it is useful, and often necessary, for researchers who take a feminist approach to big data to reach out to participants for important context, recruitment is also not a straightforward process; rather it is a reciprocal and ongoing part of the process of Internet research that involves critically reflexive recruitment practices that continually interrogate the researcher’s particular “view from somewhere.”

About the author

Corinne Jones (Ph.D.) is a visiting lecturer at the University of Central Florida whose research interests include digital rhetoric and circulation, social media, social justice, digital humanities, and digital research. E-mail: leah [dot] jones [at] ucf [dot] edu

Acknowledgments

This research was supported by the College of Graduate studies at the University of Central Florida. Though my decisions were my own, I would like to thank my dissertation advisors, Dr. Mel Stanfill and Dr. Dustin Edwards, as their insights and suggested readings were invaluable during the research process. I would also like to thank Lauren Rouse for reading over an early draft of this article, and I would like to sincerely thank the insightful and helpful reviewers of my original and revised submission.

Notes

The view from somewhere: Critically reflexive recruitment heuristics for big data


8. Dadas, 2016, p. 64.


11. Instagram Stories are images or short videos that people are disappear after 24 hours or after a “friend” has viewed the story. Instagram Stories are colloquially understood to be more casual than official posts.


References

doi: https://doi.org/10.33621/jdsr.v2i1.20, accessed 3 August 2021.


Annemarie Navar-Gill and Mel Stanfill, 2018. “‘We shouldn’t have to trend to make you listen’: Queer fan hashtag campaigns as production interventions,” Journal of Film and Video, volume 70, numbers 3–4, pp. 85–100.
doi: https://doi.org/10.5406/jfilmvideo.70.3-4.0085, accessed 1 August 2020.


---

Editorial history

Received 2 September 2021; revised 14 October 2021; revised 17 January 2022; accepted 31 January 2022.