Digital Traces of Sexualities: Understanding the Salience of Sexual Identity through Disclosure on Social Media

Connor Gilroy1 and Ridhi Kashyap2

Abstract
The authors analyze the expression of sexualities in the contemporary United States using data about disclosure on social media. Through the Facebook advertising platform, the authors collect aggregate counts encompassing 200 million Facebook users, 28 percent of whom disclose sexuality-related information. Stratifying by age, gender, and relationship status, the authors show how these attributes structure the propensity to disclose different sexual identities. There is a large generational difference; younger social media users share their sexualities at high rates, while for older cohorts marital status substitutes for sexual identity. Consistent with gendered expectations, women more often express a bisexual interest in men and women; men are more explicit about their heterosexuality. The authors interpret these variations in sexuality disclosure on social media to reflect the salience of sexual identity, intersected at times with availability. This study contributes to the sociology of sexuality with a quantitative analysis, using novel digital data, of how sexuality is signaled socially.

Keywords
digital trace data, disclosure, sexual identity, sexual orientation, social media

Digital social media hold the potential to expand quantitative knowledge about the disclosure of sexuality. Systematically understanding how people who are lesbian, gay, bisexual, or queer (LGBQ) express their orientations and identities poses a number of challenges, from the sheer smallness of the populations in question to the long history of stigma facing these groups. The benefits of new data about sexuality are not confined to sexual minorities, however. This same lens of identity and expression can also apply toward understanding heterosexuality. Indeed, heterosexual identity may take on heightened salience at a time when LGBQ people are more and more visible (Dean 2014). Here, we consider in tandem how heterosexual and LGBQ people express their sexualities online.

Sexuality is signaled socially. A sexual identity is not inherently visible or self-evident and can be shared or withheld in different social settings. Whether a sexual identity is shared depends on its salience to a given social situation or to a given individual’s self-identity (Doan and Mize 2020). Working against social disclosure, sexuality has long been treated as something stigmatized, shameful, secret, or simply private; for sexual minorities, concealing sexual identity may be a matter of safety (Goffman 1963). Yet sexual identity is also a core social and demographic trait. In a place like the contemporary United States, sexuality is a central part of people’s identities that affects individual worldviews, life experiences, and trajectories (Schnabel 2018). Although it is less often studied or measured than other socially salient attributes, such as gender or race, survey data have shown marked generational shifts in the underlying distributions of different sexual identities, with young people increasingly identifying as sexual minorities (England, Mishel, and Caudillo 2016; Gates 2014, 2017; Jones 2021). These generational shifts have occurred against a backdrop of a growing digitalization of life and the life course, where digital social media are an active and salient space of social interaction and expression. In turn, the digitalization of our lives has also generated new data opportunities that offer a lens through which to understand this shifting terrain.

1University of Washington, Seattle, WA, USA
2University of Oxford, Oxford, UK

Corresponding Author:
Connor Gilroy, University of Washington, Department of Sociology, Box 353340, Seattle, WA 98195, USA
Email: cgilroy@uw.edu
When people use social media platforms, they create “digital traces” of their activities and identities. These digital traces offer the unique opportunity to study how people manage and disclose information about themselves, including their sexualities, in a relevant social context, one that now forms an integral part of our lives. Our focus here is on the online disclosure of sexual identities as a social process, rather than an unmediated expression of underlying identities or “authentic selves” (Haimson and Hoffmann 2016). We do not consider these digital trace data as a straightforward measure of the demographic prevalence of different sexual identities, although we contextualize social disclosure using such measures from offline survey data sources.

We examine the disclosure of sexuality, and how it varies by other social attributes, using aggregate data from the population of Facebook users available from its advertising platform. Facebook is ideal for this purpose because it is the largest social media platform in the United States; 69 percent of all American adults have Facebook accounts (Auxier and Anderson 2021; Gramlich 2021). To obtain aggregate counts of users disclosing different sexualities, stratified by other characteristics we expect to be salient, we use the targeted advertising capabilities of the Facebook advertising platform. In this way, these data can be conceptualized as a type of “digital census” of the online population of Facebook users (Cesare et al. 2018). Existing work has used Facebook advertising data to model demographic and social indicators linked to migration (Alexander, Polimis, and Zagheni 2019; Zagheni, Weber, and Gummadi 2017), male fertility (Rampazzo et al. 2018), and Internet access gender gaps (Fatehkia, Kashyap, and Weber 2018; Kashyap et al. 2020), and to understand the demographic biases of this online population by validating against “ground truth” measures (Alexander, Polimis, and Zagheni 2020; Ribeiro, Benevenuto, and Zagheni 2020). We contribute to this growing body of literature by using these data for understanding the expression of sexuality. In contrast to more demographic approaches, however, we emphasize the interpretive opportunity offered by these data and consider the sociological implications of the patterns we find.

Our study contributes to the sociology and demography of sexuality by leveraging quantitative digital traces arising from a real social context as a unique lens for observing the social process of disclosure. These data can add to existing descriptive knowledge about sexual minorities, while also demonstrating the varying salience of heterosexual identity. We show at scale and in detail how the disclosure of these sexualities interacts with other social markers such as age, gender, and marital status. Our results corroborate and extend findings from surveys and qualitative interviews and suggest new directions for research on sexualities.

We proceed as follows. We begin by discussing how sexuality is disclosed and expressed in general and how people can express themselves on social media specifically. We then describe our approach to collecting data from Facebook’s advertising platform and to modeling those data in a way that conveys uncertainty. Next we present our results broken down by different characteristics, then reassemble them into a full picture of how sexuality is disclosed on Facebook in the United States. We close by considering both the implications of our findings and the potential futures of this sort of digital trace research.

**Background**

When people share information about their sexualities on social media, what might they mean to express? Social media platforms offer structured and unstructured ways for people to share their sexualities. On Facebook, this includes a profile field labeled “interested in,” for which the options are “men,” “women,” or “men and women.” But if a Facebook user’s profile shows that they are, for instance, “interested in men and women,” it is not clear a priori which aspect of sexual orientation this statement might signal. Sexual orientation is multifaceted, encompassing behavior, attraction, and identity (Laumann et al. 1994). On the basis of prior research about social disclosure of sexuality (Doan and Mize 2020) and the systematic patterns we find, we argue that in this digital context “interested in” is primarily a marker of sexual identity. As we will show, for people in some social positions this profile field operates as an unambiguous signal of identity, and for others identity intersects with related considerations such as sexual or romantic availability.

A fundamental distinction shaping social disclosure is that minority sexual identities are marked categories, while the majority identity of heterosexuality is taken for granted (Brekhus 1996; Zerubavel 2018). This means that people are generally presumed to be straight, and LGBQ people must disclose their nonheterosexuality in social situations by coming out. Because sexual minorities have long faced stigma, this repeated process of disclosing a nonnormative sexuality in new contexts can be difficult and fraught (Goffman 1963; Orne 2011, 2013). Counterbalancing that stigma, factors such as identity commitment positively mediate disclosure for sexual minorities (Doan and Mize 2020). Because of the emotionally charged, sometimes risky nature of outing oneself, and because of the potential strength of their identification with their sexual identities, we expect that LGBQ people would, on the whole, find this process of disclosure to be highly salient. In other words, they are accustomed to actively managing information and impressions about their sexualities.

Even though heterosexuality is unmarked, heterosexuals still also manage impressions about their sexualities. Heterosexuality is a privileged and normative social identity. Critical scholars of heterosexuality have analyzed this normative appeal of straight culture, showing how heterosexuals can lay claim to a heterosexual identity even when other aspects of their sexual orientation, such as behavior, do not completely align (Budnick 2016; Carrillo and Hoffman 2018; Silva 2017, 2018; Ward 2015, 2020). The cultural meanings of heterosexuality also vary by gender, and this shapes the meaning of disclosure for heterosexual men and heterosexual women. For men, avoiding being perceived as
gay forms a crucial part of heterosexual masculinity (Mishel, Bridges, and Caudillo forthcoming; Pascoe 2011; Ward 2015). Although homophobia is no longer central to all expressions of heterosexual masculinities (Dean 2014), men continue to experience pressure from peers to perform compulsory heterosexuality (Duckworth and Trautner 2019). Women, by contrast, are culturally afforded more flexibility in terms of both identity and behavior. Sociologists have argued that this is one reason for the rise in nonheterosexual identities and activities among young women (England et al. 2016; Mishel et al. 2020). What this means is that explicit indications of sexuality have a different valence, depending on whether someone is part of a minority or the majority, and also depending on their age and gender. These varied potential meanings of online disclosure are reflected in the variations in disclosure by social categories (e.g., age, gender) that we will show in our results, which can be interpreted through an intersectional lens (Crenshaw 1989).

People express their sexual identities online because digital social media have become key venues for self-expression and impression management more generally. Where online spaces were once niche and separate social contexts, they increasingly overlap with offline social worlds (boyd 2014; Jurgenson 2011; Orne 2017). This context collapse means that LGBQ people must manage their sexual identities online as they do offline. As Duguay’s (2016) interviews of queer British youth about their use of Facebook reveal, young LGBQ people are generally mindful of Facebook as a potential medium for disclosing sexualities and yet vary in how visibly they signal their sexualities, with about half of these queer young people using the “interested in” profile field to signal their identities. The remainder declined to use it, either out of privacy concerns or because they found it too rigidly binary to express queer identities. There is less prior research about the expression of heterosexual identities online, though Pascoe and Diefendorf (2018) show that men use homophobic language online to express a heterosexual style of masculinity, consistent with their offline behavior.

Our digital trace data provide a quantitative tool for studying the social disclosure and expression of sexualities, with unique strengths. We obtain a census of the complete population of U.S. Facebook users, including those who do not disclose their sexualities. These comprehensive data equip us to examine the social process of identity disclosure in a novel way. Of course, online data are also constrained by their origin as found data (Salganik 2018). Technology companies decide how to structure the data they collect, and researchers must work within that structure when using social media data. For instance, the way Facebook collects gender data and makes those data available constrains our analysis of sexualities. Although Facebook allows individual users to select from a range of custom genders to appear on their profiles, Bivens (2017) showed that in internal databases, user-specified genders are reduced down to three categories: women, men, and nonbinary people. Advertisers, subsequently, may target advertisements only toward women, toward men, or toward people of all genders (Bivens and Haimson 2016). Because advertisements cannot be targeted toward nonbinary people at all, nor toward more specific binary gender identities (e.g., only cisgender men, only transgender men), data cannot be collected from Facebook’s advertising platform for these groups. We therefore cannot analyze sexuality disclosure among these groups in this study. This is one inherent limitation of found data, and we will return to the consequences of it in our conclusion.

In the following analysis, we investigate the disclosure of sexuality on Facebook, with the goal of assessing which characteristics shape the disclosure of sexual identities and how. We collect and model count data about U.S. Facebook users, then aggregate predictions from the model to explore variations by age, gender, and relationship status. For some combinations of characteristics, rates of social disclosure are high across all sexual identities; in other cases identity disclosure is more conditional and varied.

Data

We collect data from Facebook using its advertising platform. Facebook links social, demographic, and behavioral information about individual Facebook users to categories advertisers can use to select accounts that will be targeted with advertising. Potential advertisers on social media platforms such as Facebook can specify a desired audience for their ads on the basis of targeting criteria, such as gender, age, geography, and other characteristics. For some characteristics (e.g., political preferences), these attributes are algorithmically inferred categorizations, and the relation to concrete user-disclosed information is opaque. In other cases, including ours, specific user profile fields are directly linked to targeting options in the Facebook Ads Manager. Before an ad is actually launched, Facebook’s Ads Manager shows aggregate counts of audience sizes of the queried targeting attributes; these are the data we collect. Figure 1 shows both sides of this system, from user and advertiser perspectives. The figure highlights our main variable operationalizing sexuality, the “interested in” profile field and targeting option.

“Interested in” is an optional field Facebook users may fill out or leave empty, stating whether they are interested in men, women, or both men and women. In a user profile, this field is found alongside gender and pronouns, under “basic information.” From an individual user perspective, the field has no explicit definition or description beyond the options presented. For advertisers, the field is listed in the “Demographics” section, under “Relationships” alongside “Relationships Status.” The advertiser-facing descriptions are quite open ended. They read, for instance, “People who are interested in Men and Women for friendship, dating, relationships or networking.” Despite the vagueness of that definition, each of these contexts supports the assertion that “interested in” relates to sexual identity.

We use the Facebook advertising platform to collect aggregate data by intersecting the “interested in” field with other
attributes. The example in Figure 1b shows how all advertisements must be stratified by age, gender, and geography in some fashion, and optionally by other characteristics as well. In the example, the audience for a hypothetical advertisement is men age 20 in the United States who are interested in men. The estimated audience size for this group, the “potential reach” highlighted in the figure, is 25,000 people.

By querying this system, we systematically collect estimates for each possible combination of our variables, to build up a complete multiple-way contingency table. We do not obtain these estimates by manually querying the Ads Manager user interface. Instead, we automate data collection, retrieving estimates programmatically using the Facebook Marketing application programming interface (API). We access the API using the facebookads software development kit for Python, which Facebook officially develops and releases for use by registered third-party developers. At the time of data collection in September 2017, this registration process was relatively open to anyone with a verified Facebook account.

Our primary data set consists of estimates for the number of adult Facebook users in the United States for every combination of sexuality, gender, relationship status, and age. These data represent counts of monthly active users of Facebook. From the perspective of social media users, aggregate data such as these present fewer risks and ethical concerns than individual-level data (Fiesler and Proferes 2018). Nevertheless, a few aspects of these data have implications for our analyses. First, all of the estimates returned from the

Figure 1. The “interested in” profile field from user and advertiser perspectives. Screenshots from 2017 by one of the authors. (a) The user perspective. (b) The advertiser perspective, with estimated count highlighted.
Facebook Marketing API are rounded to two significant figures, regardless of the magnitude of the estimate. At the time of data collection, the minimum value that could be returned was 20 individuals. This minimum helps preserve k-anonymity (Sweeney 2002), preventing reidentification and protecting individual privacy. Combined, the minimum value and rounding to two digits mean that the precise true number of users is more uncertain for larger categories. We minimize the impact of these features by collecting more stratified estimates and adopting a modeling strategy that accounts for some uncertainty.

Second, Facebook does not guarantee that these estimates correspond to external population values, nor that they are useful for anything beyond advertising. However, more than most social media platforms, Facebook attempts to enforce a principle of “authenticity,” whereby each person has only a single user account (Haimson and Hoffmann 2016). Still, human users with multiple accounts, as well as nonhuman or “bot” users, remain potential sources for systematic error in our estimates.

Finally, we note that Facebook’s advertising platform is in considerable flux. Categories or fields that can be accessed, how these categories are defined in the Ads Manager, as well as minimum counts of audience sizes change continually and with little notice. These changes sometimes occur in response to public critiques, especially with regard to political advertisements or discriminatory advertising practices (Goldman and Himel 2018), but in other cases have occurred with minimal explanation. Relevant to this study, in 2018, Facebook removed the “interested in” profile field as a targeting option for advertisements, without any indication of the specific factors or circumstances motivating this change. To a large extent this unpredictability is not unique to Facebook’s advertising platform but applies more broadly to digital trace data sources. These developments shape the opportunities for and limitations of research with these types of data sources, and we will return to the implications of these changes in the conclusion.

Methods

Our aim is to investigate how the disclosure of sexuality, and of specific sexual identities, on Facebook intersects with three other variables: age, gender, and relationship status. To examine these associations, we treat the stratified aggregate estimates we obtained as the outcome of a statistical regression model. All four variables, including the focal variable of sexuality disclosure, are covariates. Because the estimates are count data, we use a log-linear model with the associations between variables expressed as interaction terms (Agresti 2012). Such interaction terms are one quantitative and intersectional approach for examining intercategorical complexity (McCall 2005).

The primary motivation for constructing a statistical model from the data is to investigate which covariates can be viewed as independent from each other, conditional on the other variables. However, our best-fitting model ultimately includes all two-way and three-way interactions. If a simpler model had fit the data equally well, then it would have been possible to conceptually simplify the relationships among sexuality, gender, relationship status, and age. Instead, as we show in the “Results” section, none of these factors can be disentangled from the others.

Two additional considerations motivate us to build and present a statistical model, rather than simply describe the underlying data. First, a model regularizes the data by smoothing out noisy variation. This helps address potential issues of data quality. For instance, Facebook users at certain ages are more numerous than expected, so we have more confidence in our regularized estimates in those cases than in the original numbers. In this sense, our model is an alternative to nonparametric methods such as locally estimated scatterplot smoothing. Second, a model generates a range of potential outcomes, providing a measure of uncertainty. This allows us to assess the strength of our evidence. We can see how likely the differences we observe are to be substantively meaningful.

We include the four variables in our model as follows:

Sexuality: Our measure of sexual identity, the “interested in” profile field, is structured by binary gender. Combining responses to this field with the user’s gender as reported to advertisers, we recode this measure from “interested in men” and “interested in women” to “interested in the same gender” and “interested in a different gender.” We leave the remaining two categories, “interested in men and women” and “not specified,” unaltered. “Not specified” is an informative category, representing users who do not disclose any sexuality.

Gender: The gender categories available to advertisers are “women,” “men,” and “all,” but we present results only for women and men separately. Because of rounding, it is not possible to recover the numbers of nonbinary people or people with custom genders from the combined estimates of all Facebook users.

Relationship status: We restrict our data to the most common relationship statuses: “single” (20 percent), “in a relationship” (11 percent), “married” (28 percent), and “not specified” (41 percent). Together, these statuses account for 95 percent of Facebook users. We exclude individuals who specify other relationship statuses, for instance, “engaged,” “divorced,” or “it’s complicated.” This simplifies our model and avoids issues of data quality; for sexual minorities, especially at older ages, the numbers in those remaining categories are too small to produce stable or meaningful estimates.

Age: Although our other variables are categorical, we treat age as a continuous variable. Age has a nonlinear association with our outcome estimates, and a fifth-order polynomial produces the best fit. Only Facebook users between the ages of 18 and 64 are included in our model. Facebook’s advertising platform groups all people aged 65 and older into a single “65+” category, which is not directly compatible with a continuous
The operationalization of age; the simplest approach for modeling and interpretation is to exclude this 65+ category. Although Facebook is open to anyone older than 13, we did not gather data for any users younger than 18.

With these variables, we fit a Bayesian negative binomial model using the rstanarm package in R (Goodrich et al. 2020). A negative binomial model is more appropriate than a Poisson model because the counts are overdispersed. We take a Bayesian approach to estimation, with weakly informative priors, for several reasons. Adopting a Bayesian framework facilitates the simulation of potential data, allowing us to aggregate and transform our results when we present different aspects of our findings, while preserving quantile-based interval measures of uncertainty. Bayesian models have the further advantage of straightforward extensibility for future work that might incorporate measurement error or other data sources. Related Bayesian modeling strategies have been fruitfully applied to demographic research with Facebook advertising data (Alexander et al. 2020).

Because our model contains a large number of estimated coefficients, we do not present and interpret the parameters of the model individually. Instead, we present posterior medians and 95 percent posterior predictive intervals of estimated counts and proportions graphically. Importantly, posterior predictive intervals are wider than intervals based on the predicted means or expected values alone, making them a conservative way to examine evidence of differences (Goodrich et al. 2020). From this single underlying model, we aggregate posterior predictions to explore how disclosure rates for different sexual orientations vary by age, gender, relationship status, and finally by all of these characteristics together. Full information about our modeling approach, including comparisons to alternative models, is presented in the Appendix.

Results

We first describe the distribution of the outcome of interest, sexuality as measured by the “interested in” profile field on Facebook. Next, we show how the disclosure of sexuality is associated with the demographic attributes of age and gender. We then explore associations with the conceptually connected variable relationship status. Our model reveals that these three factors all matter and intersect in complex ways, so we close by considering them all together. Although the majority of the data are “missing,” the distribution of disclosure and nondisclosure is itself informatively patterned. On the basis of the “interested in” field, there are three possible sexual identities Facebook users might disclose. Whether individuals choose to disclose or not is informed by which sexual identity they hold, but we present results in two stages where relevant, first collapsing into the overall tendency to disclose any sexual identity and then separated into specific identity categories.

More than a quarter of U.S. Facebook users disclose their sexualities, aggregated from the model; Table A1 in the Appendix provides these and other descriptive estimates from the underlying data. We now turn to how the distribution of disclosure and nondisclosure varies by other basic demographic characteristics. Figure 3 shows the disclosure of sexualities by age, collapsing all identities together; here, it is clear that age has a strong and nonlinear association with the propensity to disclose a sexuality. Only 20 percent of 18-year-olds disclose any sexuality, but this rises sharply to 50 percent of those in their mid-20s. It falls again to 20 percent by age 40 and declines thereafter, to 10 percent by age 60.

Figure 2. Overall disclosure of sexualities by U.S. Facebook users.

Note: Data include women and men aged 18 to 64 with the four most common relationship statuses and were collected in September 2017 from the Facebook Marketing application programming interface (API). Posterior medians and 95 percent posterior predictive intervals are shown, aggregated from the full model shown in equations A1 to A5 in the Appendix.

Compared with age, gender alone has less of an association with the propensity to disclose any sexuality. Overall, women and men disclose their sexualities at similar rates, and this similarity in disclosure rates largely hold across the life course: although men may disclose at slightly higher rates in their late 20s and 30s, the predictive intervals for women and men overlap substantially, as shown in Figure 4a. However, the specific sexualities women and men disclose are not the same. Women are much more likely than men to be interested in both men and women, and they are slightly more likely to be interested in their own gender. Many more men, by contrast, explicitly express interest only in a different gender. Reframed in the language of sexual identity, women are more often openly bisexual or lesbian; the United States, 56.3 million (28 percent) specified the genders in which they were interested, while 143 million (72 percent) did not disclose information about their sexuality in this way. Among Facebook users who disclose their sexualities, we identify 4 million as sexual minorities: 1.68 million (0.8 percent) interested in their own gender and 2.21 million (1.1 percent) interested in both men and women. We construe as heterosexual the 52.4 million Facebook users (26.4 percent) who indicate an exclusive interest in a different gender. Figure 2 shows the rate of overall disclosure and what sexual identities are disclosed, aggregated from the model; Table A1 in the Appendix provides these and other descriptive estimates from the underlying data.
Disclosure rates diverge among older Facebook users. Single people at older ages continue to disclose at high rates, while disclosure rates among people in relationships fall slightly among older ages. Married users present a strong contrast: from age 40 and above, they disclose their sexualities at rates of 20 percent or below. Those who do not specify a relationship status are the least likely to specify a sexuality at all ages, though they too disclose at higher rates if they are young. Further examination reveals that these trends are driven largely by explicitly straight individuals, simply because the number of users at older ages who specify minority sexualities and relationship statuses such as “single” are quite small.

Age, gender, and relationship status interact to shape which Facebook users disclose a sexuality and which sexual identities they disclose. Young people and single people are more likely to disclose any sexuality, and women are more likely to be sexual minorities. At the same time, all these factors are conditionally dependent, which means they should be considered simultaneously. (For instance, women are more likely to be in relationships, whereas men are more likely to be single, which in turn shapes the distribution of relationship statuses across sexualities.) Figure 7 shows the full set of variables, with results presented in terms of counts rather than rates. Just like the proportions shown in previous figures, the counts are derived from a model that smooths the value of the estimates, rather than showing the underlying raw data (which are shown in the Appendix). This model is the best fit to the data; simplifying the model by dropping any interaction term, even three-way interactions, worsens model fit. Beyond the results we have already presented, Figure 7 shows various other demographic differences. For instance, although there are as many single gay men as lesbian women, there are more partnered lesbians than gay men. Across all disclosed sexual identities, relationship statuses also follow a life course pattern; the counts of single Facebook users have the youngest peak age, with older peaks for those in a relationship or married. Finally, because of the relatively high model-based uncertainty inherent in posterior predictive intervals (especially when estimated for one-year age groups), even distributions that appear to overlap somewhat have a high probability of actually being distinct. This means that, for example, young men are more likely than young women to be explicitly heterosexual and to leave their relationship status unstated; conversely, young women are more likely than young men to state that they are in a relationship or married while not disclosing their sexual identity. Numerous comparisons of this sort are possible.

Discussion

In this section, we interpret and contextualize our findings about the patterns of disclosure in this social media population using prior survey data, qualitative research, and social theory. Because we found strong interactions between different social categories, we pay particular attention to how explanations might vary according to the intersections among...
those categories (Collins 2015; Crenshaw 1989; Nelson forthcoming). As we discuss disclosure among different groups in turn—LGB and straight, younger and older, partnered and single, men and women—we expand and complicate our preceding interpretations.

Deriving estimates from responses to the “interested in” field, we found 4 million lesbian, gay, or bisexual (LGB) Facebook users. For comparison, Gallup and the Williams Institute estimate that there are 10 million lesbian, gay, bisexual, or transgender (LGBT) Americans older than 18 (Gates 2017). Of course, the categories used in the two data sources do not exactly correspond (sexual or romantic partner preference vs. a combined measure of sexual orientation and gender identity), and not all American adults are Facebook users (Auxier and Anderson 2021; Gramlich 2021). Still, an underlying population of 10 million LGB(T) people would imply a higher disclosure rate (approximately 40 percent) for LGB people on Facebook than for Facebook users overall (28 percent). This number, in fact, corresponds to the result of Pew’s nationally representative survey of LGBT Americans, which found that 4 in 10 LGBT adults overall, and 54 percent of those who use social networking sites, have disclosed their sexual identities on social media (Pew Research Center 2013). These numbers, along with the qualitative evidence from Duguay (2016), validate our belief that the “interested in” field is a strong indicator for sexual

Figure 4. Disclosure of any sexuality and specific sexual identities on Facebook, by gender and age.

Note: Data include the four most common relationship statuses and were collected in September 2017 from the Facebook Marketing application programming interface (API). Posterior medians and 95 percent posterior predictive intervals are shown.
identity among LGB people. Compared with heterosexuals, LGB people disclose their sexual identities on Facebook at relatively high rates. From this comparatively high disclosure rate, we conclude that many LGB people do find sexuality a salient facet of identity to manage and disclose in the online social context of their Facebook profiles; for them, the salience of sexual identity outweighs potential stigma or risk of sharing their status as a sexual minority. Of course, a large fraction of LGB people on Facebook still do not disclose their identities through the “interested in” field.

Accounting for age adds important nuance to this interpretation. Several recent surveys show that young people identify as LGB at high and increasing rates (England et al. 2016; Gates 2014, 2017; Jones 2021), and we find that young people are also disproportionately likely to disclose sexuality-related information on Facebook, which could drive part of the high LGB disclosure rate we previously discussed. What this generational divide in online disclosure might mean more broadly is not clear cut, and the possibilities are not mutually exclusive. Young people could be willing to share their sexualities online because sexual identity is generally more salient to them, or because sexuality overall is less subject to stigma for younger cohorts. Sharing sexuality information online may also be a practical matter for those in a life stage when seeking sexual and romantic partners is common. By contrast, older cohorts on Facebook may not find sexuality a salient axis of identity to express, may perceive expressing information related to sexuality to be taboo, or else may not find sharing their sexual identity to be practically relevant.

The youngest adult Facebook users, especially those younger than 20, present a puzzle. Rather than disclosing at the high rate of the cohort just ahead of them, they appear more similar to those in their mid-30s and beyond who share their sexual orientations on their profiles less often. However, we do not think 18- and 19-year-olds find sexuality irrelevant or inappropriate to disclose in social contexts. Some young people may find sexual identities expressed strictly in terms of binary gender to be unnecessarily limiting, preferring to identify as queer or pansexual rather than LGB or straight (Duguay 2016; Hammack et al. forthcoming), though bisexual identity is even more common among the youngest generations (Jones 2021). More significantly, we suspect that, either because of privacy concerns or because of disengagement with Facebook as a social media platform, these youngest Facebook users are more generally reluctant to share personal information in their Facebook profiles. Where previous cohorts of youth may have struggled with “context collapse” online between their peers and their parents (boyd 2014), teenagers may have become increasingly adept, on average, at online impression management, presenting more curated facets of their identities to parents and other adults (Rafalow 2020).

Examining the interaction of age and relationship status helps disambiguate the practical and expressive aspects of disclosure: on Facebook, young people do not condition whether they disclose their sexuality on the basis of whether they are presently single or partnered. Among both LGB and straight users, young people use the “interested in” field to express sexual identity, not to signal availability. Of course, the relationship statuses of young people may also be in greater flux, making them less likely to micromanage their “interested in” profile field accordingly.

Unlike people in their 20s, older cohorts’ willingness to signal their sexual identity is highly dependent on other characteristics. In particular, among many older users, being married appears to serve as a substitute signal of sexual identity. Because the majority of marriages are heterosexual, marital status itself works to convey the unmarked status of
heterosexuality. The fact that older single Facebook users disclose like their younger counterparts suggests that, for them, the “interested in” field signals not identity but availability.

Gender differences add yet another layer of complexity to our explanation above. For sexual minorities, the gender differences we observe in online disclosure are consistent with estimates of prevalence from nationally representative surveys. Specifically, a large number of those who express LGB sexual identities on Facebook are young women interested in both men and women, far outnumbering young bisexual men. This result echoes a key finding from the 2002–2013 National Survey of Family Growth, in which England et al. (2016) showed that the increasing number of young women identifying as bisexual is large enough to drive an observed overall increase in LGB identity over time. Previously, we observed a close correspondence between national surveys and our data about sexual minorities, because of the relatively high salience of sexual identity disclosure for LGB people online. In this case, we also conclude that our finding about the large number of young bisexual women reflects not only a sociological fact about disclosure but also a demographic fact about the distribution of sexual identities in the population.

Although young women are more likely to disclose a sexual minority identity than men, we find that at all ages men are more likely than women to explicitly signal that they are heterosexual. We argue that demographic prevalence is not the primary driver of this heterosexual gender disparity: the number of additional bisexual and lesbian women does not equal the magnitude of the gap between explicitly heterosexual men and explicitly heterosexual women. Instead, we attribute this disparity to the powerful interaction between heterosexuality and masculinity. As Pascoe (2011) showed, a central goal of the public social performance of heterosexual masculinity is to enable straight boys and men to avoid being perceived or labeled as gay. Ward (2015) and Carrillo and Hoffman (2018) went further, showing that even when

---

**Figure 7.** Estimated counts of U.S. Facebook users of each sexual identity, relationship status, age, and gender.

*Note: Data were collected in September 2017 from the Facebook Marketing application programming interface (API). Posterior medians and 95 percent posterior predictive intervals are shown.*
heterosexual men engage in homosexual or homoerotic behavior, they aim to hold onto the privileges associated with straight identity. Our finding that young men who explicitly indicate their interest in women also disproportionately disclose that they are single, or omit their relationship status, fits squarely into this framework of masculinities studies and critical heterosexualities studies. These men are using the “interested in” field as an opportunity to signal their heterosexual identity and often their heterosexual availability. The audience for this social signaling is not only women who might be potential sexual or romantic partners but also male members of their social worlds who might otherwise call their masculinity into question. (In fact, it could be primarily the latter.) “Interested in women” lays claim to straight privilege (Dean 2014) and meets potential peer pressure for publicly performing heterosexuality (Duckworth and Trautner 2019). This social media profile field affords a low-effort way to signal heterosexuality, while potentially avoiding costs that may now come with being perceived as homophobic (Dean 2014). Altogether, the demands of heterosexual masculinity reasonably explain why straight men might be unusually motivated to disclose their sexualities.

We see the relative absence of openly heterosexual women through the flip side of the same lens. Both online and offline, women may face violent, sexist, or sexual harassment from men in public and private social spaces (Amundsen 2021; Nakamura 2019; Rubin, Blackwell, and Conley 2020). This risk for harassment shapes how and whether women, and heterosexual women in particular, disclose their sexualities. We find that a disproportionate number of women who leave their “interested in” information unspecified also report themselves to be in relationships or married. Although (explicitly) straight men choose the opposite responses (single or relationship status unspecified) to reinforce their heterosexuality and signal their availability, women are more likely to avoid signaling availability. This gap in sexuality disclosure between women and men is a uniquely heterosexual phenomenon, rather than purely a gendered one. As we discussed previously, LGB disclosure rates are generally high, there are slightly more lesbian women than gay men, and there are many more bisexual women than bisexual men. In a heterosexual context, the aggregate behavior of potentially straight women strikes us not as an anomaly, but as a rational response to the behaviors of straight men.

To sum up, our findings should be understood in both demographic and sociological terms. Relatively speaking, for millions of LGB people and young people, Facebook’s “interested in” field represents sexual identity, and the salience of this identity appears to outweigh the potential stigma of disclosing a sexual minority identity. Consequently, for these subpopulations Facebook profile information provides a reasonable demographic proxy about sexual identity. Given the wide uptake of Facebook in the United States, our findings correspond to population trends beyond the platform’s users (Ribeiro et al. 2020). By contrast, among heterosexual or older Facebook users, the “interested in” question is sociologically entangled with relationship statuses, revealing as much about gender dynamics and sexual or romantic availability as it does about sexuality or identity.

**Conclusion**

We have analyzed the disclosure of social information about sexuality in a digitally mediated social context. Tens of millions of people in the United States have used Facebook’s “interested in” profile field to disclose their sexualities. Digital trace data allow us to observe the disclosure behavior not only of LGB youth (Duguay 2016), for instance, but also of heterosexual people whose sexual identities may be taken for granted. And although the design of the “interested in” field is outside of our control as researchers, as digital trace data generally are (Salganik 2018), it is a vehicle for people to express their sexualities online. As the digitalization of our lives has generated new spaces for social expression and interaction, we have shown that careful measurement and analysis of digital traces can uncover some of the complexity of social lives and identities.

There are several important limitations to this work. First, the data are cross-sectional, descriptive, and aggregate. This means that we cannot discern the meanings of social disclosures at the individual level. Nor can we analytically disentangle age, period, and cohort effects from cross-sectional data alone. For instance, we might expect the mid-20s cohort to maintain their high sexual identity disclosure rate as they age and form partnerships, while 18-year-olds may continue to find Facebook profiles less salient as a vehicle for identity disclosure as they age. Accordingly, a second limitation is that our findings are specific to Facebook and the time period of our data collection (2017); they do not necessarily generalize to other platforms. General social trends such as attitudinal changes toward privacy and data collection, or migration to new social media platforms with different architectures, might affect future disclosure patterns.

Finally, fundamental data limitations constrained which characteristics we were able to measure, limiting our ability to account for the full complexity of sexual identity. For instance, we could not distinguish asexual identities from nondisclosure. Asexual identities are increasingly visible and salient (Carroll 2020), but our study reinforces their invisibility. Nor could we account for other characteristics known to intersect with sexual orientation in the United States, such as race (Silva and Evans 2020). Gender identity also interacts with sexual identity (James et al. 2016), but we could not separate out cisgender and transgender women and men; nonbinary people were excluded from our count data entirely (Bivens 2017; Bivens and Haimson 2016). This last limitation means that the present study reinforces a binary understanding of gender, furthering what Bivens (2017) labeled symbolic violence against nonbinary people. This is a serious shortcoming, albeit one shared by many quantitative studies.
of sexual orientation and gender identity. Others have argued that this issue of “data violence” (Hoffmann 2018) may be intrinsic and insurmountable (Keyes 2019).

Nevertheless, our work gives us a foundation for assessing which directions for future research are promising and which might be less viable. For instance, platform changes undermine the possibility of continuous monitoring for temporal comparison. The removal of the “interested in” field as an ad-targeting option as of early 2018, together with decreasing researcher access to APIs more generally in the “post-API age” (Freelon 2018), forecloses the possibility of conducting an ongoing digital census of sexuality data. As a potential solution, we believe that platform-based surveys, whether microtargeted or aimed at characterizing a platform’s entire user base, hold greater potential moving forward as a way to understand their user populations. These surveys could combine the best of digital and traditional methods (Salganik 2018); to survey sexual identity and its expression on social media, they would also need to incorporate best practices in survey design for sexual orientation and gender identity (Lagos and Compton 2021; Westbrook, Budnick, and Saperstein 2021).

Algorithmically inferred categorizations are on the rise on social media platforms (Simpson and Semaan 2021), but we do not believe these will be as useful for studying sexual identity in digital contexts as explicitly user-disclosed information. For example, a social media user might be classified as liking “LGBTQ community.” This could be a basis for targeted advertising, but it would tell us little about deliberate disclosure or conscious identification. Moreover, the attendant harms of algorithm ascription for LGBTQ+ people have already been noted in rising algorithm-centric platforms such as TikTok (Simpson and Semaan 2021). To study social disclosure, researchers might focus instead on how marked (i.e., LGBQA+) sexual identities are made visible in online social contexts, such as through explicit identification in symbols, emojis, or unstructured text. Initial promising work in this vein already exists (Andalibi 2019; Haimson, and Veinot 2020).

Of equal importance are the experiences of the unmarked heterosexual majority. We interpret gendered disparities in heterosexual disclosure behavior in terms of theories of masculinity and critical heterosexuality studies. We believe that further work should consider interviews with heterosexual social media users to confirm or complicate this interpretation, with implications for understanding when heterosexuality becomes salient to disclose, and for promoting the well-being of straight women online.

We began this project optimistic about the potentials of digital trace data for extending the bounds of knowledge about sexuality. Even in the relatively short time since we collected the data presented here, the world has changed. Detailed data collection that fuels targeted advertising still undergirds much of the contemporary social Internet, but public attitudes have shifted, and it is unclear if this corporate model is viable in the long term. For researchers, this may mean disruptive changes to digital data sources and how publicly accessible they are, even as online and offline lives come to increasingly overlap for individuals. We believe that these changes make it all the more urgent and valuable, at any given moment, to document and understand online social life through digital trace data. And though digital data ethics is an unsettled subject that will continue to evolve beyond this writing, we believe we have undertaken this work with minimal symbolic or material harm, especially toward sexual minorities and other marginalized social groups. We hope to have shown that even when data are partial and flawed, the effort to make meaning out of them remains worthwhile.

Appendix: Descriptive Statistics and Model Comparisons

Table A1 summarizes univariate descriptive statistics for the data presented in the main text. In aggregate, 199 million U.S. Facebook users are included in the analysis, and 32.9 million more are included in the collected data but excluded from the analysis.

The final data set comprises 1,504 cells of count values, stratified by sexuality (four possible values, including “not specified”), gender (two values), relationship status (four values, including “not specified”), and age (47 values). The cell values are counts rounded to two significant figures. The largest observed estimate in our data is 1.9 million, representing a true count value between 1,850,001 and 1,949,999 inclusive. The smallest observed estimate is 280, representing a value between 275 and 285 inclusive. This implies that stratifying Facebook Marketing API queries into smaller categories results in more precise estimates. Accordingly, all broader aggregations in the main text are derived from the original stratified estimates, rather than queried anew from the API.

As discussed in the main text, nonbinary Facebook users and users with other custom gender values cannot be targeted through the Facebook advertising platform, and so are excluded from our estimated counts. Because of rounding, it is not possible to recover the number of nonbinary Facebook users by subtracting the counts for men and women separately from the counts of all Facebook users. From the data we collected, we also exclude less common relationship statuses (“engaged,” “divorced,” “widowed,” and so on, constituting nine additional categories) and users age 65 and older from our model and presentation in the main text.

We make claims about an online population of user accounts in terms of human behavior. The main plausible systematic source of error in the data would be inauthentic user accounts, that is, accounts not corresponding to individual humans. We in fact observe unusually large numbers of users with limited account information—both sexuality and relationship status unspecified—at ages ending in seven (27, 37, and so on), corresponding to birth years in 1990, 1980, and so on. For these observations we believe our model provides a more reliable estimate than the underlying data. We considered removing and interpolating those
observations, but judged the smoothing induced by the polynomial age coefficients to be adequate.

To the four-way contingency table of the data, we fit a log-linear model (Agresti 2012). A log-linear model is a generalized linear model for count data, with an analytic focus on the interactions between variables. The goal is to determine which variables are associated with each other in the data and which variables might be mutually, jointly, or conditionally independent. What we found, however, is that none of the relationships between sexuality (S), gender (G), relationship status (R), and age (A) can be simplified or removed. (See Table A2 for Bayesian information criterion [BIC] comparisons of different models.) Following Agresti (2012), we represent our best-fitting model with the shorthand notation SGR, SGA, SRA, GRA, which nests all lower-order interactions and main terms. This is model 13 in Table A2, with the lowest BIC value, written in expanded notation in equation A2. A four-way interaction, SGRA, is unnecessary according to this measure of model fit. (Note that in a four-variable model, this interaction term would saturate the model, if age is encoded categorically, and cannot be fit with maximum likelihood estimation. Compare model 15 with model 16 in Table A2.) We compare candidate models fit with maximum likelihood estimation for speed and efficiency and refit our final selected model in a Bayesian estimation framework.

We extend the basic log-linear modeling framework in two ways. First, we use a negative binomial distribution rather than a Poisson distribution to model the response variable $y_{hijk}$, the count in each cell, because the counts are overdispersed; the variance is larger than the mean. In the final Bayesian model, the reciprocal dispersion parameter $\frac{1}{\phi}$ is estimated to be 67.4 (95 percent credible interval = 62.2–72.8); an estimate near zero would have indicated no overdispersion. Second, although log-linear models typically include only categorical covariates, we model age as a continuous variable with nonlinear terms. This results in a more parsimonious model, one that is preferred by measures of model fit such as the BIC (see Table A2 for further comparisons).

We then fit the following Bayesian model, using the weakly informative priors recommended by the Stan developers (Goodrich et al. 2020):

$$y_{hijk} \sim \text{NegativeBinomial}(\mu, \phi)$$

$$\log(\mu_{hijk}) = \lambda_0 + \lambda_S^S + \lambda_G^G + \lambda_R^R + \lambda_{SG}^S + \lambda_{SR}^R + \lambda_{GR}^G + \lambda_{SGR}$$

(1)
\[ \lambda_0 \sim \text{Normal}(0, 10) \]  
(3)

\[ \lambda_{(k)(i)(j)(e)} \sim \text{Normal}\left(0, \frac{2.5}{s_x}\right) \]  
(4)

\[ \frac{1}{\phi} \sim \text{Exponential}(1) \]  
(5)

\( \lambda_0 \) is the intercept, and the other \( \lambda \) coefficients represent indicators for categories and their interactions. Parameters for reference categories are constrained to equal zero:

\[ \lambda^S_{HJ} = \lambda^G_J = \ldots = \lambda^{SG}_{HI} = \ldots = \lambda^{SGR}_{HIJ} = \ldots = 0 \]

The scale parameter for each coefficient \( \lambda_{(k)(i)(j)(e)} \) is rescaled by dividing by the standard deviation of the corresponding centered covariate \( s_x \). The polynomial terms are orthogonal polynomials.

To fit this model, we draw samples from the posterior distribution with a Markov-chain Monte Carlo method, specifically the No U-Turn Sampler, using the rstanarm R package (Goodrich et al. 2020). The resulting fit consists of 4,000 draws from four Markov chains, each run for 2,000 iterations; the first half of each chain is discarded as a warm-up. Model diagnostics indicate acceptable convergence and model fit. \( R \) values for every parameter are less than 1.01 (the largest is 1.003); parameter effective sample sizes are generally greater than 0.5 (i.e., 2,000 effective samples). According to the pareto-\( k \) diagnostic, the model fits 11 observations poorly, with \( k > 0.7 \) (out of 1,504 observations). A further 36 observations have \( k \) values between 0.5 and 0.7. These observations are at the tails of the age distribution, especially the upper tail, suggesting that the polynomial specification for age in the model may be misspecified for ages very close to 18 or 64. For computational efficiency, we do not replicate every comparison in Table A2, in part because Gelman et al. (2020) suggested it is not necessary to estimate models already known to fit poorly with Markov-chain Monte Carlo. However, select Bayesian model comparisons with PSIS-LOO and 10-fold cross-validation confirm that this model is the best fit among the models under consideration.

| Model | Notation | Description | Age Encoding | Parameters | Deviance | BIC  |
|-------|----------|-------------|--------------|------------|----------|------|
| M1    | (Intercept only) | Null model | —            | 1          | 2,001.51 | 36,966.50 |
| M2    | S, G, R, A | No interactions | Fifth-order polynomial | 13 | 1,601.13 | 33,281.04 |
| M3    | SG, SR, GR, SA, GA, RA | All two-way interactions | Fifth-order polynomial | 63 | 1,519.78 | 30,322.38 |
| M4    | SGR, SA, GA, RA | One three-way interaction | Fifth-order polynomial | 72 | 1,520.09 | 30,228.41 |
| M5    | GRA, SG, SR, SA | One three-way interaction | Fifth-order polynomial | 78 | 1,519.26 | 30,323.39 |
| M6    | SGA, SR, GR, RA | One three-way interaction | Fifth-order polynomial | 78 | 1,518.58 | 30,130.17 |
| M7    | SRA, SG, GR, GA | One three-way interaction | Fifth-order polynomial | 108 | 1,520.25 | 29,851.11 |
| M8    | SGR, SGA, GRA | Three three-way interactions | Fifth-order polynomial | 102 | 1,519.16 | 29,980.28 |
| M9    | SGR, SRA, GRA | Three three-way interactions | Fifth-order polynomial | 132 | 1,527.59 | 29,358.77 |
| M10   | SGR, SGA, SRA | Three three-way interactions | Fifth-order polynomial | 132 | 1,535.08 | 29,129.07 |
| M11   | SGA, SRA, SRA | Three three-way interactions | Fifth-order polynomial | 138 | 1,534.65 | 29,287.21 |
| M12   | SGR, SGA, SRA, GRA | All three-way interactions | Third-order polynomial | 101 | 1,516.17 | 30,309.65 |
| M13   | SGR, SGA, SRA, GRA | All three-way interactions | Fifth-order polynomial | 147 | 1,551.24 | 28,872.35 |
| M14   | SGR, SGA, SRA, GRA | All three-way interactions | Categorical | 1,090 | 1,708.46 | 33,763.13 |
| M15   | SGR | Four-way interaction | Fifth-order polynomial | 192 | 1,559.93 | 28,971.68 |
| M16   | SGR | Saturated model | Categorical | 1,504 | 0.00 | — |

Table A2. Model Comparisons.
To present model results, we combine the samples from the posterior distribution. We summarize the predicted value for each combination of variables with the posterior median using the tidybayes R package (Kay 2020). We use the posterior predictive distribution to convey uncertainty, through 95 percent quantile-based posterior predictive intervals. Compared with credible intervals, posterior predictive intervals incorporate the additional uncertainty from the likelihood of the generative model; Lynch and Bartlett (2019) among others have discussed the merits of this approach. To present higher level summaries, we simply aggregate the samples before calculating the point estimates and intervals. Where appropriate, we transform these point and interval values from counts to proportions by dividing by group sums. To convey the fit of our model graphically, Figure A1 replicates Figure 7 in the main text and additionally shows the actual estimates in the data alongside the model predictions.

All research code for data collection, data processing, statistical modeling, and visualization is available at https://github.com/ccgilroy/digital-traces-sexualities.

Acknowledgments

We would like to thank Emilio Zagheni for the initial inspiration to work with this type of digital trace data. We are indebted to the intellectual guidance of Kate Stovel, through extensive conversations and detailed comments, which led to a stronger sociological understanding and framing for this article. Mark Igra, Emma Mishel, Tristan Bridges, Hannah Curtis, Daiki Hiramori, and panelists and audiences at Population Association of America and American Sociological Association conferences, also provided valuable feedback on various iterations of this project. We thank two anonymous reviewers at Socius for their thoughtful comments. We would also like to gratefully acknowledge the 2017 Russell Sage Foundation Summer Institute in Computational Social Science for providing a vibrant space for

Figure A1. Underlying data estimates of U.S. Facebook users of each sexual identity, relationship status, age, and gender (black lines), overlaid on full model estimates from Figure 7. Note: Data were collected in September 2017 from the Facebook Marketing application programming interface (API). For model estimates, posterior medians and 95 percent posterior predictive intervals are shown.
the exchange of knowledge and ideas, including the genesis of this work.

Declaration of Conflicting Interests

Connor Gilroy was a research intern at Facebook in the summer of 2020, with no overlap with this project. All data were collected and all analyses performed before the start of the internship, and no work on the writing of the article was undertaken during the duration of the internship. No nonpublic knowledge or materials from Facebook are contained in this article.

Funding

Connor Gilroy’s participation in this project was funded by a National Institutes of Health Big Data to Knowledge traineeship (grant 3T32HD007543-15S1), through the Center for Studies in Demography and Ecology (CSDE) at the University of Washington, and also partially supported by a Shanahan Endowment Fellowship, a Eunice Kennedy Shriver National Institute of Child Health and Human Development research infrastructure grant (P2C HD042828) to the Center for Studies in Demography & Ecology, a Training Grant (T32 HD101442-01) and the Schmidt Endowment. Ridhi Kashyap would like to acknowledge support from the Leverhulme Grant (T32 HD101442-01) and the Schmidt Endowment. Connor Gilroy was a research intern at Facebook in the summer of 2020, with no overlap with this project. All data were collected and all analyses performed before the start of the internship, and no work on the writing of the article was undertaken during the duration of the internship. No nonpublic knowledge or materials from Facebook are contained in this article.

Research Materials

Code and materials are available at https://github.com/ccgilroy/digital-traces-sexualities.

ORCID iD

Connor Gilroy https://orcid.org/0000-0003-0367-9373

References

Agresti, Alan. 2012. *Categorical Data Analysis*. Somerset, UK: Wiley.

Alexander, Monica, Kivan Polimis, and Emilio Zagheni. 2020. “Combining Social Media and Survey Data to Nowcast Migrant Stocks in the United States.” *Population Research and Policy Review*. Retrieved June 26, 2021. https://link.springer.com/article/10.1007/s11113-020-09599-3.

Amundsen, Rikke. 2021. “‘A Male Dominance Kind of Vibe’: Approaching Unsolicited Dick Pics as Sexism.” *New Media & Society* 23(6):1465–80.

Andalibi, Nazanin. 2019. “What Happens after Disclosing Stigmatized Experiences on Identified Social Media: Individual, Dyadic, and Social/Network Outcomes.” Pp 1–15 in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI ’19*. New York: Association for Computing Machinery.

Auxier, Brooke, and Monica Anderson. 2021. “Social Media Use in 2021.” Retrieved June 3, 2021. https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/.

Bivens, Rena. 2017. “The Gender Binary Will Not Be Deprogrammed: Ten Years of Coding Gender on Facebook.” *New Media & Society* 19(6):880–98.

Bivens, Rena, and Oliver L. Haimson. 2016. “Baking Gender into Social Media Design: How Platforms Shape Categories for Users and Advertisers.” *Social Media + Society* 2(4):1–12.

boyd, danah. 2014. *It’s Complicated: The Social Lives of Networked Teens*. New Haven, CT: Yale University Press.

Berekhs, Wayne. 1996. “Social Marking and the Mental Coloring of Identity: Sexual Identity Construction and Maintenance in the United States.” *Sociological Forum* 11(3):497–522.

Budnick, Jamie. 2016. “‘Straight Girls Kissing?’ Understanding Same-Gender Sexuality Beyond the Elite College Campus.” *Gender & Society* 30(5):745–68.

Carrillo, Héctor, and Amanda Hoffman. 2018. “‘Straight with a Pinch of Bi’: The Construction of Heterosexuality as an Elastic Category among Adult US Men.” *Sexualities* 21(1–2):90–108.

Carroll, Megan. 2020. “What Can Asexuality Offer Sociology? Insights from the 2017 Asexual Community Census.” *SocArXiv*. Retrieved June 26, 2021. https://osf.io/preprints/socarxiv/bh7t3/.

Cesare, Nina, Hedwig Lee, Tyler McCormick, Emma Spiro, and Emilio Zagheni. 2018. “Promises and Pitfalls of Using Digital Traces for Demographic Research.” *Demography* 55(5):1979–99.

Collins, Patricia Hill. 2015. “Intersectionality’s Definitional Dilemmas.” *Annual Review of Sociology* 41(1):1–20.

Crenshaw, Kimberlé. 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):139–67.

Dean, James Joseph. 2014. *Straights: Heterosexuality in Post-closed Culture*. New York: NYU Press.

Doan, Long, and Trenton D. Mize. 2020. “Sexual Identity Disclosure among Lesbian, Gay, and Bisexual Individuals.” *Sociological Science* 7:504–27.

Duckworth, Kiera D., and Mary Nell Trautner. 2019. “Gender Goals: Defining Masculinity and Navigating Peer Pressure to Engage in Sexual Activity.” *Gender & Society* 33(5):795–817.

Duguay, Stefanie. 2016. “‘He Has a Way Gayer Facebook Than I Do’: Investigating Sexual Identity Disclosure and Context Collapse on a Social Networking Site.” *New Media & Society* 18(6):891–907.

England, Paula, Emma Mishel, and Monica L. Caudillo. 2016. “Increases in Sex with Same-Sex Partners and Bisexual Identity across Cohorts of Women (but Not Men).” *Sociological Science* 3:951–70.

Fatekhia, Masoomali, Ridhi Kashyap, and Ingrid Weber. 2018. “Using Facebook Ad Data to Track the Global Digital Gender Gap.” *World Development* 107:189–209.

Fiesler, Casey, and Nicholas Proferes. 2018. “‘Participant’ Perceptions of Twitter Research Ethics.” *Social Media + Society* 4(1):2056305118763366.

Freelon, Deen. 2018. “Computational Research in the Post-API Age.” *Political Communication* 35(4):665–68.

Gates, Gary J. 2014. “LGBT Demographics: Comparisons among Population-Based Surveys.” *Los Angeles: Williams Institute, UCLA School of Law*.
Gates, Gary J. 2017. “In US, More Adults Identifying as LGBT.” Retrieved April 29, 2021. https://news.gallup.com/poll/201731/lgbt-identification-rises.aspx.

Gelman, Andrew, Aki Vehtari, Daniel Simpson, Charles C. Margossian, Bob Carpenter, Yuling Yao, Lauren Kennedy, Jonah Gabry, Paul-Christian Bürkner, and Martin Modrak. 2020. “Bayesian Workflow.” arXiv. Retrieved April 29, 2021. http://arxiv.org/abs/2011.01808.

Goffman, Erving. 1963. Stigma: Notes on the Management of Spoiled Identity. Englewood Cliffs, NJ: Prentice-Hall.

Goldman, Rob, and Alex Himel. 2018. “Making Ads and Pages More Transparent.” Retrieved April 29, 2021. https://newsroom.fb.com/news/2018/04/transparent-ads-and-pages/.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2020. “rtstanarm: Bayesian Applied Regression Modeling via Stan.” Retrieved April 29, 2021. https://mc-stan.org/rtstanarm.

Gramlich, John. 2021. “10 Facts about Americans and Facebook.” Retrieved June 3, 2021. https://www.pewresearch.org/fact-tank/2021/06/01/facts-about-americans-and-facebook/.

Haimson, Oliver L., and Anna Lauren Hoffmann. 2016. “Constructing and Enforcing ‘Authentic’ Identity Online: Facebook, Real Names, and Non-normative Identities.” First Monday 21(6).

Haimson, Oliver L., and Tiffany C. Veinot. 2020. “Coming Out to Doctors, Coming Out to ‘Everyone’: Understanding the Average Sequence of Transgender Identity Disclosures Using Social Media Data.” Transgender Health 5(3):158–65.

Hammack, Phillip L., Sam D. Hughes, Julianne M. Atwood, Elliot M. Cohen, and Richard C. Clark. Forthcoming. “Gender and Sexual Identity in Adolescence: A Mixed-Methods Study of Labeling in Diverse Community Settings.” Journal of Adolescent Research. https://doi.org/10.1177/07311214211000315.

Hoffmann, Anna Lauren. 2018. “Data Violence and How Bad Engineering Choices Can Damage Society.” Retrieved April 29, 2021. https://medium.com/s/story/data-violence-and-how-bad-engineering-choices-can-damage-society-39e44150e1d4.

James, Sandy, Jody Herman, Susan Rankin, Mara Keisling, Lisa Mottet, and Ma’ayan Anafi. 2016. “The Report of the 2015 US Transgender Survey.” Washington, DC: National Center for Transgender Equality.

Jones, Jeffrey M. 2021. “LGBT Identification Rises to 5.6% in Latest U.S. Estimate.” Retrieved April 29, 2021. https://news.gallup.com/poll/329708/lgbt-identification-rises-latest-estimate.aspx.

Jurgenson, Nathan. 2011. “Digital Dualism versus Augmented Reality.” Retrieved April 29, 2021. https://thesocietypages.org/cyborgology/2011/02/24/digital-dualism-versus-augmented-reality/.

Kashyap, Ridhi, Masoomali Fatehkia, Reham Al Tamime, and Ingmar Weber. 2020. “Monitoring Global Digital Gender Inequality Using the Online Populations of Facebook and Google.” Demographic Research 43(27):779–816.

Kay, Matthew. 2020. “tidybayes: Tidy Data and ‘Geoms’ for Bayesian Models.” Retrieved June 26, 2021. https://cran.r-project.org/web/packages/tidybayes/index.html.

Keyes, Os. 2019. “Counting the Countless.” Real Life, April 8. Retrieved April 29, 2021. https://reallifemag.com/counting-the-countless/.

Lagos, Danya, and D’Lane Compton. 2021. “Evaluating the Use of a Two-Step Gender Identity Measure in the 2018 General Social Survey.” Demography 58(2):763–72.

Laumann, Edward O., John H. Gagnon, Robert T. Michael, and Stuart Michaels. 1994. The Social Organization of Sexuality: Sexual Practices in the United States. Chicago: University of Chicago Press.

Lynch, Scott M., and Bryce Bartlett. 2019. “Bayesian Statistics in Sociology: Past, Present, and Future.” Annual Review of Sociology 45:47–68.

McCall, Leslie. 2005. “The Complexity of Intersectionality.” Signs: Journal of Women in Culture and Society 30(3):1771–1800.

Mishel, Emma, Tristan Bridges, and Mónica L. Caudillo. Forthcoming. “Google, Tell Me. Is He Gay? Masculinity, Heterosexuality, and Gendered Anxieties in Google Search Queries about Sexuality.” Sociological Perspectives. https://doi.org/10.1177/07311242211001906.

Mishel, Emma, Paula England, Jessie Ford, and Mónica L. Caudillo. 2020. “Cohort Increases in Sex with Same-Sex Partners: Do Trends Vary by Gender, Race, and Class?” Gender & Society 34(2):178–209.

Nakamura, Lisa. 2019. “Gender and Race in the Gaming World.” In Society and the Internet: How Networks of Information and Communication Are Changing Our Lives, edited by M. Graham and W. H. Dutton. Oxford, UK: Oxford University Press.

Nelson, Laura K. Forthcoming. “Leveraging the Alignment between Machine Learning and Intersectionality: Using Word Embeddings to Measure Intersectional Experiences of the Nineteenth Century U.S. South.” Poetics. https://doi.org/10.1016/j.poetic.2021.101539.

Orne, Jason. 2011. “‘You Will Always Have to “Out” Yourself’: Reconsidering Coming Out through Strategic Outness.” Sexualities 14(6):681–703.

Orne, Jason. 2013. “Queers in the Line of Fire: Goffman’s Stigma Revisited.” Sociological Quarterly 54(2):229–53.

Orne, Jason. 2017. Boystown: Sex and Community in Chicago. Chicago: University of Chicago Press.

Pascoe, C. J. 2011. Dude, You’re a Fag: Masculinity and Sexuality in High School. Berkeley: University of California Press.

Pascoe, C. J., and Sarah Diefendorf. 2018. “No Homo: Gendered Dimensions of Homophobic Epithets Online.” Sex Roles 80:123–36.

Pew Research Center. 2013. “A Survey of LGBT Americans.” Retrieved April 29, 2021. https://www.pewresearch.org/social-trends/2013/06/13/a-survey-of-lgbt-americans/.

Rafalow, Matthew H. 2020. Digital Divisions. Chicago: University of Chicago Press.

Randazzo, Francesco, Emilio Zagheni, Ingmar Weber, Maria Rita Testa, and Francesco Billari. 2018. “Mater Certa Est, Pater Sterilis: What Can Facebook Advertising Data Tell Us about Male Fertility Rates?” Proceedings of the International AAAI Conference on Web and Social Media 12(1, 1).

Ribeiro, Filipe N., Fabricio Benevenuto, and Emilio Zagheni. 2020. “How Biased Is the Population of Facebook Users? Comparing the Demographics of Facebook Users with Census Data to Generate Correction Factors.” Pp 325–34 in 12th ACM Conference on Web Science. Southampton, United Kingdom: Association for Computing Machinery.

Rubin, Jennifer D., Lindsay Blackwell, and Terri D. Conley. 2020. “Fragile Masculinity: Men, Gender, and Online Harassment.” Pp 1–14 in Proceedings of the 2020 CHI Conference on
Salganik, Matthew J. 2018. *Bit by Bit: Social Research in the Digital Age.* Princeton, NJ: Princeton University Press.

Schnabel, Landon. 2018. “Sexual Orientation and Social Attitudes.” *Socius* 4. Retrieved June 26, 2021. https://journals.sagepub.com/doi/full/10.1177/2378023118769550.

Schnabel, Landon. 2018. “Sexual Orientation and Social Attitudes.” *Socius* 4. Retrieved June 26, 2021. https://journals.sagepub.com/doi/full/10.1177/2378023118769550.

Silva, Tony. 2017. “Bud-Sex: Constructing Normative Masculinity among Rural Straight Men That Have Sex with Men.” *Gender & Society* 31(1):51–73.

Silva, Tony. 2018. “A Quantitative Test of Critical Heterosexuality Theory: Predicting Straight Identification in a Nationally Representative Sample.” *Sexuality Research and Social Policy* 15:353–66.

Silva, Tony, and Clare R. Evans. 2020. “Sexual Identification in the United States at the Intersections of Gender, Race/Ethnicity, Immigration, and Education.” *Sex Roles* 83(11–12): 722–38.

Simpson, Ellen, and Bryan Semaan. 2021. “For You, or For ‘You’? Everyday LGBTQ+ Encounters with TikTok.” *Proceedings of the ACM on Human-Computer Interaction* 4(252):1–34.

Sweeney, Latanya. 2002. “k-Anonymity: A Model for Protecting Privacy.” *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 10(5):557–70.

Ward, Jane. 2015. *Not Gay: Sex between Straight White Men.* New York: NYU Press.

Ward, Jane. 2020. *The Tragedy of Heterosexuality.* New York: NYU Press.

Westbrook, Laurel, Jamie Budnick, and Aliya Saperstein. Forthcoming. “Dangerous Data: Seeing Social Surveys through the Sexuality Prism.” *Sexualities.* https://doi.org/10.1177/1363460720986927.

Zagheni, Emilio, Ingmar Weber, and Krishna Gummadi. 2017. “Leveraging Facebook’s Advertising Platform to Monitor Stocks of Migrants.” *Population and Development Review* 43(4):721–34.

Zerubavel, Eviatar. 2018. *Taken for Granted: The Remarkable Power of the Unremarkable.* Princeton, NJ: Princeton University Press.

**Author Biographies**

**Connor Gilroy** is a PhD student in sociology at the University of Washington. He uses digital and computational methods to study experiences of community, visibility, and belonging, with a substantive focus on queer communities and LGBTQ populations. His research projects examine change in gay neighborhoods, identities and interactions on social media, and the meaning of community.

**Ridhi Kashyap** is an associate professor of social demography and fellow of Nuffield College at the University of Oxford. Her research spans different topics in demography and sociology, including questions linked to mortality and health, gender inequality, partnerships and family, and migration and ethnicity. As coleader of the strand on digital and computational demography at the Leverhulme Centre for Demographic Science at Oxford, she is working to advance the use and applications of digital and computational innovations, in terms of both new types of data sources (e.g., digital trace data) and methods (e.g., agent-based models) in social and demographic research. She is also substantively interested in analyzing the impacts of digital technologies for demographic outcomes linked to population health, human development and gender inequality drawing on different types of quantitative data sources.