Societal Pressures, Safety, and Online Labeling: Investigating LGBTQ Self-Identification in an Online Space

Irissa Cisternino, Stony Brook University [Department of Sociology]

Jason Jones, Stony Brook University [Department of Sociology]

Abstract

When and how to come out are difficult choices. In this research, we examine one form of disclosure: the inclusion of an LGBTQ keyword in one's online social media profile. We construct daily time series of the prevalence of American Twitter users whose self-descriptions contain LGBTQ keywords. Further, we construct daily time series of inferred Add and Delete events - i.e. we make best estimates for how many users per day make an edit to include a previously absent word or remove a word previously present. These we compare to relevant annual and one-time events such as LGBTQ Pride Month and the date of the Pulse Nightclub shooting. We confirm two pre-registered hypotheses and explore several others.

Introduction

The use of the internet, and of social media sites specifically, has become a staple of social life in the 21st century. According to recent estimates, the number of adults who use the internet in the United States has increased from 52% in 2000 to 90% in 2018, with even higher percentages among younger, more educated, and more affluent adults (Pew Research Center, 2019). The use of social media sites has become especially popular; 69% of US adults are users of Facebook, 37% use Instagram, and 22% use Twitter (Perrin & Anderson, 2019). Social media platforms allow users to connect with each other, share information about their lives or topics they care about, and present information about themselves through the creation of an online profile (Ellison, Steinfield & Lampe, 2007). It is this third function of social media that we explore in this paper.

Previous research has demonstrated the ability of the internet and social media to function as a tool for identity reconstruction (Magnuson & Dundes, 2008; Huang, Kumar, & Hu, 2018). Though research has shown a relative similarity between the online presentation of identity and the offline (Back et al, 2010), other research has suggested that the construction of identity online is strategic, with people attempting to conform to an ideal (Manago, et. al, 2008). Even online, individuals are often not willing to disclose stigmatized identities or experiences unless there is some expectation of anonymity (Pavalanathan & Choudhury, 2015; Andalibi et al, 2018). In the context of online dating, strategic self-presentation is the name of the game, its *sine qua non*. Individuals are careful about what aspects of themselves they show online in the hopes of attracting a more desirable partner (Bruch & Newman, 2018).

In the case of people who identify as LGBTQ, an identity that is still heavily stigmatized (eg. Friedman et al, 2014), the internet functions as an important tool in the development of identity. For LGBTQ people in particular, the internet functions as a prominent learning environment (Fox & Ralston, 2016), and space in which to seek community and validation (Miller, 2017). Research has shown that concealing one's LGBTQ identity can lead to a range of

negative outcomes, including psychological distress and suicidal tendencies (Morris, Waldo & Rothblum, 2001), and thus the ability for LGBTQ individuals to safely disclose their queer identity online is an important one. LGBTQ people are increasingly using the internet as a tool to come out (Chester et al, 2016), and we seek to understand the factors that may influence these disclosures.

Anonymity influences the decision to disclose, as discussed above, but disclosures also occur on social media in which one's identity is public. Andalibi and Forte (2018) describe a system of network-level reciprocal disclosures, in which a disclosure by an individual increases the chances of disclosure of those adjacent to them in their social network. The use of social media to come out is becoming increasingly common (State & Wernerfelt, 2015), and we focus our analysis on Twitter as a site of LGBTQ disclosure. The public nature of Twitter (anyone can view a Twitter profile unless a user specifically opts to list their account as "private") makes this an ideal site for studying what influences the decisions to come out in a public context, rather than just to close friends or family.

Organized, annual events exist which are meant to promote tolerance, inclusion, and visibility of LGBTQ identities. Among these are Transgender Day of Visibility (Human Rights Campaign, 2019), National Coming Out Day (Asmelash, 2019), and other visibility days for various LGBTQ identities. Previous work has examined the impact of particular pride festivals on offline self expression (Kaygalak-Celebi et al, 2019; Browne, 2007), but we know less about how digital expression is affected. Though the goals of these inclusion events are positive, we currently cannot quantify their impact on the visibility of various LGBTQ identities, or on individual decisions to disclose queerness beyond a limited context. Previous work has demonstrated that a pride festival can lead to a brief surge in temporary online mentions (hashtags in tweets) (Duguay, 2016), but little has been documented about how societal events affect more permanent disclosure decisions (such as deciding to include one's queerness in a public online profile). In our data, the median Twitter user updates their profile biography once per year. Thus, we conceptualize the Twitter biography as a mostly static, but occasionally updated measure of identity.

In contrast to events meant to promote positivity among LGBTQ people, other events may increase the desire to keep one's identity hidden. Prominent among these are violent events targeting the LGBTQ community, such as the 2016 shooting at the Pulse nightclub in Orlando, Florida (Zambelich, 2016). In addition, legal decisions related to LGBTQ rights may also impact individual disclosure decisions. For example, one might expect legal victories for the LGBTQ community - such as the Supreme Court ruling on same-sex marriage which took place in 2015 (Liptak, 2015) - to increase the average LGBTQ user's willingness to publicly divulge their identity.

Though previous work has studied the motivations for disclosures of stigmatized experiences online (Andalibi & Forte, 2018), the extent to which specific societal events influence disclosure of a stigmatized identity has not been fully examined. In this paper, we attempt to map the extent to which societal events related to the LGBTQ community influence the use of LGBTQ identifiers online, on the social media platform Twitter. We examine the profile biographies of Twitter users in order to (1) understand whether the use of LGBTQ identifiers is becoming more common on the platform, and (2) whether users make the decision

to strategically edit their profile to include or omit LGBTQ keywords in response to societal events. We construct daily time series of the prevalence of LGBTQ keywords in the Twitter biographies of American users in order to answer these questions.

Theoretical Perspectives

One of the prominent sociological theorists on identity is Erving Goffman. In his influential work *The Presentation of Self in Everyday Life*, (1956) Goffman argued that individuals present an idealized image of themselves, known as a front, for the audiences that observe them. The concept of impression management entails projecting certain aspects of our emotions, identities, or experiences and hiding others from view. Goffman views social actors as akin to performers, and the act of projecting a socially desirable self presentation as beneficial. In his view, actors make an effort to hide aspects of identity that do not conform to societal norms or expectations.

This perspective is particularly useful when thinking about the construction of digital identities. The ability for social media users to pick and choose which aspects of their identity they want to display online and which to leave out means that impression management and strategic decisions about what to disclose and when play a key role in digital identity construction. In addition, the potential for a large, public audience for one's social media profile makes the decision of whether or not to disclose a stigmatized identity a difficult one. The presentation of an ideal self online can be beneficial in some contexts (Bruch & Newman, 2018). Thus, thinking about disclosure decisions as a component of strategic identity formation on the internet is a useful framework.

Scholars studying sexualities have long emphasized the importance of "coming out," or disclosing one's queer identity to others. Recently, sexuality scholars have begun to re-examine this concept, and the factors that influence one's decisions to disclose a queer identity. Jason Orne (2011) puts forth the concept of "strategic outness," and argues that, rather than a one-time process, coming out is a continuous and complex process of identity management. LGBTQ people are constantly managing and re-negotiating coming out, as they are forced to come out multiple times to new people they meet, or manage parts of their queer identity in new or different contexts. This perspective suggests that LGBTQ people are continually thinking about strategies to express or conceal their identity at different times, and that they employ a number of different methods to do so. Strategic outness is an extension of the identity management perspective, and these theoretical concepts play a central role in this project.

One strategic factor that may play a role in online disclosures is the potential offline implications. Larger societal factors may make individuals more or less likely to disclose something publicly in an online forum. One popular framework dealing with the impact of societal events on free speech and online expression, deals with so-called "chilling effects." This term can be particularly useful in the context of this work. The phrase has long been a subject of debate in the legal literature, especially as it pertains to First Amendment law (Schauer, 1978). In recent years, though, the term has taken on a life outside of the legal context, and we draw on this framework to inform our analysis of motivations behind online disclosures.

In a legal context, the term “chilling effect” refers to the possibility that the existence of a legal statute may inhibit constitutionally protected free speech in some way—in other words, “chilling” free expression. The term has featured in a variety of court cases, but has long remained a subject of debate among scholars (Penney, 2016). More generally, the concept of a chilling effect refers to the existence of some controversy, surveillance, or legislation that may potentially inhibit behaviors of the public based on a fear of sanction or punishment.

Outside of the legal system, scholars have examined chilling effects in a variety of different contexts. Some have studied a type of chilling effect that exists within familial relationships, in which the existence of coercive power or the perception of aggression by family members leads people to conceal secrets for fear of negative outcomes (Afifi and Olsen, 2005; Afifi, Olsen and Armstrong, 2005). Other research has found that chilling effects can extend to a scientific context. Scientists have been shown to engage in self-censorship when their research engages with political controversy, demonstrating a type of chilling effect (Kempner, 2008).

This paper studies a type of online chilling effect, in which we hypothesize that LGBTQ people will be less likely to disclose their identity on a public social media platform after a harmful event. Empirically, we know that people choose to limit their online behaviors due to fear of retaliation or harm. Previous work monitored Wikipedia use before and after the reveal of widespread NSA surveillance, and found that users were less likely to access information potentially seen as controversial in the wake of government surveillance (Penney, 2016). Other work on online chilling effects has shown that users may even limit their behavior offline in order to avoid projecting a problematic online image (Marder et al, 2016).

Previous work has demonstrated that users limit stigmatized disclosures online for fear of harm. Individuals are more likely to disclose a stigmatized experience or identity if they are doing so anonymously (Andalibi and Forte, 2018) or if they receive support from those within their social network (Haimson et al, 2015). Given this, it is likely that the existence of certain societal “focusing events” that are harmful to LGBTQ people may lead to a type of chilling effect in which people are less likely to publicly disclose their identity. In addition, we might expect a sort of reverse chilling effect in which societal events designed to promote the visibility of LGBTQ people and identities may increase the rate at which these identities are disclosed.

Hypotheses

In order to study the effect of offline events on individual disclosure decisions, we group these events into 3 categories: Inclusion & Visibility Events, Legal Events, and Anti-LGBTQ Violent Events. We developed the following hypotheses, and pre-registered them before conducting any analysis:

H1: The prevalence of accounts whose profile description contains at least one LGBTQ keyword will be increasing over time.

According to a 2017 estimate by Gallup, the number of adults who identify as LGBTQ rose from 8.3 million in 2012 to 10.05 million in 2016 (4.1% of the population) (Gates, 2019). Because the proportion of LGBTQ identifying adults is increasing, we expect to see the proportion of people using these labels on social media to increase as well.

H2: An *Add event* occurs when a user edits their profile biography from text not containing an LGBTQ keyword to text containing a keyword. A *Delete event* is the opposite - the user edits text containing a keyword to remove the keyword. We will observe greater-than-baseline Add events and lower-than baseline Delete events during Inclusion and Visibility Events. We expect the same on June 26, 2015 – the date the US Supreme Court decision regarding same-sex marriage was announced. We expect that the magnitude of the Add effect will be greater than the magnitude of the Delete effect.

Inclusion and visibility events are meant to promote positive awareness for LGBTQ people and to give them increased support. On days meant to promote visibility of LGBTQ people, such as National Coming Out day, it is reasonable to expect increased affirmation and disclosures of LGBTQ identities online. Previous research on social media has argued in favor of a system of “network-level reciprocal disclosures,” in which a disclosure by one person on a social networking site leads to increased disclosures by others. We hypothesize that events intended to promote visibility will have their intended effect.

We predict we will see the same pattern of increased Add events and decreased Delete events following the announcement of the US Supreme Court decision in *Obergefell v. Hodges* on June 26, 2015. This decision affirmed the right to marry for same-sex couples and was widely viewed as a landmark in reducing the stigmatization of LGBTQ identities (Pearson et al, 2015; DeVogue and Diamond, 2015). While the intention of the court was not necessarily inclusion and visibility, we believe individuals would feel less fear and stigma being publicly out after the decision.

H3: We will observe lower-than-baseline Add events and greater-than baseline Delete events in response to Anti-LGBTQ Violent Events. Specifically, we expect this pattern to follow June 12, 2016, the date of the Pulse Nightclub Shooting. Initially, we expect that the magnitude of the Delete effect will be greater than the magnitude of the Add effect. However, we also predict this date will be followed by a diffuse, delayed, compensatory period of increased Add events.

Inclusion and Visibility events are meant to promote awareness of LGBTQ issues and visibility of LGBTQ people, while Anti-LGBTQ violent events are meant to target the LGBTQ community and make LGBTQ people feel less safe. The fear of being targeted or stigmatized after an Anti-LGBTQ violent event may make one feel less comfortable being publicly out online and may lead people to remove LGBTQ keywords from their online profiles. Because of this stigma, we expect to see greater-than-baseline Delete events, and lower-than-baseline Add events in response to Anti-LGBTQ violent events. Our test case will be the Pulse Nightclub shooting. However, we predict there will be a delayed period of increased Add events in a show of solidarity with and support for the LGBTQ community.

H4: We will observe greater-than-baseline Add events and lower-than baseline Delete events on Visibility Days that are specific to the corresponding keyword. For instance, on Bi Visibility Day, we will expect the keyword “bi” to be affected, but the keyword “lesbian” to NOT be affected.

The existence of subgroup-specific visibility events provides an opportunity to test this hypothesis of selective effect. Essentially, this hypothesis is the same as H2, however, rather

than estimating the effect of LGBTQ events on LGBTQ prevalence, we can estimate the effect of more narrowly targeted messages on specific identifiers.

Methods

Data

The data for this paper were created using Longitudinal Online Profile Sampling (Jones, 2019; Rogers & Jones, 2019). This method makes use of the availability of vast numbers of individual public profiles online. The goal is to unobtrusively observe the self-presentation of identity for millions of people. Because users maintain these profiles by occasionally editing them, it is also possible to infer temporal trends by combining observations over several years of available data.

This manuscript specifically focuses on the Twitter profiles of 3,051,877 American users. Twitter is a microblogging service used by 22% of American adults (Perrin & Anderson, 2019). Using the website or smartphone app, users create and maintain a profile. The profile consists of a picture, name, location and - of interest in this work - a free response text field we call the “bio.” The bio field (what Twitter calls the “user description”) is the user’s response to the prompt: *Describe yourself in 160 characters or less*. Users are free to edit their bio at any time. Each time a user tweets, a snapshot of their profile (including bio) at that moment is published along with the tweet. We use a continuous stream of a random 1% sample of all public tweets to construct time series data for this research.

Users were selected for analysis if they met these criteria:

1. They had chosen an American timezone for their local time.
2. They had chosen English as their interface language.
3. They tweeted at least once in 2015, 2016, 2017 and 2018.

Twitter has become a valuable source of data for the social sciences (Crosset, Tanner & Campana, 2019; Sinnenberg et al, 2017; Simonetto, 2016). Much hand-wringing has been performed over the fact that Twitter users are not a representative sample of the human or American population (e.g. Mellon & Prossner, 2017). This is true. However it is also true that a population-representative sample of self-descriptions for millions of people over multiple years has never been collected and likely never will be. Additionally, most research occurs on non-representative samples. Candidates for President are not a representative sample of Americans, and yet their words and actions are extensively studied by academics. Sophomore psychology students at research universities are not a representative sample and yet nearly all human behavior experimental evidence comes from such samples. The ice around Amundsen-Scott South Pole station in Antarctica is not geographically representative of the whole continent’s ice, and yet studying it is still a valid approach to glaciology. Twitter users - filtered as described above - are a convenience sample, not a random sample. But, we reiterate the point that no other method collects frequent, longitudinal, multi-year data on millions. We would go further and argue that because of its wide usage, Twitter is likely more inclusive a sample than most social science datasets. For current purposes, Twitter is the perfect data source; this study intends to make claims about social media users and the way they choose to present themselves online.

LGBTQ-Related Events

We compiled LGBTQ-related events into three categories: Inclusion & Visibility events, Legal Events, and Anti-LGBTQ violent events. Inclusion & Visibility events are meant to celebrate LGBTQ identities and promote inclusion for LGBTQ-people. Examples include events relevant to all sections of the LGBTQ community, such as National Coming Out Day, and others that are meant to promote visibility of certain identities, such as International Transgender Day of Visibility. Legal events include prominent legislative decisions that adjudicate LGBTQ rights. The legal decision we focus on is the US Supreme Court decision in *Obergefell v. Hodges*, which affirmed the right to marriage for same-sex couples nationally. Lastly, Anti-LGBTQ violent events are acts of violence that target the LGBTQ community and make LGBTQ people feel less safe. The event we examine is the Pulse Nightclub Shooting in 2016, where 45 people were killed at an LGBTQ nightclub.

Measurement and Constructing Time Series

Our primary measure of interest is purposefully simple. It is the presence or absence of an LGBTQ keyword within a bio. The keywords used in this manuscript are listed in Table 1. We search bios using regular expression matching. This means we can be confident our keyword matches are what we intended to search for, and not simple character-to-character literal matches. (See details in the table.)

Keyword	Regular Expression Search Pattern	Matches	Does Not Match (examples)
lesbian	(?i)\blesbian\b	lesbian	lesbians
gay	(?i)\bgay\b	gay	G.A.Y.
bi	(?i)\bbi\b	bi	bicycle, mobile
trans	(?i)\btrans\b	trans	trains, intransitive
queer	(?i)\bqueer\b	queer	queerly
At least one LGBTQ word	(?i)\blesbian\b \bgay\b \bbi\b \btrans\b \bqueer\b	Any bio that includes at least one of the above.	All bios that contain zero matches to LGBTQ keywords.

Table 1. Keywords of interest in this study, match patterns and examples of matching and non-matching text. The expression (?i) indicates the match should ignore case, so Queer and queER both match (?i)\bqueer\b. The expression \b indicates that a word boundary must be found - e.g. whitespace, punctuation or an emoji.

The first time series we construct is prevalence within our sample. Our sample consists of 3,051,877 accounts, and thus that number is the denominator in the prevalence calculation. The number of user accounts that contain a keyword match in the bio is the numerator. We multiply by 10,000 because keywords are rare, and it is much easier for humans to think in whole numbers than small fractions or decimals (Ni & Zhou, 2005). Formally, prevalence is defined as:

$$(\text{Count of matching users}) / (\text{Total user count}) * 10,000$$

It is important to be clear that we are always counting users and never words. A bio that reads “gay gay gay gay gay” counts as *one* user who matches the pattern `\bgay\b`, even though the pattern happens to match five times.

We also measure the rate at which users are adding and deleting keywords. This is a slightly more complicated concept, because it involves two observations per user. However, the process is still straightforward:

1. Detect keyword matches.
2. Place users into one and only one category (out of four possibilities).
3. Count users per category.

Table 2 names and illustrates the four possible outcomes given a keyword and two bio observations separated in time.

Event Name	Time 1	Time 2
Add	Bio does not contain keyword	Bio contains keyword
Delete	Bio contains keyword	Bio does not contain keyword
Keep	Bio contains keyword	Bio contains keyword
(Absent)	Bio does not contain keyword	Bio does not contain keyword

Table 2. For any two observations of a bio, one of four possible events occurs for a keyword. This table names and explains each possibility.

An *Add* event occurs when a user who did not previously use a word in their bio subsequently does. A *Delete* event occurs when a user who did previously use a word in their bio subsequently does not. A *Keep* event occurs when the word appears at both points in time. An *(Absent)* event occurs when neither observation of the bio includes the keyword. Upon reflection, it should become clear that (Absent) events will be most numerous (but least interesting) simply because most users do not use most words in their bio. They could not if they tried - bios are limited to 160 characters. Add, Delete and Keep events are relatively rare, but among millions of users, even slow event days produce dozens of events.

We construct time series at a daily level of temporal resolution. To do this requires inference, because we only observe profiles when users tweet. We cannot control how often users tweet. Most users do not tweet every day. However, we can use the available data to infer a best guess for the approximate date of change events. We do this in the simplest way possible - by interpolating bios forward in time from the chronologically first observation, backward in time from the second observation and placing the change event on the chronologically median date. For example, if a bio was observed to contain the word “gay” on February 1, and upon observation the bio did not contain the word “gay” on February 9, a Delete event would be recorded on February 5.

Our dataset ends on May 23, 2018 due to a technical change implemented by Twitter in reaction to the The EU General Data Protection Regulation (GDPR). Specifically, on that date, the Twitter API ceased to provide users' timezone. This makes our filtering process unworkable beyond that date, because we specifically filter users to those who have chosen US timezones for their profile. There are alternative methods to filter users by location. However, to keep our method of filtering simple and consistent with the method when we pre-registered our hypotheses, we restrict our analysis to tweets observed from February 15, 2015 through May 23, 2018.

Results

Figure 1 depicts the temporal trend at daily resolution for the prevalence of user bios that contain at least one LGBTQ keyword. A consistent positive trend is clear. From the beginning of available data in February 2015 to its end in May 2018, the prevalence climbs from 34 to 50 - an increase of 47%. It is important to remember these are the same 3,051,877 users observed over this timespan, so the change is best interpreted as individual, longitudinal change. It is not a cohort or replacement effect.

Figure 1. The prevalence of user bios (per 10,000 accounts) that contain any of the words *lesbian*, *gay*, *bi*, *trans* or *queer* within a multi-year, longitudinal sample of about three million users.

Figure 2 depicts the temporal trend at daily resolution for the prevalence of user bios that specifically contain a particular keyword. For all keywords, the slope of the line is positive. However, the rate of increase differs between keywords. The prevalence of *lesbian* increases modestly from 4 to 5 per 10,000, while the prevalence of *queer* doubles from 6 to 12 and *trans* more than doubles from 2 to 5.

Figure 2. Prevalence of matching bios per each keyword in the period Feb 2015 - May 2018 in a longitudinal sample of 3,051,877 users.

The above time series result from the accumulation of Add and Delete events. Figure 3 depicts the inferred daily count of such events for matches on at least one LGBTQ keyword. The dynamics are interesting. For one, there is autocorrelation - the count for a particular day appears similar to the previous day. The Add series is correlated with itself, lagged one day at value of $r = 0.47$, $p < 0.001$ and at seven days at a value of $r = 0.27$, $p < 0.001$. Second, the count of Add and Delete events also are correlated $r = 0.32$, $p < 0.001$. The broad trend of increasing prevalence is apparent through the observation that the Add count is greater than the Delete count on most days. Pride Month (June) is visually apparent through the separation of the red and blue lines - indicating many more Adds than Deletes in this period. It is also visually apparent that there is a high volume of activity in early November 2016. We attribute this to a reaction to the surprise election of Donald Trump as President. We did not anticipate this in our pre-registered hypotheses, and instead explore this activity later in this manuscript in a set of ad-hoc analyses

Figure 3. Inferred daily incidence of Add and Delete events for bios containing any LGBTQ keyword. An Add day is the interpolated day on which a bio that previously contained no keywords becomes a bio that does contain at least one. A Delete day is the interpolated day on which a bio that previously contained at least one keyword becomes a bio that does not contain any.

A figure depicting the inferred daily count of Add and Delete events per keyword is included in supplemental materials. These series exhibited the same autocorrelation as those for any-keyword matching bios. At full zoom, one can again see the positive effect of Pride Month for all keywords with the exception of *lesbian*. Overall, the series show the keywords have about the same number of “up days” (more Adds than Deletes) as down days. The keyword *queer* stands out a bit from this pattern with the highest proportion of up days - 477/731 or 65%. We will use these series to test our hypotheses.

Hypothesis Tests

Here we repeat each of our pre-registered hypotheses (Jones & Cisternino, 2019), and present the results of statistical tests corresponding to each.

H1: The prevalence of accounts whose profile description contains at least one LGBTQ keyword will be increasing over time.

The slope of the line in Figure 1 is clearly positive. A linear regression confirms this with $p < 0.001$. H1 is confirmed.

H2: We will observe greater-than-baseline Add events and lower-than baseline Delete events during Inclusion and Visibility Events. We expect the same on June 26, 2015 – the date the US Supreme Court decision regarding same-sex marriage was announced. We expect that the magnitude of the Add effect will be greater than the magnitude of the Delete effect.

To begin, we must define baseline, greater-than-baseline and lower-than baseline. In words, we consider baseline the typical value of a series. For daily values, we define *baseline* as the mean value for the year in which the day occurs. We scale series to Z-scores, so it is easier to compare across keywords and time on a familiar scale. Thus, a score of zero represents the average day (literally no deviation from the mean) and is exactly equal to baseline. Greater-than-baseline values are positive. A score of +1.96 represents a value nearly two standard deviations above the mean. This is a useful criterion for two reasons. First, it is the critical value for conventional statistical significance ($p = 0.05$ for a two-tailed Z test). Second, exceeding this threshold tells us we are looking at a rare type of day - only a few days within a 365 day year should have a value of 1.96 or higher. We use daily Z-scores to test our hypotheses concerning individual dates.

Figure 4. Boxplots summarizing the distribution of Z-scores for Add and Delete events for days in June (labeled Pride) and days not in June (labeled Not Pride).

We focus on Pride Month (June), where the data provides the clearest evidence. Figure 4 summarizes the distribution of daily Z-scores for Add and Delete events for days in June compared to days not in June. As expected, most of the data lies close to baseline (Z-score of zero) with rare outliers above and below two standard deviations away from the

mean. Comparing the boxplots on the right and left reveals the Add distribution shifts upward and the Delete distribution shifts downward during June, as predicted. The notches on each box represent confidence intervals around the median value calculated per the method described in McGill, Tukey & Larsen (1978). These provide evidence that the number of Add events is shifted upward, above zero on June days, and Delete events are shifted downward, below zero on June days. The estimates for the medians on non-June days are tightly centered near zero. This evidence supports H2, and we predict the pattern of greater-than-baseline Add events and lower-than-baseline Delete events will hold on future June days.

We can employ binomial tests to test specific propositions about the unusualness of the Add and Delete event counts within June. Here, the evidence is stronger for Adds than Deletes. Of the 90 June days in our sample, 10 scaled Add values are above 1.96. That is more than the 2 days expected ($p < 0.001$). 58 days in June have scaled Add values above zero. That is more than the 45 expected ($p < 0.01$). Both tests are not significant for scaled Delete values: 2 days below -1.96 and 50 days below zero.

Another way to examine this trend in aggregate is to examine the prevalence of LGBTQ users over time. In Figure 1, the slope of increase in prevalence appears to accelerate temporarily each June. We employ segmented (also known as broken-stick) regression to estimate the slope of increase in disjoint date intervals. Figure 5 presents estimates of the slope for each interval. All slopes are positive; June 2017 and June 2016 are the greatest in magnitude. June 2015 stands out from the surrounding date intervals. In the supplemental material we include a figure illustrating the predictions of this model overlaid on the data.

Figure 5. Results of segmented regression predicting prevalence of users with LGBTQ bios. Each point is the estimated slope (per day) in the named date interval. Lines represent $\pm$ one standard error.

Now we turn to an individual date: June 26, 2015. This was the day of the announcement of the US Supreme Court ruling affirming same-sex marriage rights. Contrary to our expectation, there is nothing extraordinary in the series on June 26, 2015. We expected to see a pronounced anomaly - for example, a several-fold increase in Add events. However, instead we see a day of less activity. For June 26, 2015, specifically we see that the Z-score for the Add event count is -0.95. For the Delete count, the Z-score is -1.69. That does place June 26, 2015 as the eighth *least* active day for Delete events in 2015. In the larger picture, however, (and inconsistent with our predictions) on June 26, 2015 users were both adding and deleting LGBTQ keywords at a modestly less-than-baseline rate. H2 for this specific date was not confirmed.

A peak that we did not anticipate appears in November 2016. The month produced the second-most net change in prevalence in that year (after June). November 10 produced the highest number of Add events of any day in 2016, but was accompanied by a high number of Delete events as well. As an ad-hoc hypothesis, we posit that the surprise election of Donald Trump to the Presidency caused many American users of Twitter to amend their biographies to either highlight or remove their LGBTQ identity.

H3: We will observe lower-than-baseline Add events and greater-than baseline Delete events in response to Anti-LGBTQ Violent Events. Specifically, we expect this pattern to follow June 12, 2016, the date of the Pulse Nightclub Shooting. Initially, we expect that the magnitude of the Delete effect will be greater than the magnitude of the Add effect. However, we also predict this date will be followed by a diffuse, delayed, compensatory period of increased Add events.

Figure 6 presents the daily scaled incidence of Add and Delete events for the keyword *gay* for a six week period centered on June 12, 2016. June 12, 2016 is marked by a vertical black dot-and-dashed line. We predicted changes in the series would be obvious at the Pulse event date. We predicted an immediate, sharp decline in Add events to below baseline that would rise slowly back to above baseline. We predicted an immediate, sharp increase in Delete events to above baseline that would quickly fall back to baseline. Those patterns do not appear in the data. One might argue instead that the date of the Pulse shooting splits the series into a period of lower activity before and higher activity afterward.

Figure 6. Inferred daily scaled frequencies of Add and Delete events for the keyword *gay* in the six week period surrounding June 12, 2016. Red and blue points represent the inferred frequencies of Add and Delete events respectively. The red and blue dashed lines represent a LOESS smoothing of the series.

H4: We will observe greater-than-baseline Add events and lower-than baseline Delete events on Visibility Days that are specific to the corresponding keyword. For instance, on Bi Visibility Day, we will expect the keyword “bi” to be affected, but the keyword “lesbian” to NOT be affected.

Table 3 contains the average Z-score and average rank (out of 365) for Add events for each combination of keyword and visibility day. For example, the top-left cell contains the mean Z-score and rank for Add events of the keyword *lesbian* on Lesbian Visibility Day (April 26). If the top-left to bottom-right diagonal contained high positive Z-scores and low ranks, this would be support for H4. If instead, a particular column showed extreme values, it would be evidence that a particular visibility day drove Add events across all keywords. If a row showed extreme values, this would indicate that a particular keyword was responsive to multiple visibility days. Three years of values are averaged: 2015, 2016 and 2017. 2018 is omitted from this analysis, because it is a partial year of data.

	Lesbian Visibility Day Apr 26		National Coming Out Day Oct 11		Celebrate Bisexuality Day Oct 23		International Transgender Day of Visibility Mar 31		International Non-Binary People's Day Jul 14	
<i>lesbian</i>	0.19	146th	-0.84	234th	-0.73	258th	1.12	57th	0.85	71st
<i>gay</i>	-0.25	205th	-0.75	265th	-0.97	289th	0.95	97th	-0.28	202nd
<i>bi</i>	0.41	139th	0.14	152nd	-0.47	222nd	-0.21	181st	0.68	130th
<i>trans</i>	-0.30	206th	0.58	106th	-1.16	303rd	1.03	99th	0.32	160th
<i>queer</i>	0.32	124th	-0.53	228th	-0.99	294th	-0.24	198th	0.63	90th
<i>At least one of any LGBTQ keyword</i>	-0.09	194th	-0.71	267th	-1.43	329th	0.84	94th	0.18	141st

Table 3. Mean Z-score and rank for the frequency of Add events for each combination of keyword and visibility day.

The data in Table 3 do not show a pattern in support of H4. The top-left to bottom-right diagonal does not contain extreme values indicating strong Add activity specific to keywords corresponding to particular days. Overall, visibility days do not stand out as days of high activity within the year. The final two columns for International Transgender Day of Visibility and International Non-Binary People's Day display more consistently positive Z-scores and lower ranks than the other columns.

One could wonder if pinpointing activity to exact dates is asking too much of the data. We do not have daily observations of each accounts' bio. Instead we *infer* daily bios from the infrequent random sampling provided by the available data. A change that happened on a particular date is more likely to be indicated in our data within a range *near* that date than on the exact particular date. To account for this, in Figure 7 we illustrate both Add and Delete activity for the 29 day range surrounding each visibility day. We average the scaled Add and Delete Z-scores for the keyword matching the visibility day. The gold line illustrates the prediction of H4, viz. specific keyword Adds should peak at corresponding visibility days while Deletes show a negative peak.

Figure 7. Mean daily Add and Delete Z-scores in the time surrounding visibility days. On the x-axis, zero marks a visibility day, -7 marks one week before and 7 marks one week after. Only corresponding matches of keyword + visibility day are included (e.g. *trans* on International Transgender Day of Visibility). Gold lines illustrate the prediction, black dots represent the data, and dashed blue lines are a smoothed LOESS approximation.

These data do not reveal a pattern in support of H4. Instead the rate of Add and Delete events fluctuates in a tight range around baseline in the time period surrounding visibility days. There is simply no evidence that visibility days affect individual disclosure decisions in the manner we predicted.

Note that these are results for *changes in biography text* in a longitudinal sample. We are not counting which accounts are active (tweeting) around these dates; we are counting how many out of our total sample of about 3 million users are making relevant changes to the content of their bios. We plan to investigate elsewhere the cross-sectional daily effect - e.g. how much more active on Twitter are profiles whose bios contain the word *bi* on Bi Visibility Day compared to other days and other bios?

Discussion

This work demonstrates the utility of online public profiles to the study of identity. We observe the aggregate effect of millions of individual disclosure decisions at high temporal resolution. The presence or absence of keywords is both an interesting measure in its own right and a straightforward operational definition of the collective willingness to be publicly out. The volume of data allows for faceted exploration while still maintaining statistical power. The data source is a continuous stream. This means researchers can make and test actual predictions - that is, predicting a quantity before that quantity is measured. In the post-replication crisis era, as social science faces its history of p-hacking, HARKing and confirmation bias, the ability to make and test true predictions is a necessity.

Two main results jump off the page as we examine the data. First is the consistent increase in the prevalence of LGBTQ terms within users' bios over the three-plus years in the data. Our hypothesis H1, pre-registered before we began analysis, was confirmed. Compare this increase to the survey data cited previously. In the Gallup survey, the estimated percentage of American adults identifying as LGBT increased from 3.5% to 4.1% - a relative increase of 32% - over 4 years. The relative increase in the prevalence as measured here (52%) is slightly higher, but on the same order. The difference could be due to many reasons, including that people have felt comfortable disclosing on an anonymous survey that they identify LGBTQ for longer than they have felt comfortable publicly disclosing online and that the bio prevalence is "catching up."

It is also interesting to consider that the prevalence in Twitter profiles (4 in 10,000) is two orders of magnitude lower than the estimated prevalence in the population (4 in 100). One reason this is not surprising is the zero-sum game users face apportioning space in their 160 character limited self-description. Users are not compelled to list every aspect of their identity within their bio and couldn't if they tried. We may compare the prevalence to other categories for some context. In other work, we have estimated the prevalence of bios containing political or religious keywords in the range of 50-90 per 10,000. Words related to sports and creative pursuits are more popular still - around 300 per 10,000. Users are much more likely to define themselves with these labels in contrast to LGBTQ identifiers. Further research should continue to compare the rate of inclusion of LGBTQ keywords relative to other aspects of identity. Doing so will allow real-time, precise measurement of social media users' strategic choices regarding public self-presentation.

The second main result is the large and positive net change in the prevalence of LGBTQ identifiers within June. Rates of Add events climb going into June and remain high, while rates of Delete events drop or remain steady. We attribute this effect to LGBTQ Pride Month. In June, LGBTQ individuals receive messaging about taking pride in their identity and social reinforcement for doing so. In this environment, we posit it is easier to feel comfortable disclosing an LGBTQ identifier publicly online and one may feel some relief from other pressures to remove that identifier. As we observe millions of users maintaining their online presentation of self, these small shifts in relative weight tip the scales clearly in one direction.

While we made pre-registered hypotheses regarding the effects of individual dates, we did not find such clearly visible evidence as we did for our other hypotheses. Future work over a longer time frame and more sophisticated time series analysis will give a better indication if these effects are subtle, small or simply null. At present, we conclude there is no evidence for our hypotheses concerning single-day, single-event effects. A new method in which we compare every user's profile biography on a day-by-day basis regardless of whether they tweet or not will also provide better measurement. However, this method will be limited to smaller, predefined user samples.

In the specific case of the June 12, 2016 Pulse nightclub shooting, we did not find the evidence we expected of a "chilling effect." Instead, June 2016 stands out in the data as the period in which LGBTQ prevalence increased at its fastest rate. (See Figures 1 and 5.) We see this as evidence that the "reverse" chilling effect of Inclusion & Visibility (specific to Pride Month) overwhelmed that solitary event.

Limitations

As we mention in the Methods section, the Twitter sample is a convenience sample, not a population-representative random sample. Population-representative random samples with similar data are not likely to be forthcoming, however, and would have to be very large and longitudinal to detect temporal trends such as we have observed.

Some researchers worry that most Twitter accounts are bots. This is simply not the case. Twitter (the company) has an incentive to detect bots and resources and data to do so. The number of bots detected is on the scale of thousands out of the many millions of accounts. Regardless, these bots would have to be altering their bios to include more LGBTQ terms (especially in June) to produce false positives due to bot activity. We are not aware of any bot that exhibits this behavior. The most likely effect of bots in our dataset is small amounts of random measurement error.

One might worry that organizations (such as a neighborhood LGBTQ center) would have a Twitter account and be mistakenly included in this sample. With over three million accounts, we have not personally inspected each. It is likely there are some organization-type accounts within the sample. A moment's reflection will reveal that it is likely there are vastly more LGBTQ individuals in the United States than there are LGBTQ organizations. The number of accounts representing individuals likely severely outnumbers the number representing anything other than an individual, and the most likely scenario is these accounts contribute small amounts of random measurement error.

Finally, whatever one's personal inclination is toward Twitter and its users, it is a highly visible and influential form of media. This sample of active accounts across several years is visible to many human users at many moments. Whatever the ultimate cause of the temporal effects we observe here, it is still empirically true that the prevalence of user accounts with profile descriptions including LGBTQ keywords is increasing over time, and the rate of increase appears to respond to some events (Pride) and not others (Visibility Days).

Conclusion

Longitudinal Online Profile Sampling (Jones, 2019; Rogers & Jones, 2019) is a useful tool to study self-presentation online. It allows precise measurement at high resolution, which is not true for most social science methods. It lends itself to iterative, open science in which researchers make genuine predictions (about the future) and test them. In this application of the method, we have demonstrated that American users of Twitter have become increasingly likely to include an LGBTQ keyword within their bio and the rate of increase is particularly high within the month of June.

References

- Afifi, Tamara D., and Loreen Olson. 2005. "The Chilling Effect in Families and the Pressure to Conceal Secrets." *Communication Monographs* 72(2):192–216.
- Afifi, Amara D., Loreen N. Olson, and Christine Armstrong. 2005. "The Chilling Effect and Family Secrets.: Examining the Role of Self Protection, Other Protection, and Communication Efficacy." *Human Communication Research* 31(4):564–98.
- Andalibi, Nazanin, Oliver L. Haimson, Munmun De Choudhury, and Andrea Forte. 2018. "Social Support, Reciprocity, and Anonymity in Responses to Sexual Abuse Disclosures on Social Media." *ACM Transactions on Computer-Human Interaction (TOCHI)* 25(5):28.
- Andalibi, Nazanin and Andrea Forte. 2018. "Announcing Pregnancy Loss on Facebook: A Decision-Making Framework for Stigmatized Disclosures on Identified Social Network Sites." Pp. 158:1–158:14 in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI '18*. New York, NY, USA: ACM.
- Asmelash, Leah. 2019. "National Coming Out Day: What It Is, and What People Are Saying about It." *CNN*, October 11
- Browne, Kath. 2007. "A Party with Politics? (Re)Making LGBTQ Pride Spaces in Dublin and Brighton." *¿Una Fiesta Con La Política? (Re)Haciendo Espacios Del Orgullo LGBTQ En Dublín y Brighton*. 8(1):63–87.
- Back, M. D., Stopfer, J. M., Vazire, S., Gaddis, S., Schmukle, S. C., Egloff, B., & Gosling, S. D. (2010). Facebook Profiles Reflect Actual Personality, Not Self-Idealization. *Psychological Science*, 21(3), 372–374. <https://doi.org/10.1177/0956797609360756>
- Bruch, Elizabeth E. and M. E. J. Newman. 2018. "Aspirational Pursuit of Mates in Online Dating Markets." *Science Advances* 4(8):eaap9815.
- Crosset, Valentine, Samuel Tanner, and Aurélie Campana. 2019. "Researching Far Right Groups on Twitter: Methodological Challenges 2.0." *New Media & Society* 21(4):939–61.
- De Vogue, Ariane, and Jeremy Diamond. 2015. "Supreme Court Rules States Must Allow Same-Sex Marriage - CNNPolitics." *CNN*. Retrieved November 19, 2020 (<https://www.cnn.com/2015/06/26/politics/supreme-court-same-sex-marriage-ruling/index.html>).
- Duguay, Stefanie. 2016. "'Legit Can't Wait for #Toronto #WorldPride!': Investigating the Twitter Public of a Large-Scale LGBTQ Festival." 25.
- Friedman, M. Reuel, Brian Dodge, Vanessa Schick, Debby Herbenick, Randolph Hubach, Jessamyn Bowling, Gabriel Goncalves, Sarah Krier, and Michael Reece. 2014. "FROM BIAS TO BISEXUAL HEALTH DISPARITIES: ATTITUDES TOWARD BISEXUAL MEN AND WOMEN IN THE UNITED STATES." *LGBT Health* 1(4):309–18.

- Gates, Gary J. 2019. "In U.S., More Adults Identifying as LGBT." Gallup.com. Retrieved October 15, 2019 (<https://news.gallup.com/poll/201731/lgbt-identification-rises.aspx>).
- Haimson, Oliver L., Jed R. Brubaker, Lynn Dombrowski, and Gillian R. Hayes. 2015. "Disclosure, Stress, and Support During Gender Transition on Facebook." Pp. 1176–1190 in Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, CSCW '15. New York, NY, USA: ACM.
- Huang, Jiao, Sameer Kumar, and Chuan Hu. 2018. "Gender Differences in Motivations for Identity Reconstruction on Social Network Sites." *International Journal of Human–Computer Interaction* 34(7):591–602.
- Human Rights Campaign, 2019 "International Transgender Day of Visibility." *Human Rights Campaign*. Retrieved December 19, 2019 (<https://www.hrc.org/resources/international-transgender-day-of-visibility/>).
- Jones, J. J. (2019, June 18). *LOPS - Longitudinal Online Profile Sampling*. OSF. <https://osf.io/rwxec/>
- Jones, J. J., & Cisternino, I. (2019, October 16). Do LGBTQ-related Events Drive Individual Online Disclosure Decisions?. Retrieved from <https://osf.io/xbqu8>
- Kaygalak-Celebi, Sonay, Sehriban Kaya, Emir Ozeren, and Ebru Gunlu-Kucukaltan. 2019. "Pride Festivals as a Space of Self-Expression: Tourism, Body and Place." *Journal of Organizational Change Management* 33(3):545–66.
- Kempner, Joanna. 2008. "The Chilling Effect: How Do Researchers React to Controversy?" edited by P. Singer. *PLoS Medicine* 5(11):e222.
- Liptak, Adam. 2015. "Supreme Court Ruling Makes Same-Sex Marriage a Right Nationwide." *The New York Times*, June 26.
- Manago, A. M., Graham, M. B., Greenfield, P. M., & Salimkhan, G. (2008). Self-presentation and gender on MySpace. *Journal of Applied Developmental Psychology*, 29(6), 446-458.
- Magnuson, Melissa Joy and Lauren Dundes. 2008. "Gender Differences in 'Social Portraits' Reflected in MySpace Profiles." *CyberPsychology & Behavior* 11(2):239–41.
- Marder, Ben, Adam Joinson, Avi Shankar, and David Houghton. 2016. "The Extended 'Chilling' Effect of Facebook: The Cold Reality of Ubiquitous Social Networking." *Computers in Human Behavior* 60:582–92.
- McGill, R., Tukey, J. W. and Larsen, W. A. (1978) Variations of box plots. *The American Statistician* 32, 12-16.
- Mellon, Jonathan and Christopher Prosser. 2017. "Twitter and Facebook Are Not Representative of the General Population: Political Attitudes and Demographics of British Social Media Users." *Research & Politics* 4(3):2053168017720008.

- Miller, Ryan A. 2017. “‘My Voice Is Definitely Strongest in Online Communities’: Students Using Social Media for Queer and Disability Identity-Making.” *Journal of College Student Development* 58(4):509–25.
- Morris, J. F., C. R. Waldo, and E. D. Rothblum. 2001. “A Model of Predictors and Outcomes of Outness among Lesbian and Bisexual Women.” *The American Journal of Orthopsychiatry* 71(1):61–71.
- Ni, Yujing & Yong-Di Zhou (2005) Teaching and Learning Fraction and Rational Numbers: The Origins and Implications of Whole Number Bias, *Educational Psychologist*, 40:1, 27-52, DOI: 10.1207/s15326985ep4001_3
- Orne, Jason. 2011. “‘You Will Always Have to “out” Yourself’: Reconsidering Coming out through Strategic Outness.” *Sexualities* 14(6):681–703.
- Pavalanathan, Umashanthi and Munmun De Choudhury. 2015. “Identity Management and Mental Health Discourse in Social Media.” *Proceedings of the ... International World-Wide Web Conference. International WWW Conference 2015(Companion)*:315–21.
- Pearson, Michael, Ray Sanchez, and Michael Martinez. 2015. “How America Reacted to Marriage Equality Ruling - CNNPolitics.” *CNN*. Retrieved November 19, 2020 (<https://www.cnn.com/2015/06/26/politics/supreme-court-same-sex-marriage-reaction/index.html>).
- Penney, Jonathon W. 2016. “Chilling Effects: Online Surveillance and Wikipedia Use.” 31:66.
- Pew Research. 2019. *Demographics of Internet and Home Broadband Usage in the United States*. Pew Research Center.
- Perrin, Andrew and Monica Anderson. n.d. “Share of U.S. Adults Using Social Media, Including Facebook, Is Mostly Unchanged since 2018.” *Pew Research Center*. Retrieved December 19, 2019 (<https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/>).
- Rogers, N., & Jones, J. J. 2019. *Using Twitter Bios to Measure Changes in Social Identity: Are Americans Defining Themselves More Politically Over Time?* <https://doi.org/10.13140/RG.2.2.32584.67849>
- Schauer, Frederick. 1978. “Fear, Risk and the First Amendment: Unraveling the Chilling Effect.” *BOSTON UNIVERSITY LAW REVIEW* 58:49.
- Simonetto, Deana. 2016. “Expanding Our Methodological Toolbox: The ‘Place’ of Twitter in the Ethnographic Endeavor.” *Qualitative Sociology Review* 12(1):98–112.
- Sinnenberg, Lauren, Alison M. Bittenheim, Kevin Padrez, Christina Mancheno, Lyle Ungar, and Raina M. Merchant. 2017. “Twitter as a Tool for Health Research: A Systematic Review.” *American Journal of Public Health* 107(1):e1–8.

State, Bogdan and Nils Wernerfelt. 2015. "America's Coming Out on Facebook." *Facebook Research*. Retrieved December 19, 2019 (<https://research.fb.com/blog/2015/10/americas-coming-out-on-facebook/>).

Supplemental Material for *Societal Pressures, Safety, and Online Labeling: Investigating LGBTQ Self-Identification in an Online Space*

Supplemental Figure 1. Daily incidence of Add and Delete events per each keyword.

Supplemental Figure 2. Segmented Regression fit to Prevalence of LGBTQ accounts. Black dots represent data. Color lines depict the best-fit slope to June and non-June periods. Prevalence is generally increasing, and the rate is faster within June (Pride) months.