Using Twitter Bios to Measure Changes in Self-Identity: Are Americans Defining Themselves More Politically Over Time?

Nick Rogers* and Jason J. Jones

Abstract: Are Americans weaving their political views more tightly into the fabric of their self-identity over time? If so, then we might expect partisan disagreements to continue becoming more emotional, tribal, and intractable. Much recent scholarship has speculated that this politicization of Americans' identity is occurring, but there has been little compelling attempt to quantify the phenomenon, largely because the concept of identity is notoriously difficult to measure. We introduce here a methodology, Longitudinal Online Profile Sampling (LOPS), which affords quantifiable insights into the way individuals amend their identity over time. Using this method, we analyze millions of "bios" on the microblogging site Twitter over a 4-year span, and conclude that the average American user is increasingly integrating politics into their social identity. Americans on the site are adding political words to their bios at a higher rate than any other category of words we measured, and are now more likely to describe themselves by their political affiliation than their religious affiliation. The data suggest that this is due to both cohort and individual-level effects.

Key words: identity; social identity theory; social media; ideology; political polarization; Twitter

1 Introduction

Because the distinctions are essential, we begin by noting what this paper does not intend to measure. This is not an analysis of whether the general American public, over the past several years, is exhibiting more politics-based behavior. By many measures, it is. Since 1970, the five presidential elections with the highest percentage of voter turnout have been the five most recent (2020, 2016, 2012, 2008, and 2004)[1]. Politically motivated consumer boycotts are on a steady rise[2]. As are campaign activities like donations, volunteering, and wearing political paraphernalia[3].

Nor does this paper explore whether Americans are becoming more likely to affiliate with a political party. Depending upon the data source, such memberships have stayed flat or slightly decreased over the past decade[4, 5]. Yet in a phenomenon known as "sorting", Americans are

becoming more segregated along ideological lines, in realms like neighborhood[6] and cultural exposure[7–9].

We do not ask whether people within those parties or ideologies are exhibiting an increase in negative emotions towards their political "others"—at an alarming rate, they are Refs. [10, 11]. According to the Pew Research Center[12], 16% of Republicans reported "Very Unfavorable" opinions of the Democratic party in 1994, but in 2017 that number had increased to 45%. Likewise, Democrats who report "Very Unfavorable" opinions of the Republican Party increased from 16% in 1994 to 44% in 2017.

What we do seek to measure is the extent to which Americans are increasingly *defining* themselves by political affiliations. We measure whether individuals are changing their sense of *identity*, in a way that saliently incorporates their politics.

This distinction—identity—is important. It represents a threshold beyond mere attitudes or behavior. Identity is an organizing, all-encompassing sense of self that *informs* those attitudes and behavior: I must hold these values and feelings *because* of this identity; I must

● Nick Rogers and Jason J. Jones are with the Department of Sociology, Stony Brook University, Stony Brook, NY 11794, USA. E-mail: Nicholas.L.Rogers@gmail.com.
∗ To whom correspondence should be addressed.

engage in these behaviors because it is fundamentally who I *am*.

An individual might endeavor to eat a plant-based diet, but nonetheless not subjectively identify as a "vegetarian". When this is the case, such an individual's eating behavior is likely to be more flexible and context-specific than an individual who has decided to claim vegetarianism as an identity[13]. Similar comparisons can be made between a "Gun Owner" and someone who merely happens to own a firearm; or a "Bisexual" and someone who occasionally has intercourse with someone of their own gender. In each case, if a person has embraced a certain attribute as central to their identity, such identity will guide their future behavior. Individuals seek to behave in ways that are consistent with their identity; behaviors (and feelings and values) that are perceived as inconsistent will be avoided or suppressed. A person's identity is a constraint on what they are "allowed" to choose, in any given situation[14, 15].

In the political context, a person for whom politics is not part of their self-perceived identity may still engage in political activity (online discussions, campaign contributions, following the daily news, etc.), but would feel relatively free to abandon those activities if they become dull, exhausting, or otherwise problematic. By contrast, an individual who defines himself as an Activist, or a Politico, or a Liberal, or a Republican, will continue to engage with current events, form firm opinions, and "perform" their group membership(s), even if it becomes detrimental to other aspects of their life. (This is because identity, as we will further elaborate upon, is resistant to change.)

The "identity threshold", then, is conceptually subtle but of major practical impact. If Americans define themselves increasingly by their political allegiances, their feelings towards political "others" can be expected to become more negative, and debate on matters of policy will become more emotional and intractable. Traditional ideas of political persuasion may be of little use, because changing someone's mind on a particular issue would require an adjustment to an entire sense of group identity. Existing political polarization becomes more likely to deepen than subside.

Numerous scholars have speculated, quite reasonably, that the identities of Americans are becoming more politicized[3, 16, 17]. But we are aware of no compelling attempt to *quantify* this phenomenon to the extent that it exists. This gap in knowledge is largely due to the fact that the concept of identity—being abstract, subjective, and dynamic—is notoriously difficult to measure empirically. In the context of politics, how may a scientist distinguish between a person who merely engages in political behavior, and a person who defines himself by it?

To explore whether political affiliation is indeed becoming a more prominent aspect of individuals' sense of social identity, we look to microblogging platform Twitter, where a user can write a "bio"—a brief, amendable personal synopsis that informs other users of biographical basics. We treat these bios as a proxy for an individual's sense of identity—the way the individual wishes to be seen in the social world. If Twitter users are editing their bios over time to add cues to their political affiliation, then we may tentatively conclude that such affiliation is becoming more tightly woven into the fabric of those users' identity.

In examining millions of bios across several years, we conclude that Americans are indeed amending their individual identities to reflect their politics. But before presenting these findings, some discussion of the relevant concepts is necessary.

2   Background

One of the core concepts of social psychology, a person's *identity*, exists in the intersection of how the person views himself, and how society views them. Identity is the "character" that an individual presents to the world, provided that such presentation is accepted by the "audience" of society[18, 19]. A person's sense of identity is therefore their fundamental concept of who they *are*, within the social world.

There are different classifications of identity. An individual has certain traits that make them feel distinct from other people—I am particularly tall, or I am unusually generous. This is known as *personal identity*. But people also define themselves by group memberships and placement within networks of roles—I am a mother; I am a doctor; and I am a Christian. This is *social identity*. Personal identity gives an individual a sense of differentness; social identity provides a sense of sameness, or belonging. It is social identity on which we focus here.

Social identity is often shaped by an individual's group memberships[20]. Although one might expect that an individual's groups memberships are constructed around pre-existing preferences and behaviors (e.g., I

admire Jesus Christ so therefore I shall be a Christian), in fact the converse is commonly true (I wish to be seen as a Christian, so therefore I will seek out things to admire about Jesus.). Individuals often adopt a desired group identity (or accept an ascribed one like race or gender), and allow their behaviors and preferences to be shaped by the dictates of that group. In political science experiments, for example, participants are easily persuaded to express support for a policy that is antithetical to their personal ideology, merely by being told that the policy was authored by their own political party[21]. For many people, then, beliefs are dictated by party membership, rather than party choice being dictated by beliefs.

2.1　Identity salience

Because every individual is situated in a number of groups and a number of roles, everyone has a multitude of social identities. Each identity is evoked and performed in some situations, and dormant in others. Sometimes the appropriate identity is suggested by the setting: Be a doctor at the hospital; a mother at home; a mischievous friend at the bachelorette party. But how does one choose which identity to perform, in environments with no clear cue? A trip to the grocery store, for example, or a cocktail party?

This dilemma is addressed by Stryker and Serpe[22], who demonstrated that individuals maintain a subconscious hierarchy of their identities, organized according to which ones have the most general utility. This concept is known as identity *salience*. A particular identity is more salient if it is seen as being more useful in times of social ambiguity. The more salient a particular identity becomes, the more often it takes precedence over other competing identities. It has been suggested that political affiliation is becoming a "mega-identity"[3] that organizes the other identities underneath it[23, 24]. We seek here to investigate this idea, by quantifying the extent to which individual Americans are incorporating political affiliation more saliently into their identities.

2.2　Tribalism: Group polarization and the backfire effect

Why might it matter, if identities are indeed becoming more politicized? Because people with salient group identities often fall prey to *tribalism**. The description of American politics as tribal is currently trendy in the popular press[26–28]. But a coherent, unified definition of that term is a bit elusive. Here, we mean tribalism as a condition in which individuals subjectively perceive each of the following: (1) Affiliation with a group defined by clear boundaries; (2) the existence of one or more groups of clearly-defined outsiders; (3) a zero-sum competition for resources (power, status, money, etc.) among the groups; and (4) ingroup loyalty as the paramount value in the competition.

The trouble with tribalism is that, even when the differences among the groups are minimal (or even arbitrary) and the stakes of competition are very low, intense (and irrational) animosity tends to define the intergroup dynamics. In a classic experiment by Sherif et al.[29], members of a boy's summer camp (Robbers' Cave) were randomly divided into two groups prior to their arrival, and initially kept separate from the other group. The campers were told that they would be engaging in competitions with the other group during the duration of camp, and that small trinkets would be awarded as prizes.

Benign ingroup pride was immediate—each group named themselves (the Eagles and Rattlers) and constructed a flag. But outgroup animosity started quickly thereafter, and escalated from subtle to alarming. Having still never met the rival campers, each group formed elaborate negative stereotypes of the type of boys they imagined the other group to be—cheaters, sloths, and idiots. They behaved rudely to each other when finally introduced, and within a day or two of competition, they were having to be physically restrained by the researchers from fist fights. Each group refused to share a bus with the other group, for field trips. They refused to dine at the same table. They accused each other of sabotaging their respective camp areas, only to be reminded by the researchers that each group had accidentally done the damage themselves. Group affiliation rapidly becomes internalized into social identity, which is then performed enthusiastically and defended fiercely. Judgment and perception are clouded in a fog of tribal loyalty.

Also relevant to the political context is the phenomenon of *group polarization*: Once a group

---

* Although we treat tribalism here as problematic, the ability to discern allies from outsiders may have been of benefit to group survival in our species' hunter-gatherer past, and could in fact be an evolved adaptation[25].

membership is established, mindsets within that group tend to homogenize and radicalize. As summarized by Sunstein[30]: "In a striking empirical regularity, deliberation tends to move groups, and the individuals who compose them, toward a more extreme point in the direction indicated by their own predeliberation judgments. For example, people who are opposed to the minimum wage are likely, after talking to each other, to be still more opposed; people who tend to support gun control are likely, after discussion, to support gun control with considerable enthusiasm; people who believe that global warming is a serious problem are likely, after discussion, to insist on severe measures to prevent global warming."

In other words, if a person is initially inclined towards a particular viewpoint, that viewpoint tends to become more exaggerated and deeply held when the person is exposed to a likeminded group, even if no single group member's viewpoint was exaggerated or deeply held to begin with[31].

If an individual is presented with information that discredits their group, or runs contrary to its views, the individual paradoxically becomes *more* entrenched within the group. This is known as the *backfire effect*—if a person is sufficiently invested in a belief (often because the belief is woven into their identity), they will respond to counterevidence by becoming even more certain of their initial position[32].

These phenomena combine to produce a predicament: If political affiliation is becoming a more salient aspect of people's social identity, and salient social identities tend to radicalize beliefs, foster animus, and resist change, then society might be losing the most important prerequisites for rational civic debate. Traditional methods of political persuasion may be of little use. Changing someone's mind on a particular issue would require an adjustment to an entire sense of identity.

Yet this is a difficult proposition to confirm, because of the challenges in measuring identity. To solve this problem, and conduct an empirical study of the relationship between political ideology and social identity, we look to popular microblogging platform Twitter.

3 Data

On Twitter, users post brief messages—"tweets", limited to 280 characters—which may be read by other Twitter users. Although users are able to adjust their privacy settings so that tweets are only viewable by specific people, Twitter's default setting is public, and most accounts remain publicly accessible. Although many academic studies have analyzed the content of tweets, little attention has been paid to another fundamental feature of Twitter: *personal bios*. When initially opening an account, each user is prompted to "Describe yourself in 160 characters". This biographical statement is then publicly visible to other users. While users may type anything they want into their bio field, some conventions have developed. Individuals typically list nouns that correspond to socially defined roles (e.g., father, mother, scientist, and superfan). It is common for bios to be comprised solely of such appellations. While in office, former US President Barack Obama's bio, for instance, read "dad, husband, President, citizen". Crucially for this study, Twitter bios are also freely amendable, but remain relatively stable. In our data, the average Twitter user posts a tweet roughly once a month, but amends their bios only once every year. This is consistent with scholarship finding that identity tends to resist major changes over time because their sole purpose is to provide a sense of continuity or sameness across social contexts and time[33] (If bios are analogous to identity, tweets themselves might be similar to a person's *moods*, which often fluctuate rapidly.).

Beginning in March 2015, we created a computer program to collect an automated, random 1% sample of publicly visible tweets, as well as the user information associated with each tweet as it appeared at that time. For this project, we filtered those tweets to only those from users with an American time zone as their time setting and their interface language set to English[†]. To date, this has resulted in observations from about 20 million distinct users.

We then copied the text of the bios (ignoring the tweets themselves) into a searchable Structured Query Language (SQL) database. We repeated this process for each year from 2015 to 2018, such that each year has its own database with its own sample of Twitter bios (When the same user was observed more than once per year, we chose one observation at random to keep.).

[†] This method undoubtedly includes some citizens of Canada in the samples. However, because the population of the United States is about nine times that of Canada, we think the observed trends are most substantially attributable to Americans.

Consequently, we may contrast the bios from each year's sample to measure changes in the words users choose when describing themselves. We call this set of four annual databases our "cross sectional" dataset. Of the 20 million total users who appear in the sample, about 3.5 million users were captured in all four years (2015–2018) of our study. We therefore have a "longitudinal" subset of our data, in which we can track specific users' bios across multiple years.

4 Method

The Twitter bio is similar to a well-established clinical measure of identity, the *Twenty Statements Test* (TST). Developed in 1954, the TST prompts a respondent to write out, in stream-of-consciousness, twenty brief self-descriptions[34]. Those descriptors are then coded in various ways depending on the aims of the researcher, to provide a picture of the respondent's sense of identity. For example, the researcher might estimate a respondent's level of social integration by comparing the number of social roles listed (e.g., mother, doctor, or Christian) with the number of personal traits (tall, intelligent).

We suggest here a loose analog to the TST, suited for the age of big data and online social media. We treat the Twitter bios as a self-report of users' sense of identity, and track over time the ways in which that sense changes. We term this methodology *Longitudinal Online Profile Sampling* (LOPS).

Consistent with our idea that political affiliation is becoming a more salient aspect of people's social identity, and using our large data set of Twitter bios, we explore the first of three hypotheses:

H1: Between 2015 and 2018, the prevalence of users who include explicit political keywords in their bios will increase.

To test this hypothesis, we face a challenge of deciding what words constitute "explicit political keywords". We confine our terms to those that convey group membership in a political collective with broad (rather than issue-specific) ideology. Although our resulting wordlist, presented in List 1, may be imperfect, we believe it is a fair representation of the spectrum of common, modern political group identities.

Of course, ideology and political affiliation might be signaled in any number of ways that are not explicit. Such *implicit* signals may in fact be more valuable, because they permit a person to indicate

List 1 Explicit political keywords.

Socialist
Communist
Marxist
Anarchist
Leftist
Liberal
Progressive
Democrat
Conservative
GOP
Republican
Libertarian
Alt-right

affiliation to fellow group members, without suffering a social penalty from "outsiders" with an opposing worldview. (Political scientists have long documented a phenomenon wherein individuals identify themselves as "moderate" or "independent" despite having consistently partisan views, presumably to avoid alienation from outgroup peers[35, 36].) Therefore we make a second hypothesis:

H2: Between 2015 and 2018, the prevalence of users who include implicit political keywords in their bios will increase.

Here, the challenge of constructing a wordlist is considerably more daunting. The number of ways an individual might *implicitly* signal their political ideology is limitless, even when confined to the written word. Although future scholarship might introduce more "objective" methods of compiling such a wordlist, we trusted our intuition and experience as political sociologists to choose the keywords in List 2.

At this point, we needed to validate our keywords: Are users indeed intending them in a political sense? Some of the keywords, in theory, might be used in other contexts. For example, "woke" might be used as a

List 2 Implicit political keywords.

Feminist
Woke
Activist
Red pill
Men's rights
Deplorable
Nasty woman
LGBTQ
Black lives matter
Blue lives matter
The 99%
BLM
MAGA

sleep-related verb rather than a politics-related adjective; "GOP" might merely be a typo for a word like "hop" or "got". To assess this, we took a random sub-sample of 100 bios with an explicit political keyword and 100 bios with an implicit political keyword. We read each bio and coded the relevant keyword usage as either "clearly political", "clearly non-political", or "ambiguous". The analysis reveals that users are employing the keywords in the context we expected. Of the sub-sample of *explicit* political keywords, 96% of them were "clearly political", 2% were "ambiguous", and only 2% were "clearly non-political". Of the sub-sample of *implicit* political keywords, 96% of them were "clearly political", 3% were "ambiguous", and only 1% were "clearly non-political".

As a final design matter, we needed to provide some measure of comparison to contextualize our results. A longitudinal increase in political keywords in Twitter bios may not be particularly meaningful, for instance, if there has been an equal increase in keywords from non-political realms. We investigate a third hypothesis:

H3: Between 2015 and 2018, any increase in the prevalence of political keywords will outpace the increase in comparable non-political keywords.

Here again, assembling such a list of terms presents problems of subjectivity. In an effort to compensate for this, we investigate keywords in three different realms: art, sports, and religion. Each of these realms is a useful comparison to politics, in that it seems likely to be incorporated into an individual's sense of identity but unlike immutable characteristics such as race and national origin, it is fluid enough that it might change in salience over time. The keywords for each realm are listed below in List 3.

List 3 Non-political keywords by category.

Art	Sport	Religion
Artist	Sports	Christian
Painter	Athlete	Jewish
Dancer	Golf	Muslim
Sculptor	Tennis	Buddhist
Designer	Basketball	Hindu
Filmmaker	Baseball	Baptist
Musician	Football	Methodist
Poet	Soccer	Catholic
Composer	Boxing	Mormon
Comedian/	MMA	Episcopal/episcopalian
comedienne	Hockey	Lutheran
Performer	Softball	Presbyterian
Actor/actress	Volleyball	Atheist
Playwright		

By comparing the prevalence rates for these keywords between categories and across time, we characterize recent trends. We report prevalence in units of *user counts per 10 000 users*. Thus, "40" corresponds to 40 users out of every 10 000, i.e., 80 000 matching users within 20 000 000 distinct observed users. It is important to note we are counting users, not words. A user counts towards a category once and only once if their bios contain at least one keyword. The bio "Conservative GOP Republican" counts *once* as a user matching the explicit keyword category.

5 Finding

Our results generally confirm our hypotheses. In Fig. 1, we show the prevalence of keywords in users' bios, by category and year. These results are for the cross-sectional dataset.

Indeed, Twitter users are becoming more likely to include a political keyword, either explicit or implicit, as part of their bio. For comparison, Fig. 1 also graphs the prevalence over time for religious keywords. The other categories have much higher baseline rates of prevalence, and so including them all on the same figure makes the slope of any given line too difficult to visually discern. The prevalence values for each category of keywords are listed below in Table 1.
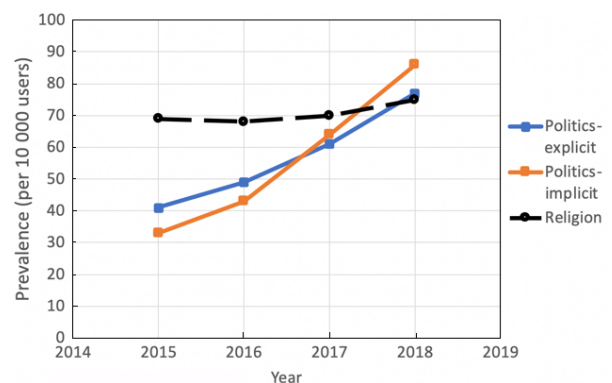

Fig. 1   Prevalence of bio keyword categories, by year–cross-sectional sample.

Table 1   Prevalence of bio keywords by category and year.

Year	Politics-explicit	Politics-implicit	Art	Sport	Religion
2015	41	33	339	332	65
2016	49	43	351	311	63
2017	61	64	361	298	64
2018	77	86	372	314	69

Note: Each value represents the number of incidences per 10 000 bios ("prevalence"). n = Approximately 10 million per year.

The usage of keywords in the category of religion and sports has remained essentially flat since 2015, and art-related keywords have climbed only a small degree relative to their overall usage. But the usage of political keywords has climbed steadily. Of particular interest, in 2018 Twitter users were more likely to describe themselves in political terms than in religious terms. In a nation that has traditionally been viewed as uniquely religious among its peer countries, this is notable.

We are interested not only in baseline usage rates of our keywords, but also in the extent to which prevalence is *changing*. Figure 2 plots the annual change for each category of keywords.

As Fig. 2 visualizes, the non-political categories have shown only moderate and fluctuating changes in prevalence since 2015. Both explicit and implicit political keywords, by contrast, consistently became more common. Twitter bios were 19.5% more likely to have an explicit political keyword in 2016 than 2015; prevalence grew by 24.5% in 2017, and 26.2% in 2018. Implicit keywords became 30.3% more prevalent in 2016, 48.8% in 2017, and 34.4% in 2018.

Although the raw number of Twitter users defining themselves by political affiliation is not as high as the number defining themselves by their arts or their sports participation, it is becoming increasingly common, at a much higher rate than any of the categories we measured.

But is this effect attributable to individual users amending their bios, akin to real identity change? Or is it rather caused by new users joining Twitter who are more political than the previous users, akin to a cohort effect? We can answer these questions by limiting our sample to the same Twitter users for each year—our longitudinal sample of about 3.5 million users. Figure 3 reports the prevalence and growth of the bio keyword categories, for that longitudinal sample.

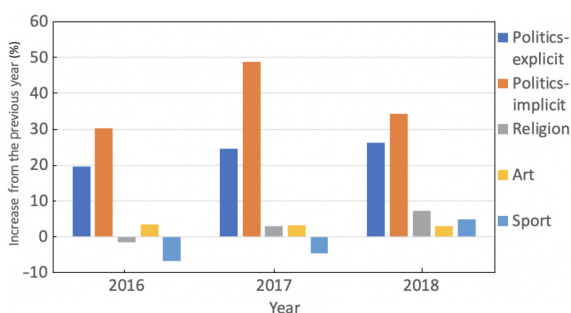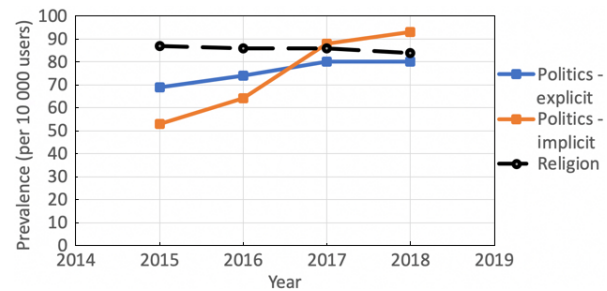As Figs. 3 and 4 illustrate, the trends in the cross-

Fig. 3 Prevalence of bio keywords categories, by year—longitudinal sample.
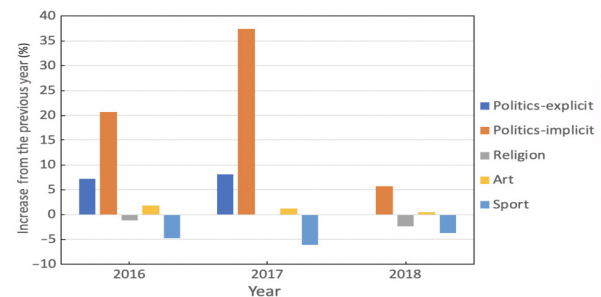

Fig. 4 Growth rate of bio keywords categories, by year—longitudinal sample.

sectional sample are also evident in the longitudinal sample. Over time, Twitter as a whole is gaining new users who are more political than the older users, but individual users are *also changing* to become more political.

As Fig. 5 illustrates, this effect is also evident when the keywords are disaggregated. Comparing 2015 bios to 2018 bios, *none* of the specific political keywords we chose showed a decrease in prevalence. Seven of the 26 political keywords remained flat, but these were keywords with a prevalence near zero (i.e., almost no one was using them in their bios at any point). The remaining 19 keywords all rose in prevalence across the 4-year
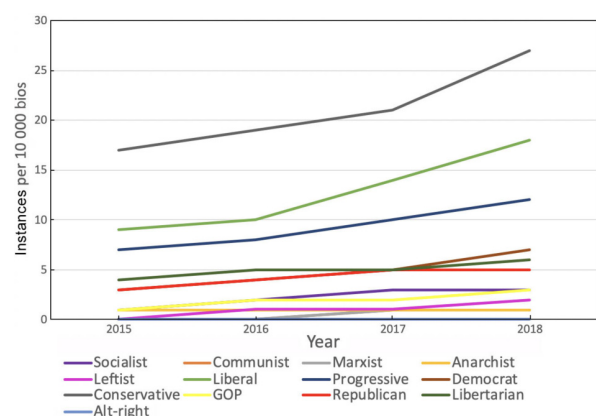

Fig. 2 Growth rate of bio keywords categories, by year—cross-sectional sample.

Fig. 5 Prevalence of "explicit polities" bio keywords, by year—cross-sectional sample.

period.

As Figs. 5 and 6 illustrate, the upward trend of political keywords as category is not attributable to any single keyword, but rather is common to nearly all of them.

For the differences we measure here—both across years, and across categories—we do not conduct tests for statistical significance. The first reason for this is theoretical. Because these are binary "count" variables rather than sample means, we know no ideal way to calculate variance within each variable and infer a statistically likely range of expected values. Also, because each set of observations might be partially dependent on prior observations (e.g., because the same person is captured in sequential samples, or because a person at Time 2 is socially influenced by Twitter bios they witnessed at Time 1), a fundamental assumption of standard significance tests would be violated. The second reason is a practical one: because of the sheer number of observations within each sample (i.e., millions of users), almost any observed difference between or among samples will be "significant" in the statistical sense. It is up to readers to determine for themselves whether the differences are substantively *meaningful*. We believe they are.

6   Discussion and Conclusion

To the extent that a person's Twitter bio is a valid measure of their sense of identity, Americans are defining themselves more saliently by their politics. This is important, because the formation of a group identity tends to change individual behavior in powerful ways. Through the phenomenon of "group polarization", people who begin with vague, weakly-held opinions tend to become more radical and dogmatic when put

Fig. 6   Prevalence of "implicit polities" bio keywords, by year—cross-sectional sample.

into like-minded groups. They also quickly develop hostile feelings towards outgroup members. Rational, evidence-based dissent tends to lose effectiveness within the groups, and in fact make group members even *more* invested in their original opinion.

To what may this increase in prevalence of political group identity be attributed? Is a more politically-engaged set of people joining Twitter for the first time, making the aggregate site more political than it was in prior years? Or are existing Twitter users amending their profiles to add a political keyword where formerly there were none? In other words, is this a generational/cohort effect, or is change occurring within individual identities? As our data reveal: both. Comparison between the cross-sectional and the longitudinal data suggests that (1) new entrants are more politically-oriented than the older participants they are joining or replacing, and also (2) individual people are amending their identities to be more political.

This dual nature of the phenomenon, as well as the effects it is likely to produce, portend a national polarization that is more likely to deepen than subside, in the short term. As Americans define themselves increasingly by their political allegiances, their feelings towards political "others" can be expected to become more negative, and debate on matters of policy will become more emotional and intractable.

Traditionally, a solution to the problem of tribalism has been found in the concept of "superordinate goals". Rival groups can put aside their perceived zero-sum differences when presented with a shared obstacle that requires cooperation to surmount. In the Robbers' Cave experiment, the Rattlers and Eagles were able to work together and even form intergroup friendships, once they were presented with obstacles that required cooperation for shared benefit[37]. Particular to our political context, some experimental research has suggested that priming a national identity (American) can mitigate partisan bias[38]. The attacks of September 11, 2011, for example, led to a period of bipartisan focus on international terrorism. Yet in the current political climate, such agreed-upon goals seem rare. Democrats and Republicans seem to diagnose distinct social maladies from each other, unable to even agree on shared definitions of problems.

Limitations and future inquiry. Although we believe our method provides a useful, digital-age measure of individual identity that is similar to the
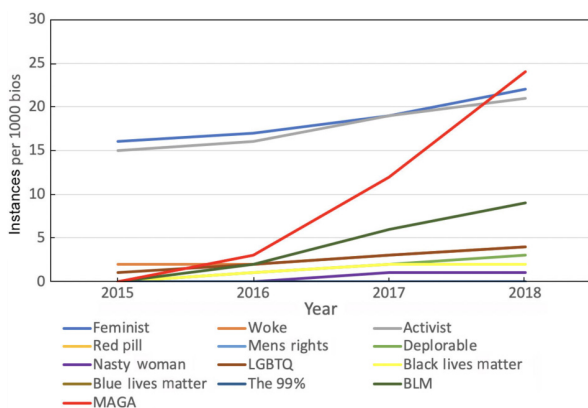
seminal Twenty Statements Test, there are imperfections worth noting.

First is the potential influence of "bots". It is well-established that Russian intelligence sought in 2016, and continues to seek, to influence American political discourse through the creation of social media accounts that pose as American users and spread divisive (and often fabricated) political content[39]. It is conceivable that our documented increase in prevalence of political keywords in bios is partially attributable to a growing number of these bots. However, our best evidence suggests any such influence is minimal. To investigate this possibility we tested random subsamples of our data using "Botometer", an automated tool to detect automated "bot" accounts. Almost all accounts received low scores. The mean for accounts in the longitudinal sample was 0.6 on a scale of 0 (probably not a bot) to 5 (probably a bot). The growth rate of botlike accounts fluctuated across our study period and could not account for the increases in political identity reported here. A full account of this analysis is included as the Appendix.

A second concern: Are our findings generalizable to the American general public, or is the politicization specific to Twitter users? To be sure, a sample of Twitter users is not the same as a random sample of Americans. In a recent study by Pew Research Genter[40], Twitter users are discovered to be younger, wealthier, and more educated than the United States at large. They are also modestly more liberal and more likely to say that voted in the last election. So it is conceivable that Twitter users are also more likely to adopt political identities than the general population. More data would be necessary to resolve this ambiguity. But we think that a general politicization of social identity is consistent with the other measures of politicization that we referenced in Section 1—voter turnout, affective polarization, cultural sorting, and so on.

Further, our sampling method samples tweets rather than users. Users who do not use tweets—who may have an account only to receive information or direct message—are thus not observed. These users may be systematically different from our sample of users who do use tweets, and the present method cannot speak to whether their self-identification is changing or not.

A third issue is the construction of our lists of keywords. We were sensitive to the possibility that certain "trendy" keywords could increase in prevalence not because individuals are defining themselves more

politically, but rather because the keywords themselves are becoming more popular and supplanting "outdated" keywords that are not in our lists. For example, a hypothetical Twitter user might have had an Obama-supportive "Yes We Can" phrase in their bio in 2015, but swapped it out in 2016 for a "Nasty Woman" reference to Hillary Clinton. Because the former phrase is not in our list, and the latter phrase is, our method would give the misleading impression that the user had "politicized" their bios, when in fact it was political all along.

We considered a number of methods that might limit the amount of subjectivity of that process. We searched for an adequate pre-existing keyword set, to no compelling avail. We analyzed the Twitter bios of several dozens of popular political figures, to see what descriptors they commonly employed. To our surprise, these individuals rarely used words that were even implicitly partisan, in their bios[‡]. We contemplated various Natural Language Processing techniques, to obtain frequently-used words on political hotbeds such as Reddit's r/politics subreddit. But ultimately we concluded that the utility of such methods would be outweighed by the drawbacks and complications. Future research may build upon these results by constructing more comprehensive (or selective) banks of keywords.

It would also be fruitful to expand upon these descriptive data and incorporate more layered analyses. With demographic information on our Twitter users, for example, we could conduct models to determine which characteristics are most correlated with changes in political identity. We could also analyze the users' tweets over time (rather than merely their bios), and analyze what sorts of rhetoric tends to portend or reflect a recent change of identity.

Continued inquiry on the matter is important: It is crucial to understand the dynamics underlying American political polarization. The stability of a people is

---

‡ For example, at the time of writing, Donald Trump's bio reads, "45th President of the United States" with a flag emoji. Sean Hannity identifies himself as "TV Host Fox News Channel 9 PM EST. Nationally Syndicated Radio Host 3-6 PM EST. http://Hannity.com Retweets, Follows NOT endorsements! Due to hackings, no DM's!" Joe Biden's bio is "Senator, Vice President, 2020 candidate for President of the United States, husband to @DrBiden, proud father & grandfather. Loves ice cream, aviators & @Amtrak." MSNBC host Chris Matthews defines himself as "Host of @hardball M-F at 7PM ET on @MSNBC and author of 'Bobby Kennedy: A Raging Spirit'."

dependent on some sense of unifying solidarity. Without it, order is imperiled and chaos invited.

7   Appendix: On the Question of Bots

In using social media bios as a proxy for individuals' sense of identity, we make a fundamental assumption: the bios are crafted in good faith by actual human beings. Yet it is clear that the assumption is not completely true — the existence of "bot" accounts is both empirically documented, and anecdotally obvious to anyone who is sufficiently active on social media[§]. The relevant question, then, for scientific findings from social media data is this: Is the discovered phenomenon still evident, *net of* the influence of bots? Put another way: assuming part of the findings are attributable to bots, is there nevertheless a significant part that is *not*?

In the current work, there is special reason to wonder whether the bios of bot accounts are becoming increasingly political. A Special Counsel's investigation into the 2016 US Presidential Election reported that Russian Intelligence created a number of bots across social media platforms, to sow discord among the American public and weaken the support for candidate Hillary Clinton[39]. Since these bots were disseminating political content through tweets, it is reasonable to assume that many of them were conveying political content through bios as well.

Determining the *extent* of this problem is difficult. Doing so with perfect accuracy is impossible. Some bots make for convincing humans, and some real humans make for convincing bots. Twitter has estimated that less than 8.5% of its profiles are bots[41]. Some have contested this number[42]. Some of those bots are "legitimate", such as those programmed to automatically share weather updates or seismic activity. Presumably only a small number of the overall bot population are designed to influence political discourse. The platform deleted about 4500 accounts that it deemed to be bots from Russia or Iran in 2018[43], and approximately another 5000 in 2019[44]. These are insignificant numbers relative to the Twitter population as a whole, but it is safe to assume that some "perhaps many"—political

---

[§] For convenience, we use the term "bot" to include completely automated accounts, accounts that are a mix between individually-crafted content and automated content (sometimes referred to as "cyborgs"), and accounts that are entirely human-crafted but are operated in bad faith by foreign actors disguising themselves as American citizens.

bot accounts evade discovery by Twitter. How many? Enough to materially influence our findings that bios are becoming more political over time?

To analyze this question, we conducted an additional inquiry into our data. We began by taking random 100-account subsamples from the longitudinal group and from each year of these categories of our cross-sectional sample: (a) Users whose bio included an explicit political keyword; and (b) users whose bio included an implicit political keyword. This resulted in 9 separate 100-user subsamples.

For each subsample, we entered each User Identification Number into the online tool "Botometer". Botometer (formerly "Bot or Not", available at https://botometer.iuni.iu.edu) is a project by network and computer scientists at the University of Indiana. (For an explanation of its methodology see Varol et al.[42]) The program checks the activity of a Twitter account and gives it a score based on how likely the account is to be a bot. Higher scores are more bot-like. Scores are reported on a color-coded scale from 0–5, with scores between 3.0–4.0 ("orange") being viewed as suspicious, and scores above 4.0 ("red") assumed to be bots.

Botometer was not always able to report a score for the users in our subsamples. In some cases, the User ID Number was "not found", which likely means that the user's account has been deleted since being captured into our sample. (We have no way of knowing whether the user cancelled their account themselves, or whether it was involuntarily deleted by Twitter.) In other cases, the User ID was not authorized, which likely means that the account still exists, but has switched its privacy settings from "public" to "private". (Botometer, like us, can only analyze accounts that are publicly viewable.)

It is unclear whether either of these "missing data" designations is evidence of an account being a bot. In fact, accounts that were "not authorized" seem *less* likely than other profiles to be bots, because the purpose of bots is to build as wide (i.e., public) an audience as possible. And because Twitter users delete their accounts for all sorts of non-nefarious reasons, it would make sense that the number of "not found" accounts would grow cumulatively from year to year. Nevertheless, we include these designations in our analysis, for purposes of a conservative estimate.

In Table 2, we report the results from the Botometer analysis. For each subsample, we list the number of "orange" accounts, the number of "red" accounts, and the

Table 2   Botometer results: Relevant counts.

Group	Year	Not found	Not available	Orange	Red	Total count	Change (%)
Longitudinal group		4	7	3	1	15	
Explicit politics	2015	20	6	5	4	35	
	2016	17	7	4	3	31	–11.40
	2017	25	4	5	0	34	9.70
	2018	21	1	4	1	27	–20
Implicit politics	2015	13	8	2	3	26	
	2016	18	7	1	2	28	7.70
	2017	20	7	1	1	29	3.60
	2018	21	2	3	0	36	24.10

Note: For each group, $n = 100$. "Change" is measured from the previous year's value.

Table 3   Botometer results: Mean scores.

Group	Year	Mean	Change (%)
Longitudinal group		0.623	
Explicit politics	2015	1.03	
	2016	0.92	–10.7
	2017	0.86	–6.5
	2018	0.79	–8.1
Implicit politics	2015	0.72	
	2016	0.64	–11.1
	2017	0.68	6.3
	2018	0.77	13.2

Note: For each group, $n = 100$. "Change" is measured from the previous year's value.

number of not found/not authorized accounts. For our purposes, we are not so much interested in the *baseline* number of these accounts in any given year. Rather, we look to whether those numbers are *changing over time.* If the prevalence of political keywords in bios is growing from year to year, but the number of bots is not, we may assume our measures of identity are valid. So the final column reports the percent change from the previous year.

There is no consistent trend in the number of bots found within each group across the years. In two of the years for the explicit politics group, the number of bots actually decreased from the previous year. When we ignore the not found and not available designations, the growth of bots disappears entirely. Among the explicit politics group, the combined number of orange and red Botometer scores shrinks from 9 to 7 in 2016, then to 5 in 2017 and 2018. Among the implicit politics group, that count shrinks from 5, to 3, to 2, and rises again to 3 in 2018.

Another way of measuring the growth rate of bots is to calculate the *mean* of Botometer scores from year to year. Table 3 reports those trends.

Here again, there is only mixed evidence of the growth of political bots in our sample over time. The mean Botometer score actually decreases, in 4 of the 6 measurable years. When the percent change values of Tables 2 and 3 are compared with the growth rate graph in the body of the paper (Fig. 2), one can see that the annual increases in political keywords substantially outpaces any increase in the number of bots, for any given year. By our best estimates, then, our findings that individual Americans are showing growing signs of politicized identities between 2015–2018 are indeed net

of the influence of bots.

These methods are an imperfect match for a challenging question. Botometer scores have known limitations[¶]. But it is folly to hold that any data from social media are necessarily invalid merely because some accounts are bots. There are confounding variables of unknown value in any social-scientific endeavor. As with any other study, we acknowledge those variables and do our best to control for them.

References

[1] FairVote, Voter Turnout, https://www.fairvote.org/voter_turnout#voter_turnout_101, 2021.
[2] C. Heldman, *Protest Politics in the Marketplace*: *Consumer Activism in the Corporate Age*. New York, NY, USA: Cornell University Press, 2017.
[3] L. Mason, *Uncivil Agreement*: *How Politics Became Our Identity*. Chicago, IL, USA: The University of Chicago Press, 2018.
[4] Pew Research Center, Party identification trends 1992–2017, https://www.people-press.org/2018/03/20/party-identification-trends-1992-2017/, 2018.
[5] Gallup, Party affiliation: Historical trends, https://news.gallup.com/poll/15370/party-affiliation.aspx, 2021.

¶ The FAQ section of the Botometer website notes: Bot detection is a hard task. Many criteria are used in determining whether an account is controlled by a human or a bot, and even a trained eye gets it wrong sometimes. If this task were easy to do with software, there would not be any bots—Twitter would have already caught and banned them! Another point to keep in mind is that humans and machines have different strengths when it comes to pattern recognition. Some "obviously" bot/human accounts according to a human observer will fool a machine learning algorithm. For example, Botometer sometimes categorizes "rganizational accounts" as bot accounts. Likewise, an algorithm may confidently classify some accounts that humans have a hard time with. The best approach is to use this tool to complement, not to replace, your own judgement.

[6] B. Bishop, *The Big Sort*: *Why the Clustering of Like-Minded America is Tearing Us Apart*. Boston, MA, USA: Mariner Books, 2008.

[7] N. Rogers, Split screens: A content analysis of american liberals' and conservatives' respective television favorites, *Psychol. Pop. Med.*, vol. 9, no. 1, pp. 45–58, 2020.

[8] D. DellaPosta, Y. R. Shi, and M. Macy, Why do liberals drink lattes? *Am. J. Sociol.*, vol. 120, no. 5, pp. 1473–1511, 2015.

[9] P. DiMaggio, Are art-museum visitors different from other people? The relationship between attendance and social and political attitudes in the United States, *Poetics*, vol. 24, nos. 2–4, pp. 161–180, 1996.

[10] A. I. Abramowitz and S. Webster, The rise of negative partisanship and the nationalization of U.S. elections in the 21st century, *Elect. Stud.*, vol. 41, pp. 12–22, 2016.

[11] S. Iyengar and S. J. Westwood, Fear and loathing across party lines: New evidence on group polarization, *Am. J. Pol. Sci.*, vol. 59, no. 3, pp. 690–707, 2015.

[12] Pew Research Center, Partisanship and political animosity in 2016, http://www.people-press.org/2016/06/22/partisanship-and-political-animosity- in-2016/, 2016.

[13] J. Greenebaum, Veganism, identity and the quest for authenticity, *Food Cult. Soc.*, vol. 15, no. 1, pp. 129–144, 2012.

[14] A. D. Cast, Identities and behavior, in *Advances in Identity Theory and Research*, P. J. Burke, T. J. Owens, R. T. Serpe, and P. A. Thoits, eds. Boston, MA, USA: Springer, 2003.

[15] P. J. Burke and D. C. Reitzes, The link between identity and role performance, *Soc. Psychol. Q.*, vol. 44, no. 2, pp. 83–92, 1981.

[16] J. J. Van Bavel and A. Pereira, The partisan Brain: An identity-based model of political belief, *Trends Cogn. Sci.*, vol. 22, no. 3, pp. 213–224, 2018.

[17] L. Huddy, From social to political identity: A critical examination of social identity theory, *Polit. Psychol.*, vol. 22, no. 1, pp. 127–156, 2001.

[18] E. Goffman, *The Presentation of Self in Everyday Life*. New York, NY, USA: Penguin Random House, 1959.

[19] G. P. Stone, Appearance and the self: A slightly revised version, in *Social Psychology Through Symbolic Interaction*, G. P. Stone and H. A. Farberman, eds. New York, NY, USA: John Wiley and Sons, 1981.

[20] H. Tajfel and J. Turner, An integrative theory of intergroup conflict, in *The Social Psychology of Intergroup Relations*, W. G. Austin and S. Worchel, eds. Monterey, CA, USA: Brooks/Cole, 1979.

[21] G. L. Cohen, Party over policy: The dominating impact of group influence on political beliefs, *J. Pers. Soc. Psychol.*, vol. 85, no. 5, pp. 808–822, 2003.

[22] S. Stryker and R. T. Serpe, Identity salience and psychological centrality: Equivalent, overlapping, or complementary concepts? *Soc. Psychol. Q.*, vol. 57, no. 1, pp. 16–35, 1994.

[23] P. J. Egan, Identity as dependent variable: How americans shift their identities to align with their politics, *Am. J. Pol. Sci.*, vol. 64, no. 3, pp. 699–716, 2019.

[24] L. D. Davenport, The role of gender, class, and religion in biracial Americans' racial labeling decisions, *Am. Sociol. Rev.*, vol. 81, no. 1, pp. 57–84, 2016.

[25] L. Cosmides, J. Tooby, and R. Kurzban, Perceptions of race, *Trends Cogn. Sci.*, vol. 7, no. 4, pp. 173–179, 2003.

[26] D. Roberts, Donald trump and the rise of tribal epistemology, https://www.vox.com/policy-and-politics/2017/3/22/14762030/donald-trump-tribal-epistemology, 2017.

[27] A. Sullivan, Cam America survive tribalism? New York Magazine, http://nymag.com/daily/intelligencer/2017/09/can-democracy-survive-tribalism.html, 2017.

[28] A. Chua, The destructive dynamics of political tribalism, The New York Times, https://www.nytimes.com/2018/02/20/opinion/destructive-political-tribalism.html/, 2018.

[29] M. Sherif, O. J. Harvey, B. J. White, W. R. Hood, and C. W. Sherif, *Intergroup Conflict and Cooperation*: *The Robbers Cave Experiment*. Norman, OK, USA: University of Oklahoma Book Exchange, 1961.

[30] C. R. Sunstein, The law of group polarization, *J. Polit. Philos.*, vol. 10, no. 2, pp. 175–195, 2002.

[31] D. G. Myers and H. Lamm, The group polarization phenomenon, *Psychol. Bull.*, vol. 83, no. 4, pp. 602–627, 1976.

[32] B. Nyhan and J. Reifler, When corrections fail: The persistence of political misperceptions, *Polit. Behav.*, vol. 32, no. 2, pp. 303–330, 2010.

[33] M. Rocque, C. Posick, and R. Paternoster, Identities through time: an exploration of identity change as a cause of desistance, *Justice. Q.*, vol. 33, no. 1, pp. 45–72, 2016.

[34] M. H. Kuhn and T. S. McPartland, An empirical investigation of self-attitudes, *Am. Sociol. Rev.*, vol. 19, no. 1, pp. 68–76, 1954.

[35] C. B. Hawkins and B. A. Nosek, Motivated independence? Implicit party identity predicts political judgments among self-proclaimed independents. *Pers. Soc. Psychol. Bull.*, vol. 38, no. 11, pp. 1437–1452, 2012.

[36] B. E. Keith, *The Myth of the Independent Voter*. Berkeley, CA, USA: University of California Press, 1992.

[37] M. Sherif, Superordinate goals in the reduction of intergroup conflict, *Am. J. Sociol.*, vol. 63, no. 4, pp. 349–356, 1958.

[38] M. S. Levendusky, Americans, not partisans: Can priming american national identity reduce affective polarization? *J. Polit.*, vol. 80, no. 1, pp. 59–70, 2018.

[39] R. Mueller, *Report on the Investigation Into Russian Interference in the 2016 Presidential Election*. U.S. Department of Justice, 2019.

[40] Pew Research Center, Sizing up twitter users, https://www.benton.org/headlines/sizing-twitter-users, 2019.

[41] Twitter, Form 10-Q, filed with the U.S. securities and exchange commission, https:// www.sec.gov/Archives/edgar/data/1418091/000156459014003474/twtr-10q_20140630.htm, 2014.

[42] O. Varol, E. Ferrara, C. A. Davis, F. Menczer, and A. Flammini, Online human-bot interactions: Detection, estimation, and characterization, in *Proc. 11th Int. AAAI Conf. on Web and Social Media*, Montreal, Canada, 2017.

[43] Twitter, Enabling further research of information operations on twitter, https://blog.twitter.com/en_us/topics/company/2018/enabling-further-research-of-information-operations-on-twitter.html, 2018.

[44] Twitter, Information Operations on Twitter: Principles, Process, and Disclosure, https://blog.twitter.com/en_us/ topics/company/2019/information-ops-on-twitter.html, 2019.

Nick Rogers is a PhD candidate in sociology at Stony Brook University, and teaches courses in statistics and research methodology at New York University's Center for Global Affairs. He received the MA degree from Georgia State University in 2016, and the JD degree from Washington University in St. Louis in 2007. His research focuses primarily on political ideology and social polarization. Prior to academia, he practiced law in his hometown of Little Rock, Arkansas. In addition to scholarly publications, his writing has appeared in *The New York Times and Arkansas Life*.

Jason J. Jones is an associate professor in the Department of Sociology and the Institute for Advanced Computational Science at Stony Brook University. He received the PhD from University of California- San Diego in 2011. He leads the Computational Social Science of Emerging Realities Group and has authored 19 peer-reviewed publications. His research interests include individual identity and the effects of artificial intelligence on society.