

Male, Female, and Nonbinary Differences in UK Twitter Self-descriptions: A Fine-grained Systematic Exploration

Mike Thelwall[†], Saheeda Thelwall, Ruth Fairclough

Statistical Cybermetrics Research Group, University of Wolverhampton, UK

Abstract

Purpose: Although gender identities influence how people present themselves on social media, previous studies have tested pre-specified dimensions of difference, potentially overlooking other differences and ignoring nonbinary users.

Design/methodology/approach: Word association thematic analysis was used to systematically check for fine-grained statistically significant gender differences in Twitter profile descriptions between 409,487 UK-based female, male, and nonbinary users in 2020. A series of statistical tests systematically identified 1,474 differences at the individual word level, and a follow up thematic analysis grouped these words into themes.

Findings: The results reflect offline variations in interests and in jobs. They also show differences in personal disclosures, as reflected by words, with females mentioning qualifications, relationships, pets, and illnesses much more, nonbinaries discussing sexuality more, and males declaring political and sports affiliations more. Other themes were internally imbalanced, including personal appearance (e.g. male: beardy; female: redhead), self-evaluations (e.g. male: legend; nonbinary: witch; female: feisty), and gender identity (e.g. male: dude; nonbinary: enby; female: queen).

Research limitations: The methods are affected by linguistic styles and probably under-report nonbinary differences.

Practical implications: The gender differences found may inform gender theory, and aid social web communicators and marketers.

Originality/value: The results show a much wider range of gender expression differences than previously acknowledged for any social media site.

Keywords Twitter; Online self-presentation; Gender; Nonbinary genders; Twitter bios

Citation: Thelwall, M., Thelwall, S., & Fairclough, R. (2021). Male, female, and nonbinary differences in UK Twitter self-descriptions: A fine-grained systematic exploration. *Journal of Data and Information Science*, 6(2). <https://doi.org/10.2478/jdis-2021-0018>

Received: Dec. 23, 2020

Revised: Feb. 6, 2021

Accepted: Feb. 7, 2021



[†] Corresponding author: Mike Thelwall (Email: m.thelwall@wlv.ac.uk, ORCID: 0000-0001-6065-205X).

1 Introduction

Gender is part of our identity, influencing how we behave and interact with others (e.g. Witt & Wood, 2010; Wolf, 2000). Knowledge of gendered behaviours is important because gender is a core part of the human experience. Gender understanding may also inform those using the social web for communication purposes, from mental health professionals to marketers (Bettany et al., 2010). In particular, information about fine-grained gender expression differences may help a better understanding of how gender influences online interactions, supporting better decision-making.

Whilst human biological male and female sexes have genetic definitions, albeit with intersex and other exceptions (WHO, 2020), gender is a complex social construction with its expression varying between societies and over time. For example, higher education was an almost exclusively male trait when females were barred from UK universities but now females outnumber males in UK university degree awards (57.8% female, 42.2% male, 0.06% nonbinary in 2018) (HESA, 2019) so it has become a moderately female-associated trait. A person may declare or inhabit a gender (their *gender identity*) differing from their biological sex, such as through being transgender or nonbinary.

Male and female genders are associated with culturally-specific sets of behaviours, often similar to those for biological sexes, known as *gender expressions*. For example, a man in the UK in the early 21st century is more likely to watch sport on TV, although not all men do and some non-males watch TV sports too (Abraham, 2020). Similarly, a woman is more likely to have a feminine hairstyle, clothes designed for women and work in a caring job, but many don't. Thus, a person realising that most of their preferred activities are male-associated might feel themselves to inhabit a male gender, irrespective of their biological sex. Conversely, another person may have a predominantly male gender expression but feel themselves to be purely female because of their biological sex, so gender identity can be completely different from gender expression (Pullen Sansfaçon et al., 2020; Zucker, 2015). *Gender nonconformity* in this sense of gender expression differing from gender identity is common, but not universal, in the lesbian and gay community, for example (Mackay, 2019; Rieger et al., 2010). Thus, whilst gender choices are made in the context of tacit knowledge of a range of gendered behaviours, they do not always reflect them, and gender is essentially an internal feeling based on biological sex and/or socialised behaviours. All choices of gender identity are equally valid and to be respected, irrespective of biological sex, gender expression and sexuality (APA, 2015).



Nonbinary genders encompass a range of gender identities, not all on the spectrum of male to female, and not necessarily constant over time. A nonbinary person might fall on the spectrum between male and female (demigender, transmasculine, transfeminine, genderfluid), may have multiple genders, or may be agender (no gender) and outside the male/female spectrum. Transgender people, who identify with a gender different from their birth-assigned sex, aren't nonbinary if they identify as male or female. Nonbinary status is separate from sexual orientation. One of the few behaviours that might be expected of a nonbinary person in general (including those not on the male/female spectrum) is the preference for linguistic terms associated with their status, such as they/them pronouns (e.g. McGlashan & Fitzpatrick, 2017), although this is not universal and is language-specific (Miltersen, 2018).

It is not clear how gender influences self-presentation on the social web, beyond a few aspects that have been tested for (see below). Moreover, it is not clear that there would be *any* self-presentation styles associated with nonbinary social web users since there are many types of nonbinary gender. This article therefore seeks systematic evidence of gendered identities on the social web for nonbinary, female, and male tweeters. Twitter was investigated because it is the largest mainly public social web site that allows data harvesting for research purposes. It therefore provides the largest-scale opportunity to study social web gendered identities. Twitter.com was the 46th most visited website in December 2020, according to Alexa.com (www.alexa.com/siteinfo/twitter.com) but its primary use may be through smartphone apps. The focus is on the UK because whilst the UK and USA both have large numbers of Tweeters, extensive astroturfing (Keller et al., 2020) seems likely in the USA in 2020 in the run up to the presidential election. In contrast, the next UK parliamentary election is set to be at the end of 2024. Thus, active UK Twitter accounts in 2020 seem less likely to be fake. Tweeters can choose to express their gender identity in their tweets, profile pictures or profile biographies. Biographies are short free text segments that appear under a user's profile picture on their home page. They are most suitable for identity analyses because they primarily contain self-description information, such as age, gender, occupation, relational status, religion, hobbies and political affiliation. For practical reasons, explained below, the biographies analysed were of people that had tweeted about COVID-19. The following question drives the study.

- How does self-presentation in UK Twitter biographies differ by gender?



2 Literature review

This section introduces research into gendered behaviours to set the context for the results, with additional literature drawn upon later to discuss these results.

2.1 Gender expression differences

Although gender expression differences pervade societies, there are disagreements about their underlying causes. From a social role theory perspective (Eagly, 1987), humans discover gendered behavioural expectations from birth and thus play learned gendered roles. Gendered mass media coverage (Garg et al., 2018) is part of the gender socialisation process, but routine daily interactions and expectations are presumably more important. In contrast, the biological perspective suggests that biological sex genetic differences may be the root causes of some or all female-male gender differences. An evolutionary psychology explanation of sex differences emphasises natural selection to favour traits useful for the biological sex role divisions of labour in the hunter-gatherer societies that dominated human existence (Buss, 1995). Gender expectations vary between societies (Costa, Terracciano, & McCrae, 2001) and change over time (Eagly et al., 2020), however. In the USA, for example, expectations for females have changed the most (Diekmann & Eagly, 2000) and male traits have become increasingly acceptable for women (Twenge, 1997). The social role theory approach seems better able to explain the complexity of gender differences in modern life, but it is possible that neither explanation is completely wrong.

Previous research has theorised about the nature of underlying gender expression differences. Social role theory argues that males are socialised to become more agentic (e.g. assertiveness, independence, mastery) whereas females are socialised to be more communally oriented (e.g. concerned about others, expressive, friendly). These traits underly many gender expression differences (Eagly & Wood, 2016). From the biological perspective, human males may be intrinsically more interested in tangible inanimate “things” (e.g. machines) whereas females may be more interested in other humans: a people-things dichotomy. If true, this would explain many gender differences in vocational (e.g. male carpenters vs. female nurses) (Lippa, 1998), and academic (Su & Rounds, 2015) interests, although there are exceptions (e.g. cell biology is a female majority interest). Social role theory can explain the same differences since thing-oriented jobs tend to satisfy agentic goals and people-focused jobs tend to satisfy communal goals (Thelwall et al., 2019).

Although nonbinary genders have presumably existed for as long as people, they are under-researched (Matsuno & Budge, 2017). No systematic study has reported characteristics or preferences of nonbinary people, focusing instead on health



(Richards et al., 2016), lived experiences (Bradford et al., 2019), or anti-nonbinary harassment (Nadal et al., 2016). Whilst there is no reason to believe that there are many behavioural patterns for nonbinary genders as a whole since they reflect a range of identities, one study has suggested that the importance of self-description words and narratives is a frequent concern for nonbinary people (Bradford et al., 2019).

2.2 Self-disclosure online

People control the information they share to manage the impression that they make on others (Goffman, 1959). For example, teens sometimes post content online to appear interesting and attractive (Yau & Reich, 2019). Many factors influence the extent to which people self-disclose on the social web, including age (Nosko, Wood, & Molema, 2010) and personality (Grieve, March, & Watkinson, 2020). Lower self-esteem associates with higher self-promotion content on Facebook (Mehdizadeh, 2010). Oppressed groups, including perhaps nonbinary and transgender users, may self-censor more because of a greater need for impression management to counteract negativity towards them or negative stereotypes (Pitcan, Marwick, & Boyd, 2018). Related to this, public social media generate “context collapse”, with users unable to tailor their self-presentation to different audiences (e.g. parents, teachers, friends, colleagues, future employers, customers) (Baym & Boyd, 2012). Thus, some aspects of the self may be hidden from public profiles to appeal to a general imagined audience (Marwick & Boyd, 2011).

A content analysis of 2,633 English-language Australian/UK/US Twitter user biographies found five types of information: Relational (relationships with other people); Occupational; Political, Ethnic/religious; and Stigmatized (membership of a stigmatised group) (Priante et al., 2016). Thus, Twitter profiles contain multiple types of information as part of the owner’s online identity.

2.3 Online gender expression differences

There are gender differences in online presentation styles, although perhaps weaker than offline (Oberst et al., 2016). These are only known for males and females, and mainly concern US students and/or Facebook. For example, male Facebook photos are more likely to accentuate status whereas females are more likely to showcase family relationships (Tifferet & Vilnai-Yavetz, 2014). These differences extend to stylistic variations. For example, males were more likely than females to use the possessive pronoun “my” when describing their partner in Facebook messages in a study of 75,000 users (Schwartz et al., 2013). These Facebook users also showed age, personality and geographic differences in post language.



For Twitter, female tweets have been suggested to be more personal (Weathers et al., 2014). More personal information was also shared by females (and parents) in a study of 187 tweeters (Walton & Rice, 2013). Privacy controls are more important to females than males on social network sites, so the lack of these on Twitter compared to other sites, such as Facebook, may reduce female participation in Twitter (Kuo et al., 2013) and may hide female-associated private issues from Twitter. A study of 69 top-ranked tennis players found the men to focus on being a sports fan whereas the women spent more time on general brand management (Lebel & Danylchuk, 2012), showing that specific types of user may have their own gender differences.

3 Methods

The research design uses word association thematic analysis (Thelwall, 2021), a method to find differences between sets of texts by identifying words that occur statistically significantly more frequently in one subset than another. In this case the texts are Twitter self-descriptions and the subsets are self-descriptions created by different genders or gender groups. This is more appropriate than content analysis or thematic analysis because it can identify fine-grained differences at a level that would not be identified by human coders (Thelwall et al., 2021). The method has three stages (Figure 1).

1. Automatically create a large systematic sample of UK male, female and nonbinary Twitter users and their Twitter biographies (the top three boxes in Figure 1).
2. Automatically extract words that occur more often than expected in the Twitter biographies of each gender type (Word Association Detection).
3. Manually group the extracted words into themes (Word Association Contextualisation and Thematic Analysis).

3.1 Stage 1: Sample creation

It is not possible to directly sample UK Twitter users since there are no comprehensive public lists. Instead, users were found indirectly by searching for tweets and identifying the tweeters. Queries are needed for the search stage, and a set of COVID-19 queries was used for this. This virus affects almost everyone in the UK and may therefore generate a large sample of active tweeters. The following queries were submitted to Twitter every 15 minutes from 10 March to 30 June 2020, specifying English language: coronavirus, "corona virus", covid-19, covid19, #stayhome, #socialdistancing, #quarantinelife, #lockdown, #washyourhands,



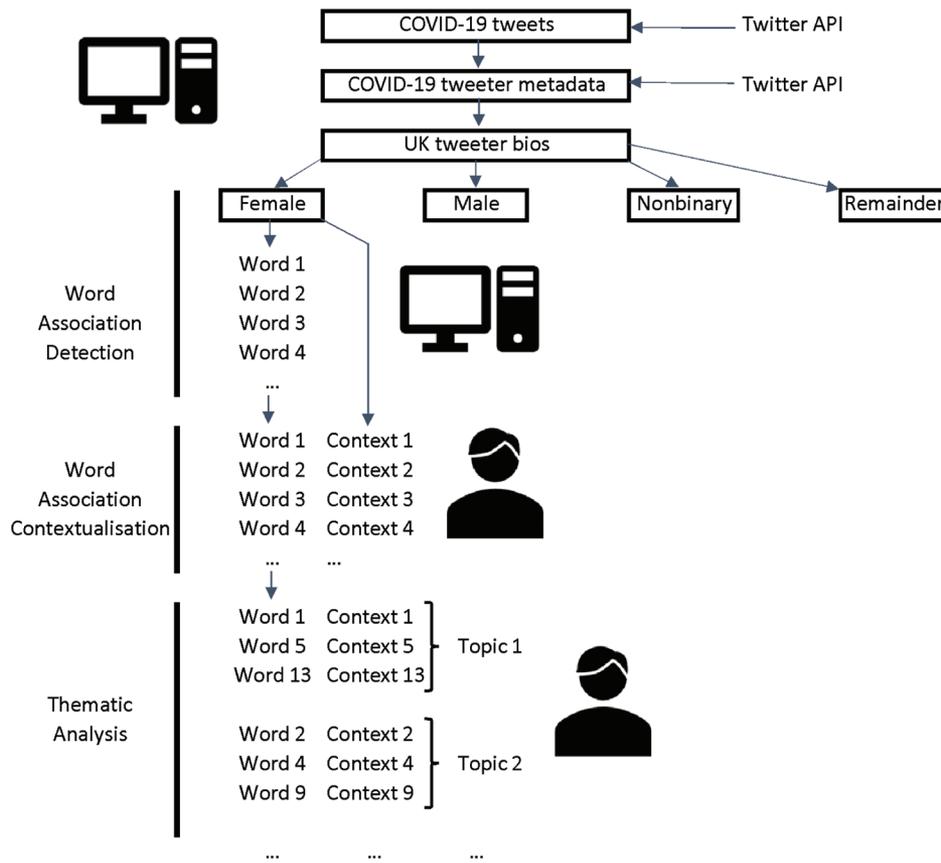


Figure 1. The research design, illustrating the multiple automated and two manual stages for the female themes.

#stayathome, #selfisolation, #workingfromhome, #quarantine. The Twitter Applications Programming Interface (API) was used through the free Mozdeh software to submit the queries and download the matching tweets. A complete list of users was extracted from the metadata associated with these tweets.

In July 2020 the Twitter API was used for a second time to download the Twitter biographies and locations of all the users identified above (i.e. users that had tweeted about COVID-19). Users specifying the UK as their location were retained and the remainder discarded. At the time, the Twitter API (v1.1) did not include user information along with tweets, but this is an option in the Twitter API v2.0 introduced later in 2020.

Twitter does not record gender, so this was inferred from usernames and pronouns in biographies. Lists of male and female gendered first names was obtained from



public US census 1990 tables of popular first names by gender, supplemented with Gender-API.com queries for additional UK names. A name was considered to be gendered if it was used at least 90% of the time by the same male or female gender. The sources used do not record nonbinary genders and there do not seem to be specific nonbinary first names, partly because people are named by their parents before they develop a gender identity. Twitter users were characterised as male or female if the first part of their username matched either the male or female name list, otherwise they were classed as unknown.

A minority of Twitter users specify in their biographies the pronouns that they prefer to be known by (e.g. “Pronouns: they/them”). They/them pronouns are often important aspects of identity for people with nonbinary genders (e.g. Bergman & Barker, 2017; Richards et al., 2016), although not all nonbinary people exclusively use they/them pronouns. This style was checked for in all user biographies, assigning additional male, female and nonbinary genders. If this conflicted with the name-based gender assignment, then the pronouns were taken as definitive. All cases of pronoun/first name disagreement were manually checked for accuracy and pronouns were always found to be reliable. The results for all the 169 nonbinary Twitter profiles were manually checked since any errors would have a greater impact on the results. All were found to be correct in the sense of consistent with the tweeter’s wider bio information.

The gender identification is sometimes wrong because: (a) nonbinary people not specifying pronouns may be assigned male or female if this is suggested by their first name, (b) males can have majority female names and females can have majority male names for a variety of reasons, including gender differences between cultures (e.g. Maria and Nicola can be Italian male names) and spelling mistakes (e.g. Lesley/Leslie), and (c) Twitter usernames may be for organisations starting with a gendered name (@MarieCurieSCO) or may not reflect the owner’s name (e.g. @NinaSimoneNo1Fan). Nevertheless, this method produces lists of users that have a high probability of having the specified gender identity. All errors in the lists would reduce the power of the method used in stage 2 but would not lead to false results because they would reduce average differences between sets of users. The final sample included 161,938 female, 247,380 male, and 169 nonbinary Twitter users with a UK location. Although the nonbinary sample is much smaller than the others, the statistical methods used in the next step take this into account so that the risk of identifying misleading differences is minimised.

3.2 Stage 2: Word association detection

Systematic differences between profiles were identified at the level of individual words for the word association detection (WAD) part of the WATA method used.



The word frequency approach of WAD identifies words that occur statistically significantly more often in the biographies of a given gender than in the remaining gendered biographies. For example, the word “grumpy” occurred more often in male biographies than in female biographies. Thus, males are more frequently bad tempered, more likely to consider bad-temperedness as part of their public identity or prefer this word to express their bad-temperedness. As this example illustrates, the words identified by the approach are fine-grained linguistic differences, reflecting an unknown combination of personal traits, linguistic style, and presentational preferences. This interpretational ambiguity is a practical limitation.

Before analysis profiles were processed to remove URLs and emojis and to convert plural words to singular (following English rules) to increase the statistical test power. Words were then identified as gendered if they (1) occurred in at least two profiles (to exclude typos) (2) occurred in a higher proportion of biographies from that gender than from other genders, and (3) passed a test of statistical significance. The statistical significance test used was a 2x2 chi-squared test, which is a standard approach to assess whether the difference between two proportions is statistically significant (Thelwall, 2021). This test was repeated for every word in every biography for a gender, generating a high chance of false positives due to multiple simultaneous tests. The Benjamini and Hochberg (1995) procedure was used to guard against this by increasing the significance threshold systematically to give a 99.9% chance that all words reflect underlying gender differences in preferences, if the test assumption is valid. The main test assumption that is questionable is independence: the tests assume that the likelihood of a word being used by one person is independent of its use by others in the same group, but there may be a degree of copying. This is commented on below when it seems relevant. Copying (or multiple profiles for the same person) is more likely to influence the nonbinary results because the smaller sample size makes each person more influential in the data.

This stage identified 1,474 words that are statistically significantly more used by UK males, females or nonbinaries, with a high degree of certainty. Nonbinary words were the rarest because of the smaller number of nonbinary users found. Female words were more common than male words despite females being a minority in the dataset, suggesting that females more likely to use general words that could meet statistical significance thresholds. Biography average lengths were similar (female: 8.6 words, nonbinary: 8.6 words; male: 8.3 words).

3.3 Stage 3: Word association contextualisation and thematic analysis

The 1,474 gendered words were clustered into groups using a modified thematic analysis (Braun & Clarke, 2013) to help identify broader patterns in the results. This was supported by an initial word association contextualisation step.



Each gendered word was manually assigned an initial theme to illustrate its typical use by the associated gender. This was a two-stage process. First, a sample of at least 10 descriptions was read (selected with a random number generator by Mozdeh) to identify the typical context of the word for its main gender user group. Second, the word was given a theme name generalizing this use, as in thematic analysis. The first stage is necessary because of polysemy and the multiple different possible contexts for a word, even when its meaning is clear. The following examples illustrate the process and the necessity for word sense and context disambiguation.

- “ig” (in the female list) usually occurred to link to an Instagram profile and was tagged “Social Media”. It was tagged “Social Media” because other social media terms were mentioned in a similar context.
- “Party” usually occurred in statements of political affiliations and was tagged “Politics” (even though in other contexts, the word could mean celebration).
- “Mitchell” always occurred in the context of employment (e.g. “clerk at Irwin Mitchell, solicitors”), and was tagged “Job”.
- “Heels” occurred in the context of shoes worn by the user (e.g. “barrister in high heels”), and was tagged “Appearance”.
- “Disabled” occurred mostly in job descriptions (rather than declaring a disability) and was therefore tagged “Job”.
- “Egg” occurred in both food preferences and self-evaluations (e.g. “all round good egg”) and was tagged “Multiple”.

As the last example illustrates, terms used in different contexts were classed “Multiple” rather than given additional themes. Two coders (ST and MT) carried out this stage, comparing the themes, checking discrepancies, removing themes judged incorrect, and adding extra themes to words when they represented different valid perspectives.

After the initial grouping of words into themes, the list was revisited multiple times to ensure that the themes were consistent and to recheck words that might fit into revised or new themes. Large themes were also rechecked to split them into smaller sub-themes (e.g. the Interests/Crafts subtheme was generated from the initial large Interests theme). This process continued until the results stabilised. Two coders (ST and MT) worked on this stage.

4 Results and discussion

The 1,474 gendered words identified were grouped into 33 themes. Each word reflects a fine-grained offline, presentational style or linguistic gender difference. Linguistically, the results also partly depend on polysemy, presence in idioms, or



common allusions/metaphors. For example, the term *heels* occurred partly because of the use of the phrase “in high heels” as an apparent metaphor for glamour or femininity within a job, rather than an interest in shoes.

Although experimental results are usually presented in numbers, graphs and tables (quantitative) or descriptive text (qualitative), here they are reported as plain lists of words because of the emphasis on fine-grained differences. The lists serve the dual functions of clarifying the theme name and showing each fine-grained difference. The methods combine singular and plural versions of words for testing and the singular versions are reported below but [s] or [es] is added when it is more intuitive to report the plural versions. The results are analyzed and compared to previous academic findings here rather than a separate Discussion section because of the wide variety of themes included.

The online supplement reports the percentage of male, female, and nonbinary descriptions using each word. There are too many terms to explain the context of each, but the online supplement gives information in cases where this is unclear and spells out acronyms when there is ambiguity (others can be looked up online). Twitter can also be searched for the terms for more context since its search matches descriptions as well as tweets. The original Twitter descriptions cannot be shared because of Twitter’s terms and conditions and privacy reasons. In many cases, terms were usually in lists, giving no extra context. For example, *Adidas* and *handbags* were often single words in lists, whereas *witch* was more often part of a phrase (e.g. “researcher witch that loves roleplay”).

4.1 Self-descriptions

Twitter users described multiple aspects of themselves in their biographies. Most themes include terms from at least two genders. The absence of nonbinary terms could be due to the smaller amount of data (fewer people in this category) or the absence of underlying differences. A complete list of themes is reported here since this study is attempting to identify all gender differences in Twitter bios.

The negative words for males may reflect a greater use of self-deprecating humour (Greengross & Miller, 2008). The extensive use of qualification terms for females is surprising given that status concerns are a male trait. It is possible that women work more in health fields (e.g. nursing) where explicitly stated qualifications are important, or that they feel a greater need to give evidence of competence in a sexist society.

The large number of gender identity words for nonbinary people is partly a side-effect of the detection method (specifying pronouns). The informal term *enby* is perhaps the equivalent of *bloke* or *gal* in this context. The nonbinary term *witch* suggests a partly female association and the word *nerd* has connotations of



obsessiveness that may relate to *autistic* in a different category. In the lists below standard font terms are for female, italic for male, and underlined for nonbinary.

- **Appearance:** *Adidas, shirt, bald, beard, bearded, beardy*, blonde, eyeliner, lipstick, redhead, shoes, heels, dress[es], handbags.
- **Self-Evaluation:** *average, bastard, cunt, curmudgeon, cynic, cynical, decent, failed, fanboy, fool, geek, git, grumpy, idiot, king, legend, tinkerer, nerd, witch, #GirlySwot, #QueenOf, arty, babe, badass, bitch, bubbly, chatterbox, crazy, curious, daydreamer, diva, dreamer, enthusiastic, fangirl, feisty, fierce, goddess, INFJ, introvert, mermaid, optimist, positive, queen, quirky, sassy, scatty, smiley, sparkle, swot, winging [it], witch.*
- **Employment/Qualification:** *alumnus, #WOW, alumna, alumni, award-winning, BA, BSC, BSL, grad, graduate, hons, MA, MBACP, MSc, newly [qualified], PGCE, PhD, prize, qualified, RCN, RGN, RMN, SCPHN, SEN.*
- **Gender identity:** *bloke, chap, dude, fella, gentleman, guy, lad, male, man, nonbinary, enby, genderfluid, non-binary, [non] binary, pronoun[s], non [binary], them, they, chick, female, gal, girl, girlie, girly, lady, lass, lassie, trans, woman.*
- **Geographic origins/location:** *born, Britain, British, Cornishman, Cymro, England, Englishman, exile, exiled, Irishman, Scotsman, UK, Welshman, Yorkshireman, Canadian, Cymrae, dweller, expat, Londoner, sea, seaside, Spanish, Yorkshirewoman.*

4.2 Illness

Females and nonbinaries seem more likely to share illness information in their Twitter descriptions. Females offline tend to be more willing to share personal information and are more likely to have disabling illnesses (Crimmins, Kim, & Solé-Auró, 2011; Gordon et al., 2017), despite living longer. This may partly explain the greater reporting of long-term serious conditions by females in Twitter biographies. Males may also hide illness because they view weakness as non-masculine (Jeffries & Grogan, 2012). Given that females on Twitter are more likely to value social support when they disclose a disease, compared to males (De Choudhury et al., 2017), there may be less reason for males to disclose their conditions.

Previous studies have sometimes suggested that there are higher illness rates amongst nonbinary people, attributing this to the additional stress of being marginalised (Lefevor et al., 2019). Nevertheless, evidence of health differences is inconclusive (Scandurra et al., 2019).



- **Illness:** *illness, severe*, #eds, #fibro, #spoonie, breast [cancer], coeliac, endo, endometriosis, fibro, fibromyalgia, MS, spoonie, survivor, warrior.
- **Illness/Mental Illness:** *autistic*, #actuallyautistic, #autism, #bpd, ADHD, anxiety, ASD, autism, syndrome.

4.3 Sentiment and support

There were gender differences in the words used to express opinions and only females disproportionately used supportive terms. Many of the evaluative terms could also reasonably have been classified as Self-Evaluation but the focus of the terms was on the interest rather than the person. For example, whilst *fanboy* seems to be a Self-Evaluation, the term *fan* was usually used in the expression “fan of ...”, emphasizing that the interest was liked. Both male and female evaluative terms sometimes contain aspects of self-evaluation (e.g. Males: “football nut”, “massive F1 fan”; females: “happiest when singing”, “lover of tea”). In contrast, only female evaluative terms were directly about the evaluated object (e.g. “love laughing”).

Females are expected to express emotions more and to be more compassionate (Eagly, 1987). In addition, females seem to express sentiment more directly in the social web (Thelwall, 2018). Thus, the gender differences reflect known patterns. The inclusion of some negative terms for males may reflect their greater tendency for self-deprecating humour (Greengross & Miller, 2008).

- **Evaluative:** *fan, fanatic, long-suffering, keen, massive, mighty, nut, sin[s], suffering*, addict, addicted, adorable, adore, amazing, beautiful, cute, fab, fabulous, fiend, gorgeous, happiest [when...], happy [when], love, lovely, lover [of...], loving, obsessed [with...].
- **Supportive:** #bekind, #kindness, advocate, caring, cheerleader, compassion, [make a] difference, happiness, kind, kinder, kindness, smile, positivity.

4.4 Interpersonal relationships

Some Twitter biographies described relationships or sexuality. Many relationship terms are gendered (e.g. sister, brother), so their presence in the list only shows that the terms are not rare. The parent relationship is interesting because the greater use by females (10.8% use a synonym of mother) than males (7.0% use a synonym of father) is probably a presentational difference rather than reflecting a demographic difference. There are substantially more mothers than fathers in the UK (UK men are 3 years older when they have children (ONS, 2019) and die 3.7 years earlier (ONS, 2020) so live 6.7 fewer years as a father. Of people in the UK that have children and are at least 13 (the Twitter minimum age), females spend 48% of their lives as mothers and males spend 42% as fathers. Based on this, and other factors



being equal, Twitter mothers could be expected to outnumber Twitter fathers by 1.15 to 1 (per female or per male) but biography-declared mothers outnumber biography-declared fathers by 1.54 to 1 (per female or per male).

The greater tendency for females to report relationships and sexuality reflects a greater tendency to share personal information on Twitter (Weathers et al., 2014). The corresponding tendency for nonbinary people does not seem to have been researched before.

- **Relationships:** *brother, grandad, granddad, grandfather, grandpa, husband, uncle, aunt, auntie, aunty, friend, girlfriend, godmother, grandchildren, gran, grandma, grandmother, granny, hubby, nan, nana, nanna, nannie niece, sister, teenage, teenager, widow, wife, wifey.*]
- **Relationships/Parent:** *#dad, dad, daddy, father, papa, son, #digitalmum, #mother, #mum, babies, breastfeeding, daughter, mam, mama, mamma, mammy, mat, maternal, maternity, mom, momma, mommy, mother, motherhood, mum, mumma, mummy, parent, pregnancy, stepmum, toddler, twin, two.*
- **Relationships/Sexuality:** *#gay, gay, [gay/bi] af [= as fuck], bi, bisexual, pan [sexual], polyamorous, queer, sex, [Thick Thigh] Thursday, nsfw, pansexual, bi, lesbian, queer.*

4.5 Beliefs

Biographies sometimes reported political or religious beliefs. Despite the possibility of astroturfing, there were no clearly fake profiles. Males are more likely to make their political affiliation public on social media (Bode, 2017), and less likely to discuss politics on Twitter (Koc-Michalska et al., 2019), but the results suggest that females are more willing to announce support for specific campaigns. The inclusion of several nonbinary terms perhaps reflects a greater politicisation of nonbinary people (e.g. Ndelu, Dlakavu, & Boswell, 2017), as a result of recent gender-related political conflicts about transgender rights, including around the term TERF (Trans-Exclusionary Radical Feminist) (Pearce, Erikainen, & Vincent, 2020; Weber, 2016).

Women have been historically more religious in the West than men (Miller & Stark, 2002), which aligns with *atheist* (men are more likely to be atheist) (Schnabel et al., 2016) and *humanist* being male-associated terms, but not *religion*. This latter term is used by males to mention the concept without claiming to be religious. Thus, the absence of religion-related terms for females and the presence of terms about religion for males are unexpected. It is possible that males are more willing to discuss religion abstractly.



- **Politics:** *#brexit, #jft96, @conservatives, @uklabour, anti, Brexit, Brexiteer, centrist, conservative, democracy, democrat, democratic, EU, Europe, European, left, liberal, libertarian, Marxist, nation, party, patriot, politics, political, politically, progressive, sceptic, socialism, tory, voted, labor, activism, activist, ally, campaigner, candidate, injustice, WASPI.*
- **Politics/Issues:** *anti-fascist, TERF, cruelty, eco, planet, #feminism, #feminist, feminism, feminist, patriarchy, #BlackLivesMatter, BLM.*
- **Religion:** *atheist, humanist, religion, spiritual.*

4.6 Social media style

Biographies sometimes linked to or mentioned other social media sites, positioning the profile as part of a networked online presence or declaring other interests beyond Twitter. The sites reflect some known gender differences in uptake, with females being more likely to use Instagram (Pew Research Center, 2019).

Some biographies contained disclaimers, such as “views/comments my own” or “retweets not endorsements”, explained who would be blocked, or finished with a kiss. Disclaimers and blocking threats seem to be most used by males. This would align with male tweets having a less personal focus so that content might offend others or express a controversial opinion.

- **Social Media:** *affiliate, Twitch, Twitter, YouTube, #blogger, #bookblogger, #socialmedia, blog, blogger, ig, Insta, Instagram, Pinterest.*
- **Social Media/Twitter etiquette:** *block, blocked, comment[s], endorsement[s], follow, necessarily, offend, opinion, reflect, represent, rt, own, view, x, xx, xxx.*

4.7 Food and drink

The home is the traditional female sphere, which may explain the greater female concern with food in profile descriptions. Women spend more time on food preparation (Tashiro & Lo, 2012), are still twice as likely as UK men to cook and shop for food (FSA, 2017) and showing an interest in food may act as a display of femininity (Rodney et al., 2017). The greater body image pressure on females may also be a contributing factor. The barbecue seems to be the traditionally most male cooking activity in some Western cultures (e.g. Molina, 2014). A male preference for spicy food has been found in some previous studies (Byrnes & Hayes, 2015; Spinelli et al., 2018) although not an association between curries and masculinity. The masculine curry association may be specifically British.

Declaring an alcohol preference may be a sociability signal in Twitter profiles. Gender differences in alcohol preferences are unsurprising since alcohol marketing is highly gendered (Atkinson et al., 2017). An early study of US students suggested



that males were more likely to declare alcohol consumption online (Peluchette & Karl, 2008).

- **Interests/Food and drink:** *BBQ, curry, #vegan, avocado, bake, baking, cake, cheese, chocoholic, chocolate, coffee, cook, cooking, cupcake, eating, feeding, food, foodie, gluten, nutrition, recipe, tea, vegan, vegetarian.*
- **Interests/Alcohol:** *ale, beer, brewer, Guinness, lager, malt, pub, whisky, cocktail, drinker, fizz, g&t, gin, prosecco, wine.*

4.8 Interests

Biographies often described one or more hobbies or non-work interests, whether to characterise the user's personality or announce likely Twitter topics to potential followers. Several groups of interests only have terms from one gender, but this is partly a side-effect of the narrow categories used. There are too many interests to comment on all, but a focus on crafts is a traditionally feminine trait because of its association with the home, females read for pleasure more (DoE, 2012) and women are more likely to be key decision makers for holidays (Mottiar & Quinn, 2004), whereas males have a greater interest in "things" (Su & Rounds, 2015), including machines, and gamble more (Gambling Commission, 2019).

The nonbinary interest in games and roleplay (including the combination) does not seem to have been noted before, although nonbinary issues have been discussed in the context of gaming (Jaroszewski et al., 2018). Roleplay may seem to be a natural interest to nonbinary people since they presumably lived the early part of their lives in a gender role that did not fit them. Cosplay (dressing up as a character) is a form of roleplay that supports both femininity and gender fluidity (Nichols, 2019).

Jobs reflect interests to some extent, although gender differences in job choices have been analysed extensively elsewhere (Cortes & Pan, 2018) and so are not discussed here. The results (see Appendix) confirm systematic employment gender differences, such as female domination of health-related careers (ONS, 2018).

- **Interests/Betting:** *bet, betting, gambler, gambling, poker, punter.*
- **Interests/Crafts:** *craft[s], crafter, crafting, crafty, crochet, crocheter, knit, knitter, knitting, quilter, sew, sewing, sewist, stitcher.*
- **Interests/Reading:** *[Jane] Austen, bookworm, Gryffindor, Harry [Potter], Hufflepuff, Jane, [Harry] Potter, Ravenclaw, reader, reading, Slytherin, [book] worm, yarn.*
- **Interests/Film or TV:** *film, jedi, movie, Disney, dramas, Outlander, [TV] soap.*
- **Interests/Machines:** *car, drone[s], motorbike, motorcycle, motorcyclist, petrol, petrolhead, rail, railway, ride, wheels.*



- **Interests/Music and dance:** *band, bass, bassist, Beatles, drummer, guitar, guitarist, hip, keyboard, metal, music, punk, rock, ska, techno, vinyl, ballet, [Gary] Barlow, choir, dancing, musical, sing, singing.*
- **Interests/Nature:** *birder, mountain, mountaineer, flower, garden, gardener, gardening, nature, wild, wildlife.*
- **Interests/Pets:** *#AdoptDontShop, cat, dog, fluffy, fur, guinea [pig], kitten, owned, pet, ponies, puppy, spaniel, animals.*
- **Interests/Travel:** *beach, holiday, travel, travelling, wanderlust.*
- **Interests/Videogames:** *FIFA, game[s], gamer, gaming, PC, PlayStation, PS4, streamer, video, videogame, Xbox, game[s], [Animal] crossing.*
- **Interests/Other:** *amateur, chess, classic, comic, freemason, gadget, joke, occasional, purveyor, retro, sci-fi, shooting, tech, anime, cosplay, cosplayer, d&d, drag, gender, TTRPG, @girlguiding, adventure, Brownies, cosplayer, fuelled, fun, girlguiding, hoarder, laugh, laughing, laughter, nap[s], shopaholic, shopper, shopping, sparkly, stationery, sunshine, tarot, unicorn[s], vintage, volunteer, volunteering.*

4.9 Physical activity

Male terms dominate the physical activity broad category, although the sport-related terms may refer mainly to spectator sports. Competitiveness is a male trait (e.g. Buser, Niederle, & Oosterbeek, 2014). The male habit of declaring sporting team affiliations aligns with a far greater interest in watching competitive sport for UK males (Abraham, 2020), but contrasts with German student social network users, from which women were more likely to declare group affiliations online, albeit social network groups rather than sporting groups (Haferkamp et al., 2012). Sports teams may also serve as identity markers for males. All the non-competitive physical activity words are female in the list. Fishing, swimming and cycling in the sports list can also be non-competitive physical activities, and two of these are male.

- **Sport:** *angler, basketball, bicycle, bike, biker, boxing, captain, carp, club, coach, cricket, cricketer, cycle, cycling, cyclist, darts, EV (motorsport), F1, FA (soccer), fisherman, fishing, football, footballer, footy, Formula, FPL, goalkeeper, golf, golfer, grassroots, ironman, league, level, [life-]long, marathon, martial, MMA, motorsport, NBA, NFL, non-league, PGA, racer, racing, referee, RFC, rugby, snooker, sport, sporting, sportsman, squash, stand, UEFA, umpire, wrestling, dressage, netball, netballer, swimmer, swimming.*
- **Non-competitive physical activity:** *#yoga, [fitness] bunny, fitness, Pilates, reiki, walking, yoga, yogi, yogini, Zumba.*
- **Sport/Supporter:** *#afc, ... Wolves, YNWA. [see Appendix for full list]*



4.10 Limitations

The results have many limitations. Gender identities are probably expressed differently in other social web sites and in other countries. Moreover, the differences found here for the UK may change over time. The first name heuristic used to identify gender will not work for people with names from minority cultures that are not in the list, as well as for people using fake or shortened names. The results are dependent on the variety of words available to express a meaning (e.g. soccer, football, footy), the specificity of the terms used (e.g. individual football team names, vs the sport name) and whether key terms have substantial polysemy. Thus, despite the systematic method used to identify gender differences, there may be gaps in the results. Additional gaps are also likely for nonbinary users because of the small non-binary sample size reduces the statistical power of the word association test used. The results are also affected by whether there are gendered versions of relevant words that could normally not be used by at least one gender. Thus, non-gendered words (e.g. parent) are more powerful indicators of underlying gender differences in the theme than gendered words (e.g. mum).

At a deeper level, word frequency differences may reflect underlying gender differences in interests, presentation styles or linguistic choice of word to represent a concept. The most likely causes of false positives (spurious gender differences) may be imitation of bio styles between users or multiple profiles for the same user.

5 Conclusions

The word comparison approach has revealed, for the first time, that there are many fine-grained gender differences in the Twitter profiles of male, female and nonbinary UK users. These differences were clustered into 33 distinct themes. In the case of males and females, most of the themes align with prior research and confirm that these traits transfer online. Other themes, such as the prominent reporting of qualifications by females, do not seem to have been noticed before in any context and could be an online-only gender expression or a previously unnoticed offline gender expression. For nonbinary Twitter users, the results are the first systematic evidence of gender expression differences. These show that, despite the many non-binary gender identities, there may be some nonbinary gender expression commonalities compared to males and females. The results also serve as list of fine-grained male, female and nonbinary gender expression differences related to interests and online presentations of the self. Of course, the lists are statistical associations and not in any sense prescriptive or ideal characteristics.

The information about gender-associated aspects of self-presentation in the social web may be useful for those that need to communicate online or analyse online



communication. For example, the lists may help organisations identify messages with unintended gender bias through addressing interests associated with one gender. They may also help generate gender-targeted messages when these are needed.

Author contributions

Mike Thelwall (m.thelwall@wlv.ac.uk): Conceptualization (Equal), Data curation (Equal), Formal analysis (Equal), Methodology (Equal), Software (Equal), Validation (Equal), Writing-original draft (Equal), Writing-review & editing (Equal); Saheeda Thelwall (s.thelwall@wlv.ac.uk): Data curation (Equal), Investigation (Equal), Methodology (Equal), Validation (Equal), Writing-review & editing (Equal); Ruth Fairclough (r.fairclough@wlv.ac.uk): Writing-original draft (Equal), Writing-review & editing (Equal).

References

- Abraham, T. (2020). YouGov research for International Women's Day shows how attitudes to men's and women's sport are changing [spreadsheet]. YouGov.com. <https://yougov.co.uk/topics/sport/articles-reports/2020/03/08/seven-ten-support-equal-coverage-womens-sport-not->
- Atkinson, A.M., Ross-Houle, K.M., Begley, E., & Sumnall, H. (2017). An exploration of alcohol advertising on social networking sites: An analysis of content, interactions and young people's perspectives. *Addiction Research & Theory*, 25(2), 91–102.
- APA (2015). Guidelines for psychological practice with transgender and gender nonconforming people. *American Psychologist*, 70(9), 832–864.
- Baym, N.K., & Boyd, D. (2012). Socially mediated publicness: An introduction. *Journal of Broadcasting & Electronic Media*, 56(3), 320–329.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, 57(1), 289–300.
- Bergman, S.B., & Barker, M.J. (2017). Non-binary activism. In *Genderqueer and non-binary genders* (pp. 31–51). Palgrave Macmillan, London.
- Bettany, S., Dobscha, S., O'Malley, L., & Prothero, A. (2010). Moving beyond binary opposition: Exploring the tapestry of gender in consumer research and marketing. *Marketing Theory*, 10(1), 3–28.
- Bode, L. (2017). Closing the gap: Gender parity in political engagement on social media. *Information, Communication & Society*, 20(4), 587–603.
- Bradford, N.J., Rider, G.N., Catalpa, J.M., Morrow, Q.J., Berg, D.R., Spencer, K.G., & McGuire, J.K. (2019). Creating gender: A thematic analysis of genderqueer narratives. *International Journal of Transgenderism*, 20(2–3), 155–168.
- Braun, V., & Clarke, V. (2013). *Successful qualitative research: A practical guide for beginners*. London, UK: Sage.
- Byrnes, N.K., & Hayes, J.E. (2015). Gender differences in the influence of personality traits on spicy food liking and intake. *Food Quality and preference*, 42, 12–19.



Research Paper

- Buser, T., Niederle, M., & Oosterbeek, H. (2014). Gender, competitiveness, and career choices. *The Quarterly Journal of Economics*, 129(3), 1409–1447.
- Buss, D.M. (1995). Psychological sex differences: Origins through sexual selection. *American Psychologist*, 50(3), 164–168.
- De Choudhury, M., Sharma, S.S., Logar, T., Eekhout, W., & Nielsen, R.C. (2017). Gender and cross-cultural differences in social media disclosures of mental illness. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (pp. 353–369).
- Cortes, P., & Pan, J. (2018). Occupation and gender. In: *The Oxford handbook of women and the economy*. OUP: Oxford, UK (pp. 425–452).
- Costa, P.T., Terracciano, A., & McCrae, R.R. (2001). Gender differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality and Social Psychology*, 81(2), 322.
- Crimmins, E.M., Kim, J.K., & Solé-Auró, A. (2011). Gender differences in health: Results from SHARE, ELSA and HRS. *European Journal of Public Health*, 21(1), 81–91.
- Diekmann, A.B., & Eagly, A.H. (2000). Stereotypes as dynamic constructs: Women and men of the past, present, and future. *Personality and Social Psychology Bulletin*, 26(10), 1171–1188.
- DoE (2012). Research evidence on reading for pleasure. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/284286/reading_for_pleasure.pdf
- Eagly, A.H., Nater, C., Miller, D.I., Kaufmann, M., & Sczesny, S. (2020). Gender stereotypes have changed: A cross-temporal meta-analysis of US public opinion polls from 1946 to 2018. *American Psychologist*, 75(3), 301–315.
- Eagly, A.H., & Wood, W. (2016). Social role theory of sex differences. In: *The Wiley Blackwell encyclopedia of gender and sexuality studies*. New York: Wiley. 1–3. <https://doi.org/10.1002/9781118663219.wbegss183>
- Eagly, A. (1987). *Sex differences in social behavior. A social role interpretation*. New Jersey: Hillsdale.
- FSA (2017). *The food and you survey: Wave 4*. Food Standards Agency. https://www.food.gov.uk/sites/default/files/media/document/food-and-you-w4-combined-report_0.pdf
- Gambling Commission (2019). *Gambling participation in 2018*. <https://www.gamblingcommission.gov.uk/PDF/survey-data/Gambling-participation-in-2018-behaviour-awareness-and-attitudes.pdf>
- Garg, N., Schiebinger, L., Jurafsky, D., & Zou, J. (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences*, 115(16), E3635–E3644.
- Goffman, E. (1959). *The presentation of self in everyday life*. London: Harmondsworth.
- Gordon, E.H., Peel, N.M., Samanta, M., Theou, O., Howlett, S.E., & Hubbard, R.E. (2017). Sex differences in frailty: A systematic review and meta-analysis. *Experimental Gerontology*, 89, 30–40.
- Greengross, G., & Miller, G.F. (2008). Dissing oneself versus dissing rivals: Effects of status, personality, and sex on the short-term and long-term attractiveness of self-deprecating and other-deprecating humor. *Evolutionary Psychology*, 6(3), 393–408.
- Grieve, R., March, E., & Watkinson, J. (2020). Inauthentic self-presentation on Facebook as a function of vulnerable narcissism and lower self-esteem. *Computers in Human Behavior*, 102, 144–150.



- Haferkamp, N., Eimler, S.C., Papadakis, A.M., & Kruck, J.V. (2012). Men are from Mars, women are from Venus? Examining gender differences in self-presentation on social networking sites. *Cyberpsychology, Behavior, and Social Networking*, 15(2), 91–98.
- HESA (2019). Higher Education Student Statistics: UK, 2017/18 – Qualifications achieved. <https://www.hesa.ac.uk/news/17-01-2019/sb252-higher-education-student-statistics/qualifications>
- Jaroszewski, S., Lottridge, D., Haimson, O.L., & Quehl, K. (2018). “Genderfluid” or “Attack Helicopter” Responsible HCI Research Practice with Non-binary Gender Variation in Online Communities. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. New York, NY: ACM Press (pp. 1–15).
- Jeffries, M., & Grogan, S. (2012). ‘Oh, I’m just, you know, a little bit weak because I’m going to the doctor’s’: Young men’s talk of self-referral to primary healthcare services. *Psychology & Health*, 27(8), 898–915.
- Keller, F.B., Schoch, D., Stier, S., & Yang, J. (2020). Political Astroturfing on Twitter: How to coordinate a disinformation Campaign. *Political Communication*, 37(2), 256–280.
- Koc-Michalska, K., Schiffrin, A., Lopez, A., Boulianne, S., & Bimber, B. (2019). From online political posting to mansplaining: The gender gap and social media in political discussion. *Social Science Computer Review*, 39(2), 197–210. <https://doi.org/10.1177/0894439319870259>.
- Kuo, F.Y., Tseng, C.Y., Tseng, F.C., & Lin, C.S. (2013). A study of social information control affordances and gender difference in Facebook self-presentation. *Cyberpsychology, Behavior, and Social Networking*, 16(9), 635–644.
- Lebel, K., & Danylchuk, K. (2012). How tweet it is: A gendered analysis of professional tennis players’ self-presentation on Twitter. *International Journal of Sports Communication*, 5, 461–480.
- Lefevor, G.T., Boyd-Rogers, C.C., Sprague, B.M., & Janis, R.A. (2019). Health disparities between genderqueer, transgender, and cisgender individuals: An extension of minority stress theory. *Journal of Counseling Psychology*, 66(4), 385.
- Lippa, R. (1998). Gender-related individual differences and the structure of vocational interests: The importance of the people-things dimension. *Journal of Personality and Social Psychology*, 74(4), 996–1009. <https://doi.org/10.1037/0022-3514.74.4.996>
- Marwick, A.E., & Boyd, D. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, 13(1), 114–133.
- Matsumo, E., & Budge, S.L. (2017). Non-binary/genderqueer identities: A critical review of the literature. *Current Sexual Health Reports*, 9(3), 116–120.
- Mackay, F. (2019, July). Always endangered, never extinct: Exploring contemporary butch lesbian identity in the UK. In *Women’s Studies International Forum*, 75, paper 102241. <https://doi.org/10.1016/j.wsif.2019.102241>
- McGlashan, H., & Fitzpatrick, K. (2017). LGBTQ youth activism and school: Challenging sexuality and gender norms. *Health Education*, 117(5), 485–497. <https://doi.org/10.1108/HE-10-2016-0053>
- Mehdizadeh, S. (2010). Self-presentation 2.0: Narcissism and self-esteem on Facebook. *Cyberpsychology, Behavior, and Social Networking*, 13(4), 357–364.



Research Paper

- Miller, A.S., & Stark, R. (2002). Gender and religiousness: Can socialization explanations be saved? *American Journal of Sociology*, 107(6), 1399–1423.
- Miltersen, E.H. (2018). De, den, hen, and the rest. *Journal of Language Works-Sprogvidenskabeligt Studentertidsskrift*, 3(1), 31–42.
- Molina, H. (2014). The construction of South Texas masculinity: Masculine space, the pico de gallo and the barbeque grill. *Identities*, 21(2), 233–248.
- Mottiar, Z., & Quinn, D. (2004). Couple dynamics in household tourism decision making: Women as the gatekeepers? *Journal of Vacation Marketing*, 10(2), 149–160.
- Nadal, K.L., Whitman, C.N., Davis, L.S., Erazo, T., & Davidoff, K.C. (2016). Microaggressions toward lesbian, gay, bisexual, transgender, queer, and genderqueer people: A review of the literature. *The Journal of Sex Research*, 53(4–5), 488–508.
- Ndelu, S., Dlakavu, S., & Boswell, B. (2017). Womxn’s and nonbinary activists’ contribution to the RhodesMustFall and FeesMustFall student movements: 2015 and 2016. *Agenda: Empowering Women for Gender Equity*, 31(3–4), 1–4.
- Nichols, E.G. (2019). Playing with identity: Gender, performance and feminine agency in cosplay. *Continuum*, 33(2), 270–282.
- Nosko, A., Wood, E., & Molema, S. (2010). All about me: Disclosure in online social networking profiles: The case of Facebook. *Computers in Human Behavior*, 26, 406–418.
- Peluchette, J., & Karl, K. (2008). Social networking profiles: An examination of student attitudes regarding use and appropriateness of content. *CyberPsychology & Behavior*, 11(1), 95–97.
- Oberst, U., Renau, V., Chamarro, A., & Carbonell, X. (2016). Gender stereotypes in Facebook profiles: Are women more female online? *Computers in Human Behavior*, 60, 559–564.
- ONS (2018). EMP04: Employment by occupation (2018 spreadsheet). <https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/datasets/employmentbyoccupationemp04>
- ONS (2019). Birth characteristics in England and Wales: 2018. <https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/bulletins/birthcharacteristic sinenglandandwales/2018>
- ONS (2020). Dataset: Life expectancy at birth and selected older ages. <https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/datasets/lifeexpectancyat birthandselectedolderages>
- Pearce, R., Erikainen, S., & Vincent, B. (2020). Afterword: TERF wars in the time of COVID-19. *The Sociological Review*, 68(4), 882–888.
- Pew Research Center (2019). Social Media Fact Sheet <https://www.pewresearch.org/internet/fact-sheet/social-media/>
- Pitcan, M., Marwick, A.E., & Boyd, D. (2018). Performing a vanilla self: Respectability politics, social class, and the digital world. *Journal of Computer-Mediated Communication*, 23(3), 163–179.
- Priante, A., Hiemstra, D., Van Den Broek, T., Saeed, A., Ehrenhard, M., & Need, A. (2016). #WhoAmI in 160 characters? Classifying social identities based on twitter profile descriptions. In *Proceedings of the First Workshop on NLP and Computational Social Science* (pp. 55–65).
- Pullen Sansfaçon, A., Medico, D., Suerich-Gulick, F., & Temple Newhook, J. (2020). “I knew that I wasn’t cis, I knew that, but I didn’t know exactly”: Gender identity development, expression



- and affirmation in youth who access gender affirming medical care. *International Journal of Transgender Health*, 21(3), 307–320. <https://doi.org/10.1080/26895269.2020.1756551>
- Richards, C., Bouman, W.P., Seal, L., Barker, M.J., Nieder, T.O., & T’Sjoen, G. (2016). Non-binary or genderqueer genders. *International Review of Psychiatry*, 28(1), 95–102.
- Rieger, G., Linsenmeier, J.A., Gygax, L., Garcia, S., & Bailey, J.M. (2010). Dissecting “gaydar”: Accuracy and the role of masculinity–femininity. *Archives of Sexual Behavior*, 39(1), 124–140.
- Rodney, A., Cappeliez, S., Oleschuk, M., & Johnston, J. (2017). The online domestic goddess: An analysis of food blog femininities. *Food, Culture & Society*, 20(4), 685–707.
- Scandurra, C., Mezza, F., Maldonato, N.M., Bottone, M., Bochicchio, V., Valerio, P., & Vitelli, R. (2019). Health of non-binary and genderqueer people: A systematic review. *Frontiers in Psychology*, 10, 1453. <https://doi.org/10.3389/fpsyg.2019.01453>
- Schnabel, L., Facciani, M., Sincoff-Yedid, A., & Fazzino, L. (2016). Gender and atheism: Paradoxes, contradictions, and an agenda for future research. In: *Annual Review of the Sociology of Religion*. Leiden, NL: Brill BV (pp. 75–97).
- Schwartz, H.A., Eichstaedt, J.C., Kern, M.L., Dziurzynski, L., Ramones, S.M., Agrawal, M., ... & Ungar, L.H. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLoS One*, 8(9), e73791.
- Spinelli, S., De Toffoli, A., Dinnella, C., Laureati, M., Pagliarini, E., Bendini, A., & Gasperi, F. (2018). Personality traits and gender influence liking and choice of food pungency. *Food Quality and Preference*, 66, 113–126.
- Su, R., & Rounds, J. (2015). All STEM fields are not created equal: People and things interests explain gender disparities across STEM fields. *Frontiers in Psychology*, 6, 1–20. <https://doi.org/10.3389/fpsyg.2015.00189>
- Tashiro, S., & Lo, C.P. (2012). Gender difference in the allocation of time: Preparing food cooked at home versus purchasing prepared food. *Food, Culture & Society*, 15(3), 455–471.
- Thelwall, M., Bailey, C., Tobin, C., & Bradshaw, N.A. (2019). Gender differences in research areas, methods and topics: Can people and thing orientations explain the results? *Journal of Informetrics*, 13(1), 149–169.
- Thelwall, M., Makita, M., Mas-Bleda, A., & Stuart, E. (2021). “My ADHD hellbrain”: A Twitter data science perspective on a behavioural disorder. *Journal of Data and Information Science*, 6(1), 13–34. <https://doi.org/10.2478/jdis-2021-0007>
- Thelwall, M. (2018). Gender bias in sentiment analysis. *Online Information Review*, 42(1), 45–57.
- Thelwall, M. (2021). *Word association thematic analysis: A social media text exploration strategy*. New York, NY: Morgan & Claypool.
- Tifferet, S., & Vilnai-Yavetz, I. (2014). Gender differences in Facebook self-presentation: An international randomized study. *Computers in Human Behavior*, 35, 388–399.
- Twenge, J.M. (1997). Changes in masculine and feminine traits over time: A meta-analysis. *Sex roles*, 36(5–6), 305–325.
- Walton, S.C., & Rice, R.E. (2013). Mediated disclosure on Twitter: The roles of gender and identity in boundary impermeability, valence, disclosure, and stage. *Computers in Human Behavior*, 29(4), 1465–1474.
- Weathers, M., Sanderson, J., Matthey, P., Grevious, A., Warren, S., & Tehan, M. (2014). The tweet life of Erin and Kirk: A gendered analysis of professional sports broadcasters’ self-presentation on Twitter. *Journal of Sports Media*, 9(2), 1–24.



Research Paper

- Weber, S. (2016). "Womanhood does not reside in documentation": Queer and feminist student activism for transgender women's inclusion at women's colleges. *Journal of Lesbian Studies*, 20(1), 29–45.
- WHO (2020). Gender and genetics. <https://www.who.int/genomics/gender/en/index1.html>
- Witt, M.G., & Wood, W. (2010). Self-regulation of gendered behavior in everyday life. *Sex Roles*, 62(9–10), 635–646.
- Wolf, A. (2000). Emotional expression online: Gender differences in emoticon use. *CyberPsychology & Behavior*, 3(5), 827–833.
- Yau, J.C., & Reich, S.M. (2019). "It's Just a Lot of Work": Adolescents' Self-Presentation Norms and Practices on Facebook and Instagram. *Journal of Research on Adolescence*, 29(1), 196–209.
- Zucker, K.J. (2015). The DSM-5 diagnostic criteria for gender dysphoria. In *Management of gender dysphoria* (pp. 33–37). Springer, Milan.



This is an open access article licensed under the Creative Commons Attribution-NonCommercial-NoDerivs License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).



Appendix: Other terms

- **Employment (87/425 female and 1/202 male terms (surgeon) are health-related; 0/425 F and 33/202 M are computing-related):** #bitcoin, accountant, actor, advice, aerospace, aircraft, analyst, analytic, android, app, architect, army, asset, astronomer, automation, automotive, aviation, azure, blockchain, broker, build, builder, building, bus, business, businessman, cameraman, carpenter, CEO, chairman, chartered, chef, civil, cloud, code, coder, comedian, commentator, commercial, company, composer, computer, computing, conductor, construction, contractor, contributor, crypto, CTO, cyber, data, database, defence, designer, dev, developer, devop, director, DJ, driver, driving, drum, economic, economist, electrical, electrician, electronic, employer, energy, engine, engineer, engineering, entrepreneur, equipment, Esport, estate, evangelist, ex, filmmaker, finance, financial, fire, firefighter, firm, fleet, former, fx, gas, graphic, greenkeeper, groundsman, guru, hacker, hardware, heating, HGV, hire, host, industrial, industry, infosec, infrastructure, insurance, internet, investing, investment, investor, iOS, IOT, JavaScript, lighting, Linux, logistic, machine, magician, maintenance, managing, manufacturing, market, match, md, mechanic, mechanical, merchant, Microsoft, military, mobile, money, motoring, musician, MVP, navy, net, official, operator, pastor, philosopher, philosophy, photographer, php, physic, physicist, pilot, plumbing, podcast, podcaster, police, politician, postman, premier, product, professor, programmer, programming, property, radio, RAF, record, renewable, retired, royal, sale, salesman, security, server, serving, software, soldier, solution, songwriter, sound, source, SQL, stack, station, stock, surgeon, surveyor, system, tax, techie, technical, technologist, technology, telecom, tenor, test, tour, trade, trader, trading, train, transport, truck, union, vehicle, venture, veteran, videographer, web, wind, windows, commission, postgrad, #actress, #comms, #diversity, #equality, #hellomynameis, #inclusion, #mentalhealth, #nhs, #pr, #sbs, #wellbeing, #women, #womeninstem, #womenintech, @graziauk, @irwinmitchell, @kingsioppn, @lshtm, @metrouk, @stylistmagazine, AAC, about, accessories, actress, acute, administrator, adolescent, advanced, AHP, alto, ambassador, ANP, anthropologist, anthropology, aromatherapy, art, artist, assistant, associate, attachment, awareness, baker, BDA, beauty, behaviour, bereavement, bestselling, biomedical, board, book, bridal, bride, businesswoman, CAMH, campaign, cancer, care, career, carer, celebrant, centred, championing, change, charity, childcare, childhood, choreographer, chronic, classes, clinical, clothe, CN, cohort, colour, comm, commissioning, committed, committee, communication, communities, community, complementary, contemporary,



coordinator, co-ordinator, copywriter, costume, counselling, counsellor, creative, creativity, criminology, critical, cultural, culture, curator, CYP, dance, dancer, debut, dementia, deputy, development, diabetes, diagnostic, dietetic, dietitian, disability, disabled, disease, disorder, diversity, doctoral, domestic, doula, drama, drawing, dysphagia, ed, education, educational, educator, emotional, empowering, empowerment, engagement, English, equality, equestrian, Etsy, event, examiner, exploring, EYF, fabric, facilitator, families, fashion, feature, fellow, fiction, Florence, florist, forensic, foundation, frailty, freelance, French, funded, fundraiser, fundraising, gender, gift, hair, handmade, head, healer, health, healthcare, healthy, HEE, heritage, historical, holistic, hospice, hospital, housewife, HR, hub, human, HVS, hypnotherapist, ICU, illness, illustration, illustrator, impact, improvement, improving, inclusion, inclusive, inequalities, inequality, infant, infection, inpatient, inspire, inspired, inspiring, integrated, interested, interior, interpreter, intersectional, intervention, intuitive, involvement, Irwin, jewellery, journalist, journo, justice, ks2, ks3, language, law, lead, learning, lecturer, librarian, libraries, library, lifestyle, lingerie, linguist, linguistic, lit, literacy, literary, literature, Macmillan, maid, makeup, make-up, making, manager, Marie, marketer, marketing, massage, matron, medical, medicine, meditation, menopausal, menopause, mental, mentee, mentor, mezzo, MFL, MH, midwife, midwifery, midwives, migration, mindfulness, mistress, Mitchell, Mslexia, MUA, museum, nail, nanny, natural, neonatal, neuro, neuroscience, nh, nightingale, NIHR, novel, NQT, NUH, nurse, nursery, nursing, nutritional, nutritionist, occupational, od, officer, oncology, OT, OU, outdoor, outreach, PA, paed, paediatric, pain, painting, palliative, paralegal, parenting, participation, partnership, passion, passionate, patient, pattern, performer, perinatal, pharmacist, pharmacy, physio, physiotherapist, physiotherapy, PICU, plant, poet, poetry, policy, postnatal, PR, practice, practitioner, prev, prevention, previous, primary, printmaker, programme, project, proofreader, PSHE, psych, psychologist, psychology, psychotherapist, public, publishing, QI, qualitative, quality, radiographer, raising, rape, receptionist, recovery, reducing, reflexologist, refugee, registered, rehab, rehabilitation, relationship, rep, rep'd, reproductive, research, researcher, researching, respiratory, reviewer, RNA, romance, romantic, safeguarding, SCBWI, scholar, school, seamstress, secondary, SENCO, senior, sex, sexual, shop, singer, skill, skincare, SLE, slimming, SLT, social, sociology, solicitor, soprano, special, specialising, specialist, speech, staff, stem, stitch, stroke, student, studies, studying, style, stylist, support, supporting, sustainability, teacher, teaching, teen, textile, theatre, therapies, therapist, therapy, thrive, trained, trainee, training, translation, translator, trauma, trust, trustee, tutor,



UHB, uni, university, veterinary, viability, violence, violinist, visitor, welfare, wellbeing, well-being, wellness, widening, women, worker, workforce, working, workshop, writer, writing, young.

- **Gendered pronoun:** *he, him, his, her, herself, she.*
- **Sport/Supporter:** *#afc, #arsenal, #avfc, #bhafc, #coyg, #coys, #itfc, #lcf, #lfc, #lufc, #mcf, #mufc, #nffc, #nufc, #saintsfc, #thfc, #ynwa, @arsenal, @lfc, @lufc, @manutd, @spursofficial, AFC, Albion, argyle, Arsenal, Aston, athletic, AVFC, Blackburn, blade, blue, boro, Brentford, CC, CCC, Celtic, Charlton, Chelsea, City, County, COYS, CPFC, Everton, Evertonian, Falcons, FC, Forest, Fulham, gooner, Ham, Hibs, holder, Hotspur, hull, Ipswich, Leeds, Leicester, Leyton, LFC, Liverpool, LUFc, Manchester, MCFC, Miami, Millwall, MUFC, Newcastle, NFFC, Notts, NUFC, Orient, palace, pompey, QPR, Rams, Rangers, Rovers, SAFC, Saints, season, Spurs, St, Sth, Stoke, Sunderland, supporter, SWFC, THFC, ticket, Tottenham, Town, UFC, United, Utd, Villa, Watford, Wednesday, West, Wolves, YNWA.*
- **Multiple:** *again, agree, any, away, back, bad, been, bio, but, die, do, don't, egg, electric, elite, else, etc, few, fly, follower, force, from, general, giant, glory, god, here, hit, if, is, it, I've, know, last, lifelong, like, lord, ltd, mainly, may, mod, motor, not, nothing, now, occasionally, odd, old, on, once, only, or, order, other, over, play, played, player, playing, poor, problem, real, regular, road, round, served, since, some, someone, Steve, stuff, than, that, the, then, these, think, those, till, time, used, war, was, we, white, chaotic, mod, abuse, adult, all, always, at, baby, be, being, body, bunnies, busy, by, child, children, chronically, colourful, currently, drag, early, fairy, glass, glitter, healing, heart, in, juggling, ld, life, little, live, living, make, miss, mixed, my, navigating, our, people, person, princess, rainbow, right, Sarah, send, sharing, slightly, small, story, ta, thing, tiny, to, tree, with, word, pro.*

