Social Media and Son Preference: Evidence from India

Draft intended for the Gender in the Economy: Women in the Digital World Conference
Please do not circulate or quote

By Praachi Kumar*[1,2] and Bruno Martorano[1,2]

[1] UNU-MERIT, United Nations University
[2] Maastricht University School of Business and Economics

Abstract

This research investigates the impact of exposure to social media platform Twitter on son-biased fertility preferences in India. We apply an instrumental variables strategy based on a popular national Twitter campaign attributed to the retirement of Indian cricketer, Sachin Tendulkar. We find that exposure to Twitter supports less discriminatory preferences for the sex of the child, as well as weaker son preference for women in India. Our findings show that hashtag activism related to #SelfieWithDaughter, an online girls' rights campaign, encourages a more potent impact on women's preferences, decreasing discrimination. This paper outlines the contribution of new-media technology and an enthusiastically adopted campaign on gender norms in a country marked by a booming population and enduring gender inequities, during a technology transition.

Keywords: Son Preference, Social Media, Gender Norms, Technology Diffusion, Child Rights, Campaigns
JEL Codes: J16, J13, O33, O12, L82

* E-mail: kumar@merit.unu.edu

1. Introduction

Son preference is one of the most severe and pervasive forms of gender-based discrimination in the world today. In India, largely regarded as the country of missing women (Sen 1992, Agnihotri *et al.,* 2002), fewer girl children are being born and surviving, and many face relentless neglect afflicting not only their health, nutrition and development, but also their dignity and identity. For instance, daughters are being given names which mean 'Unwanted', 'The Last One', or 'Missed Call', simply because a son was desired where a daughter was born (See Deshpande & Gupta, 2019). With the advent of prenatal sex determination testing, the harmful culture of discrimination against daughters has resulted in an adverse sex ratio and the stark reality of millions of 'missing daughters'[2], because they were discriminated against before they were even born[3] (Kabeer et al., 2014; see Sen, 1992). While sex-determination has been criminalised, the persistence of an adverse sex-ratio at birth illustrates that gender norms which embolden son preference are tenacious[4]. New forms of media may help to challenging these harmful cultural norms by facilitating horizontal patterns of communication and connection and steering a paradigm shift in one's worldview through new information networks (See Castells, 2009). Yet, the influence of a national technology transition, which has ushered in contemporary social networking platforms on son preference is not clearly understood, and more research is needed.

In this paper we document how social media influences gender norms in the context of India, placing emphasis on one particular mechanism of influence, endorsements. The potential role of traditional media in shaping gender norms has been extensively studied (La Ferrara *et al.,* 2012; Kearney & Levine, 2015; Arias, 2019; Pesando, 2022; Rajkhowa & Qaim, 2022). For example, media and its impact on son preference for the Indian setting is explored in Jensen & Oster (2009), identifying that among a set of indicators for women's status and gender-based discrimination, cable television diffusion led to substantial decline in preference for sons soon after introduction. Its premise is that new modes and types of information can introduce changes in gender attitudes in rural India, pointing to outside exposure as the mechanism responsible for rapid changes in beliefs. However, Iversen & Palmer-Jones (2019) partly replicate this analysis to find that for illiterate women, the effect of cable television interacted with son preference is not significant. Pande & Astone (2007) note that while mothers' education is the predominant driver for low son preference, there also exists an association

[2]Whereas the problem of missing women in India emerges from a culture of neglect towards women and girls, resulting in an adverse sex ratio (Sen, 1992). In addition to high maternal mortality, the prime cause of missing women is comparative childhood neglect in healthcare and nutrition, which jeopardise their collective survival e.g., Jayachandran & Pande (2017) attribute the short height of Indian children to be a result of favouritism towards the eldest son and birth order effects, leading to health inequalities among siblings.

[3] https://voxdev.org/topic/health/ten-facts-about-son-preference-india

[4]The Pre-Conception and Pre-Natal Diagnostics Testing (PCPNDT) Act (1994) which outlaws the detection of the sex of the child at any point before or during the pregnancy in an attempt to arrest sex-selection i.e., termination of pregnancy due to the predicted sex of the foetus, one of the leading causes of missing women (See Ahankari et al, 2015) with harsh penalties to violating practitioners, diagnostic centres and parents (Bhalotra & Cochrane, 2010; Patel et al., 2013; Jha et al., 2011). Challenges in the implementation of this act mean that the sex ratio at birth has not yet returned to its natural rate, despite declines in total fertility and improvements in both sex ratio at birth and survival for girls (Tabaie, 2017). Thus, the issue of missing women persists despite progress in offsetting the imbalanced sex-ratio at birth.

between media exposure, cinema and weaker son preference implying an influence of progressive ideas.

The effect of contemporary media on son preference is unclear. For instance, social media platforms can offer spaces where discussions may take place virtually and outside the traditional confines of the household and community[5]. Such exposure can not only introduce new ideas, but also engage users to participate in discussions taking places on the web. The persistence of a social institution, in this case harmful gender norms, is tied to not only social networks but also the representation of gender issues in discourse[6]. Social media has charted new networks and shaped how connected individuals interact with virtual worlds, leading to the possibility that the introduction of new ideas challenges the old.

A national technology transition has led 820 million Indians to the internet (Internet and Mobile Association of India, 2023). In the last decade, digital technology has diffused at a rapid pace, with more Indians owning mobile devices than toilets at the beginning of the decade (Ramani, 2016), cultivating a new era marked by virtual worlds and networks. Social media has shaped how people consume and produce information. Operators of social media platforms have turned to India to increase their user base. In 2014, India was considered as the fastest growing market for the social media platform, Twitter (formerly, renamed 'X' and henceforth retained as Twitter) (The Times of India, 2014). According to former CEO of Twitter, Dick Costello, Twitter is an 'information network' which complements traditional media with the 'widest possible coverage in the world' (Khanna, 2014). For women, social media has cast an influence beyond its role as a network[7]. However, for the Indian context where social and gender norms have a substantial power over women's rights and capabilities, the impact of social media still remains to be investigated.

The first aim of this paper is, therefore, to investigate how the exposure to Twitter, one of the most popular social media networks in the country, influences women's preferences for the sex of their child. For this purpose, we use a novel dataset combining information on preference for the sex of the child extracted from Demographic and Health Survey (DHS) with georeferenced data from Twitter's Academic Track API in 2014 and 2019.

[5] Through a virtual ethnography study, Pearce et al. (2022) find that Facebook groups can offer spaces for women in Azerbaijan to discuss son preference and sex selection, with results suggestion a greater degree of conformation than deviation from gender norms.

[6] For instance, Michalopoulos & Xue (2021) find that folklore influenced how gender norms were diffused and enforced over generations.

[7] Scholars have extensively explored the impact of social media on various economic and political outcomes (See Zhuravskaya et al. (2020) for a discussion on the political effects). Other empirical research on social media which focusses on gender relations includes Ongo Nkoa et al. (2023), covering 45 African countries and finds that social media strengthens women's political empowerment. The links between gender-based activism and social media are also distinctively witnessed through mass global movements which address violence against women and sexual misconduct in the workplace (#MeToo and #TimesUp), Femicide (NiUnaMenos), etc. Theoretically, Cheng & Siawh (2022) provide a basis to show how movements such as #metoo can have both intended and unintended effects. Battisti et al. (2022) show that Twitter influences crime rates related to gender-based violence in the United States of America, leading to their reduction and possibly by changing social norms.

A methodological concern in our analysis is that the exposure to Twitter was not random across districts. As a consequence, we might fail to account for unobserved variables that matter for explaining the observed variation in change in social norms, or those which jointly determine the number of Tweets in a specific locality and individuals' inclination towards revising their adherence to norms. To address potential endogeneity, we instrument exposure to Twitter considering the path dependent nature of technology diffusion taking inspirations from Müller & Schwarz (2023). Specifically, we focus on a campaign which brought the social media platform to the general public's eye across the country. In November 2013, India's 'god of cricket', Sachin Tendulkar retired and played his final match which was televised throughout the nation and was regarded as the highest rated match on Indian television in eight years. In honour of the retiring star last game, the Board of Control for Cricket in India (BCCI) introduced a Twitter campaign '#ThankYouSachin'. Throughout the game, the campaign was regularly advertised alongside the match score bringing national attention to Twitter as a platform for celebrating Sachin Tendulkar's cricket career. We exploit interest towards the cricketer at the height of the campaign and instrument Twitter use with an indicator of Sachin Tendulkar's popularity during the campaign using Google Tends data. Since Google Trends reports search terms indexed to total volume of searches, we construct this instrument as a dummy for districts indicating if interest in the search term 'Sachin Tendulkar' is non-zero on the index in any locality in the district during the final. We run multiple tests to probe our instrument and several tests, such as alternative definitions of the Twitter variable, to confirm the validity of our main results. All of these tests confirm the robustness of our results.

We find that social media exposure is significantly and directly linked to lower son preference. Moreover, with greater exposure to Twitter, women are less likely to discriminate on the sex of their child. The implicit hypothesis is that changes in social norms are due to women and their households' exposure to new networks and their introduction to the information society, for instance, through the observation of contemporary gender dynamics through Twitter. In this perspective, the content or narratives which women and their households are exposed to can guide their preferences in one direction or another.[8] Some studies indeed emphasize the significance of marketing techniques to improve societal and development outcomes (Chandy *et al.* 2021; Davis *et al.* 2021; Kotler & Zaltman 1971). Son preference itself was a feature of a viral campaign popularised by then recently elected Prime Minister Narendra Modi on Twitter in 2015. The campaign '#SelfieWithDaughter' was spearheaded by Sunil Jaglan, girl's rights activist, and the elected decision maker or sarpanch of Bibipur village in district Jind in Haryana, a state with a notably skewed child sex ratio at birth against girls[9]. Using Twitter's feature for posting pictures, users were encouraged to use the hashtag with a picture of themselves and their daughter. Our results show that the impact of social media exposure was

[8] However, there exists a fundamental difference between the nature of the production of gendered language on Twitter and campaign-driven narratives on their ties to preferences, that is those who define the features of an issue, who participates and how it pulsates online as substantiated by contrary outcomes to media exposure. For instance, new media has been shown to contribute to discrimination and hate, Bursztyn et al., (2019) study social media's contribution to Xenophobia; Traditional media, such as antisemitic propaganda on radio influencing those predisposed to such views (Adena et al., 2015) and film, through Ang (2023) demonstrating how a popular American film in 1915 which included racists tropes, continues to explain hate crimes today.

[9] See Ahlawat (2009) for a discussion on Haryana's skewed child sex ratio, Yasir (2023) for a profile on the movement

particularly effective in shaping preferences in districts where the #SelfieWithDaughter campaign had take-up. To explore the nature and salience of online narratives on children, we develop a classification model by fine tuning XLM-RoBERTa, a transformers architecture developed by FacebookAI to identify from the universe of Tweets in India, those which relate to children and adolescents. We find that discussions related to children have increased marginally between 2015 and 2018. Through a thematic analysis of 500 Tweets, we find that discussions related children and adolescents are thematically varied. Tweets with a neutral and progressive stance far outweigh regressive discourse around children and adolescents online.

Our paper contributes to various strands of literature. First, we contribute to the literature investigating the potential role of media technology in shaping gender norms (La Ferrara et al., 2012; Kearney & Levine, 2015; Arias, 2019; Battisti, 2022; Pesando, 2022; Rajkhowa & Qaim, 2022). Jensen & Oster (2009) specifically investigate the impact of television on son preferences. We advance this literature focusing on the role on social media in challenging harmful gender norms. To our knowledge, this is the first study investigating the impact of exposure to Twitter on women's preferences for the sex of their child. Second, investigating the effects of the campaign '#SelfieWithDaughter' on child sex preferences, we also contribute to the literature analysing campaigns to address societal issues (Chandy *et al.* 2021; Davis *et al*. 2021; Kotler & Zaltman 1971). In a recent contribution, for example, Donati, Orozco-Olvera & Rao (2022) experimentally show social media content might be very effective in influencing gender norms and reducing social acceptability of violence against women in India, even in marginalized communities. We contribute to this emerging literature focusing on a real and popular social media movement called #SelfieWithDaughter which was introduced to Twitter in 2015 with the objective of demonstrating that daughters are valued in the country and to reduce son preference. Another feature of this campaign is that it was endorsed in 2015 by the Prime Minister of India, Narendra Modi. By extending our research to hashtag activism, therefore, this analysis also contributes to the literature on endorsements (see Alatas et al., 2019).[10]

The paper is organised as five sections following this overview, Section 2 presents the data sources and methodology. Section 3 provides the main results, introduces the instrument and the results which follow. Section 4 focuses on the mechanisms of impact and heterogeneous effects. Section 5 validates the main analyses through robustness tests, the last section concludes.

2. Data and Methodology

This paper investigates the impact of exposure to the social media platform, Twitter, on son-biased fertility preferences held by women in India over two periods, 2014-15 and 2019-21. For this purpose, we create a unique dataset using three independent sources: the India

[10] Alatas et al. (2019) show that celebrity endorsements over Twitter for a public health campaign in Indonesia and their influence on changing behaviours related to seeking immunisation.

Demographic and Household Survey (DHS)[11], data from Twitter's academic track application programming interface (API) and administrative data from the 2001 Census of India. The unit of analysis is individual women. In order to account for potential changes in the shape and size of a district over the period of analysis, we use the 2015 fixed district boundaries.

2.1 Data on Twitter

To measure the exposure to social media, we use Twitter data. Tweets are preferred over other geo-data, such as self-reported user location, because we are interested in the diffusion of this social media platform and not in the adoption of its membership. Diffusion is measured using the action of tweeting through which we can observe production of content on Twitter and its penetration at the district level. Hence, the variable measures how many Tweets are being produced in contrast to how many Twitter users are living in a district.

We quantify the number of Tweets in a district by extracting spatial data on Tweets limited to India from Twitter using its former Academic Track API and the archive search feature. Nearly 600,000 Tweets were pulled from the API from December 15, 2014, to December 31, 2014[12]. The decision to collect Tweets over this period is because we aimed at measuring exposure to Twitter right before the data collection process for the 2015-2016 DHS survey, specifically January 2015. Only the Tweets with point geo-data which corresponded to the districts in the sample were used. The total final sample was thus reduced to 339,554. The timing of data mining is aligned with Twitter's policy on the collection of geolocated GPS data[13], 4 months before a mandatory opt-in for the collection of this geo-data was enabled allowing us to anonymously populate the most representative and geographically accurate possible corpus of Tweets (see Lapowsky, 2019).

We pull data from 2019 in a separate query. The method of user location identification yielded 1.2 million Tweets over from May 15, 2019, to May 31, 2019, of which 930,951 Tweets were retained and the rest dropped due to noise. These Tweets are pulled immediately prior to the 2019-2021 DHS survey round for which data collection started in in June 2019. Unlike the Tweets from 2014 which used point data, a separate approach which identifies the centroids of bounding box was used as Twitter's regulations on geodata collection meant that users would have to opt-in to have their GPS data embedded into Twitter's metadata. The bounding box is an approximate region of the location from where the Tweet was sent out, from which Tweet location is assumed as the centroid of this box[14]. Although a relatively less accurate measure than GPS data, complete precision is not necessary since we aggregate the data at the district level. We also employ several robustness checks which validate this approach (see section 5).

[11] The DHS is also known as the National Family Health Survey (NFHS); Other controls are extracted from the 1999 DHS. Individual Women's Data - Individual Recode (IR) is used to conduct the analysis.
[12] Twitter introduced an opt-in for GPS metadata collection in April 2015. This data is completely de-identified before matching and processing, containing solely point coordinates for the analysis of location.
[13] Müller and Schwarz (2023) discuss the accuracy of Twitter's point geodata sample.
[14] We use a heuristic to eliminate very large values. Only bounding boxes with area less than or equal to 0.29 degree are included. The average district is 0.43 degrees.

Figure 1 shows the diffusion of Tweets from December 15, 2014, to December 31, 2014, by districts. The large variation in the diffusion of Tweets between districts is worth noting. Tweets are highly concentrated in the Northwest, West, and Southern regions of the country, while they are less prevalent in the East and North-Eastern districts.

Figure 1: Diffusion of Tweets in December 2014



Source: Authors' calculation using Twitter API metadata.
Notes: The map plots the diffusion of Twitter from December 15, 2014, to December 31, 2014, at the district level. The darker the colour gradients of the district, the higher the diffusion as measured by the number of Tweets.

2.2 Measuring Son Preference from the Demographic and Household Surveys

Data on preferences for the sex of the child (henceforth, preferences) are extracted from the India DHS[15]. For the purpose of this analysis, two rounds of DHS for 2014-15 and 2019-21

[15] The DHS has been previously used to study fertility preferences (see for example, Pande and Astone (2007) and Jayachandran and Pande (2017)).

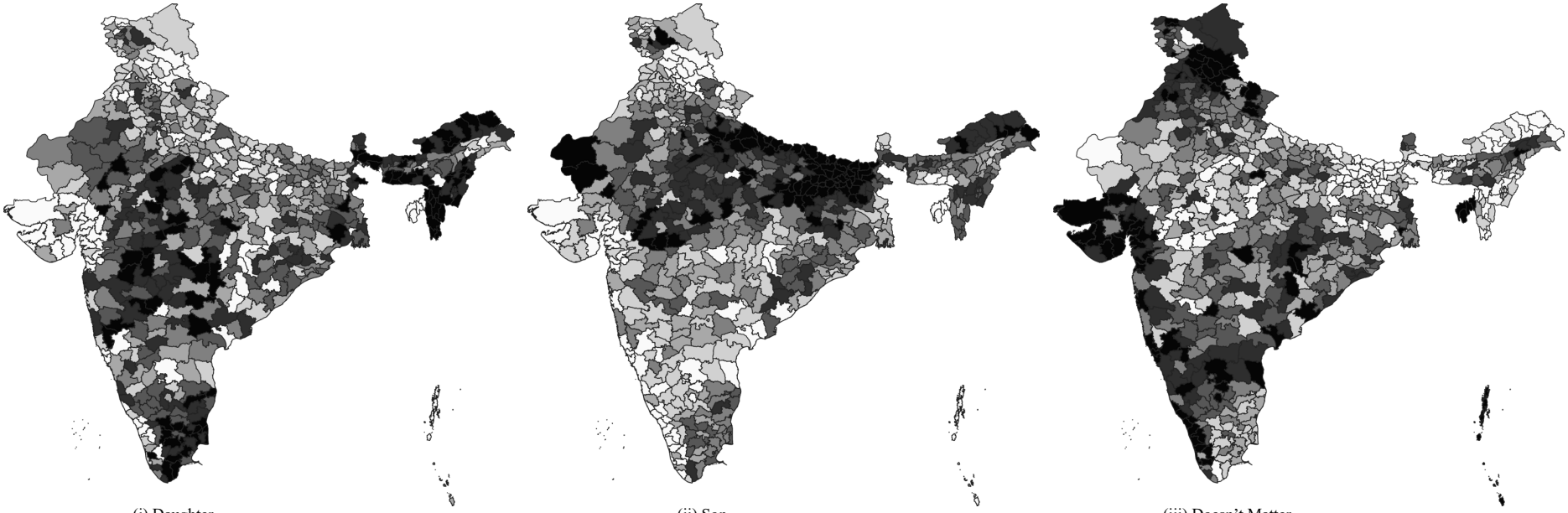
were used to ascertain preferences as a repeated cross-section[16]. The survey question we use is asked differently depending on whether or not the respondent has living children with probes for a numeric response. Those who do not have living children are asked: '*If you could choose exactly the number of children to have in your whole life, how many would that be'*. For women with living children, the question is preceded by the condition '*if you could go back to the time you did not have any children and [you could choose]…'*. To understand the preferred sex-composition of their desired number of children, the respondent is asked: *'How many of these children would you like to be boys, how many would you like to be girls, and for how many would it not matter if it's a boy or a girl?'*. Therefore, we construct three dependent variables based on questions regarding women's preferences. Each variable then represents a decomposition of the respondents' fertility preferences regarding the desired sex of their children (y^p), expressed as a ratio between the desired quantity of children of the preferred sex (q^p) and the total desired number of children (q^n):

$$y_{idt}^{p} = \left(q_{idt}^{p}\right)/(q_{idt}^{n}) \quad (1)$$

where *i*, *d* and *t* refer, respectively, to individual women, district and year. As a result, we have three dependent variables representing preference for number of boys, preference for number of girls, and preferences for children irrespective of their sex (henceforth regarded as non-discriminatory preferences). Figure 2 illustrates the distribution of preferences for the sex of the child across the different districts in India. Son preference has clear geographic demarcations and geographically represented trends in son preference correspond to gender norms. For instance, women in North India show greater degrees of son-biased fertility preferences.

[16] All district level variables from the census and earlier rounds of the DHS have been harmonised. A historical account of administrative restructuring of the country was prepared using official as well as secondary and geodata sources between 1990 and 2021. This was used to create a many-to-one match for districts which were bifurcated or divided into multiple districts, or one-to-one matches for districts which were renamed for the corresponding data year, with few districts and one state for which matching was not possible, leading to a total of 407 district-units per the 2001 Census in the final corpus. Some districts are not included in the analysis, including the state of Chhattisgarh, which was carved out of the state of Madhya Pradesh in the year 2000 and underwent extensive administrative restructuring in 2012 with districts which were formed from more than one erstwhile district. A full account of included districts and sample of matching is available in the Appendix.

Figure 2: Preferences for the sex of the child in India as a ratio of total number of children preferred



(i) Daughter (ii) Son (iii) Doesn't Matter

Source: DHS 2015, Authors calculation

2.3 Empirical Strategy

The influence of twitter on women's preferences for the sex of their child is investigated by estimating equation (2):

$$Preference_{idst} = \alpha + \beta Twitter_{dst} + \gamma X'_{2001,ds} + \varphi_s + \theta_t + \varepsilon_{idst} \quad (2)$$

where *i* refers to women living in district *d* and state *s,* surveyed in round *t*. *Preference* (Inverse Hyperbolic Sine) is the dependent variable and measures women's preferences for the sex of their child. As explained above, we measure preferences using three separate ratios: the number of preferred (i) son, (ii) daughter, or (iii) children for whom the sex does not matter, over the total number of preferred children. *Twitter* (Inverse Hyperbolic Sine) is our main independent variable and measures the exposure to Twitter using the number of Tweets recorded for district *d.* It is important to note that Twitter information refers to December 2014 and May 2019, prior to the collection of DHS data on preferences in 2015 and 2019, reducing concerns about endogeneity. X represents a vector of social and demographic characteristics measured at district level before 2001. In particular, we use information on female literacy, caste, and tribe characteristics which are sourced from the 2001 Census of India, and data on years of husband's education and religion which are obtained from the 1998-99 India DHS. Descriptive statistics of our sample are reported in Table 1. We further decrease concerns about endogeneity, including survey (θ_t) and state (φ_s) fixed effects which are expected to account for unobserved heterogeneity. We use survey weights in our regression in order to ensure that the results are representative of adult female population in the country and use robust standard errors throughout.

Table 1: Descriptive Statistics

	Mean	SD	Min	Max	N
Twitter (t=2014)	832.89	4010.96	0	34962	626
Twitter (t=2019)	1225.62	5002.90	0	104471	626
Son preference, ratio (t=2015, 2019)	.476	.213	0	1	1291917
Daughter preference, ratio (t=2015, 2019)	.407	.196	0	1	1291917
Non discriminating preference, ratio (t=2015, 2019)	.117	.314	0	1	1291917
Hindu population in district (t=1998)	.78	.256	0	1	407
Share of female scheduled castes in district (t=2001)	.149	.084	0	.502	407
Share of female scheduled tribes in district (t=2001)	.147	.252	0	.983	407
Husbands' education in years (t=1998)	6.149	1.767	.345	11.975	407
Female literacy in district (t=2001)	.442	.141	.145	.844	407

Notes: descriptive statistics with weights for DHS; un-weighted Census and Twitter data

3. Empirical Results

3.1 Main Results

The results of our main analysis, which links Twitter to women's preferences for the sex of their child are reported in Table 2. We find that greater exposure to Twitter significantly influences women's preferences for the sex of their child (Table 2, columns 1-3). In particular, the model outcomes suggest two relationships. Firstly, exposure is related to greater preferences for a child irrespective of their gender*, i.e.* non-discriminatory ('Either'). Secondly, both preferences for sons and daughters are negatively correlated with the exposure to Twitter. Interestingly, the impact is bigger in the case of son preference than daughter preference. All results are statistically significant at 1%. The direction and significance of these results are not sensitive to the inclusion of control variables (Table 2, Panel B).

Although we find this result conveys that exposure to Twitter is directly and meaningfully linked to greater non-discriminatory preferences and weaker son preference, the magnitude of the coefficient is small. Considering the full in-sample distribution of point estimates, we estimate that that average effect size of Twitter on non-discriminatory preference is 2%.

Table 2. OLS regressions, linking Twitter and women's preferences for the sex of their child

	(1) Either	(2) Son	(3) Daughter
Panel A: OLS Estimation, No Controls			
Twitter Diffusion	0.00851*** (0.000307)	-0.00629*** (0.000213)	-0.00314*** (0.000207)
Observations	1,291,869	1,291,869	1,291,869
R-squared	0.035	0.043	0.012
Controls	NO	NO	NO
State FE	YES	YES	YES
Year FE	YES	YES	YES
Panel B: OLS Estimation, With Controls			
Twitter Diffusion	0.00387*** (0.000337)	-0.00293*** (0.000233)	-0.00138*** (0.000225)
Observations	1,079,207	1,079,207	1,079,207
R-squared	0.038	0.047	0.013
Controls	YES	YES	YES
State FE	YES	YES	YES
Year FE	YES	YES	YES

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 respectively. Survey weights are applied. IHS transformation applied to all dependent variables and the Twitter variable. The discrepancy in the sample sizes in Panel A and Panel B is a result of matching loss for districts in 1999-99 DHS and 2001 Census from which district level control variables are sought.

3.2 Endogenous Exposure to Twitter

Twitter use does not occur randomly and is likely to be contingent upon other unobserved factors which raises concern that our estimates could include bias due to omitted variables which might explain the uptake of Twitter in a locality. We respond to this endogeneity issue by extending the baseline model using an instrumental variables (IV) strategy.

Our approach takes inspiration from Müller & Schwarz (2023) which exploits the path dependence of technology diffusion in instrumenting Twitter. Müller & Schwarz (2023) locate an event which explains Twitter's adoption trajectory in the United States of America. In the same vein, we capture interest shown towards an event which launched a national Twitter campaign in India during a time when Twitter was not yet widely popular in the country.

Twitter's rise in India can be attributed to a campaign in 2013 because of which thousands of Indians set up an account on the platform (Mitter, 2015). The Board of Control for Cricket in India (BCCI) announced that cricketer Sachin Tendulkar will be retiring from international cricket with a final match on November 16^{th,} 2013. Tendulkar is regarded as a national icon with enormous celebrity throughout India, a public figure and prolific cricketer, his acclaim is validated by the nickname 'God of Cricket' (Kidambi, 2011; Time Magazine, 2012). Tendulkar's retirement was marked not only by an enormous interest as it was the highest rated match on television in eight years, but also by a sentimental populace and a moving farewell speech by the retiring batsman. Conversation and appreciation were amplified through digital connections with many Indians turning to Twitter to publicly thank Tendulkar, however the choice of platform was not by chance. On the 6th of November 2013, BCCI launched a campaign for fans of Tendulkar to instantaneously receive an automated 'digital' autographed photograph with the Twitter users' first name, a message from Sachin and his signature if they Tweeted with the hashtag '#ThankYouSachin' and tagged the '@BCCI' Twitter handle. However, over the course of the week the campaign took on a new momentum when Tendulkar's retirement was announced with fans reaching out in support of his career. The campaign was supported by Twitter India and lead to more than 3 million Tweets (Twitter, 2013).

The official television broadcast displayed '#ThankYouSachin' alongside the match score, with some initial instructions to use the account handle and hashtag to encourage engagement through Twitter. We hypothesis that people interested in participating in the campaign looked for additional information through a search engine. Hence, we measure the attention captured by the twitter campaign using google trends data (our instrument). Indeed, Indian google trends data for 15th to 17th November for the search term 'Twitter' show that 'twitter sachin' was the second most related search term to Twitter among top searches, the first being 'twitter login'[17]. Searches for Sachin Tendulkar were at a historical high in this period.

[17] Google Trends provides 'Top Searches' which are search terms most frequently searched for in tandem with the given query.

The IV is based on the identifying assumption that the campaign introduced the act of tweeting to fans of cricket and those interested in either sharing a token of appreciation or receiving a personalised digital keepsake of the retiring star.

According to Twitter officials, the campaign was the 'biggest splash' in Indian Twitter's history as it not only created new usage records with a Tweet every 6 seconds on match days but also led to thousands of Indians joining Twitter, with international sports personalities such as Lewis Hamilton and Roger Federer tweeting #ThankYouSachin (Mitter, 2015). Prior to the outbreak of this campaign in late 2013, India was not considered an important market for Twitter and the event was a breakthrough moment for the platform in India. It was at this juncture that Twitter introduced its first ever blog for India with the title 'Hello India, #ThankYouSachin[18]', a testament to the popularity of the campaign.

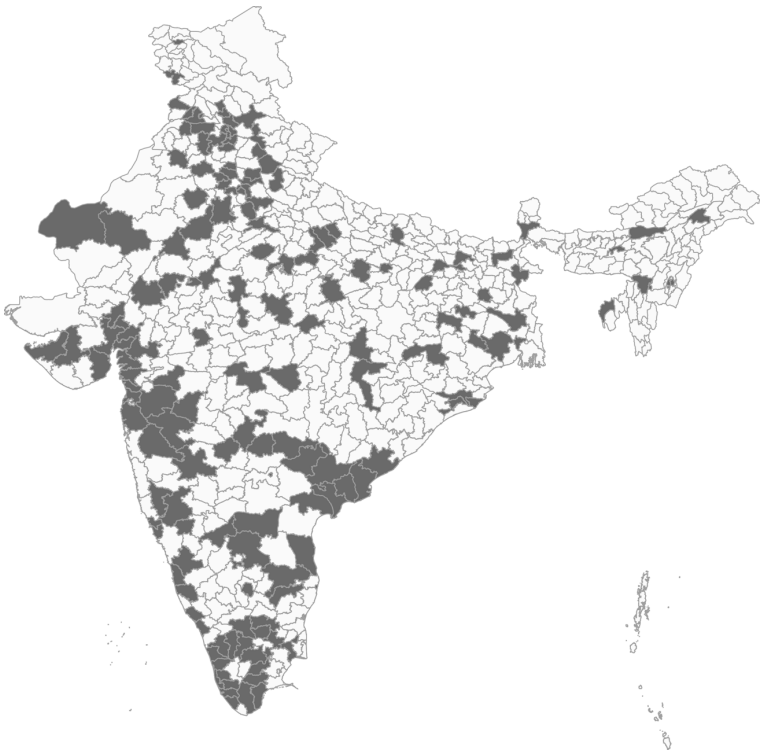
The nudges to use Twitter for discussing Tendulkar's retirement were present across mediums. Through televised reinforcement and online campaigns, Tendulkar's retirement led to the first large-scale Twitter campaign in India, so much so that Twitter changed how Indians obtain information and communicate. Other network effects, such as heightened discussions and word of mouth related to the campaign, may also have influenced relative and easier diffusion of innovations in the platform *i.e.,* the rate of knowledge about the social media platform, its network influence leading to its adoption and continued use (See Rogers, 1995, pp.235). On the other hand, we believe that discrimination on the sex of the child is unaffected by interest in Tendulkar's retirement because cricket is where Indians finds common ground. The sport is integral to modern India's cultural and national esteem. it is often viewed in a group and played in every corner of the country. Devan (2012) suggests that cricket creates a 'single national identity', a unifying force and which connects Indians from various backgrounds and acts as a source of national pride. Sachin Tendulkar's celebrity, cricket's role in social cohesion, the universal preoccupation with cricket in the country, and a campaign geared at bringing sentimental Indians onto Twitter to celebrate the cricketer allow us to satisfy the exclusion restriction. We are thus able to measure the impact of Twitter on son preference reliably.

To measure virtual interest in Sachin, we limit Indian google trends data to three days to account for the peak when people were more likely to engage in appreciative expression online. The day of the match when Sachin Tendulkar entered the cricket ground, 15th of November, the retirement speech on the 16th of November and one day after, the 17th of November when print media covered the event. The keyword used is 'Sachin Tendulkar' and the steps to composing the variable are as follows. Firstly, we pull data at the city level including low search volume to account for wider interest, then we match 265 cities to the districts based on the 2011 Census, which is identical to the 2015 DHS district fixed boundaries. The database of cities includes less populous divisions or 'taluks' as well as large cities. Finally, since google trends reports a normalised index value which measures the search for Sachin as a proportion

[18] https://blog.twitter.com/en_in/a/2013/hello-india-thankyousachin

of the total search volume, the variable is constructed as a dummy for interest in Sachin[19]. The aggregated outcomes from Google Trends are presented in figure 3 and interest is diffused in pockets of districts throughout the country.

Figure 3: Google Trends data for Districts with Searches for Sachin



Source: Google Trends, Authors Calculation

3.3 IV results

As reported in Table 3, the instrument is positively correlated with the treatment *i.e.,* the exposure to Twitter (First stage, column 1). The F-statistics for each of the estimations is above 10 and the reported Kleibergen-Paap F-statistics indicate that the instrument is relevant. The validity of the estimations is further supported by the results of the Anderson-Rubin Wald tests. The second stage of the IV strategy (columns 2-4) are concomitant with the prior results, in direction and in significance. Non-discriminatory preferences are a consistent finding, along with the result that exposure reflects weaker son preference than daughter preference for women. The coefficients for son preference and non-discriminatory preferences retain their significance, and direction with the addition of controls (Column 2, 'Either') although with lower magnitudes than the base IV specification. Preferences for daughters loses significance

[19] Any interest is '1', negligible [*i.e.,* a 0 value] or no interest is '0'. Trends solely provides data for popular terms, so a value of '0 is given to search terms with low volume or an unpopular term, as a result we include all districts, even if some cities in other districts are not explicitly reflected in this search.

with the addition of controls (Appendix, Table A1.5, Column 4). The direction of these results is consistent with findings using the OLS estimator.

Table 3. IV regressions, linking Twitter and women's preferences for the sex of their child

	(1) First Stage (Twitter)	(2) Either	(3) Son	(4) Daughter
Panel A: First Stage Instrumental Variables, With Controls				
Interest in Sachin Tendulkar (t=2013)	0.950***			
	(0.00612)			
Panel B: Instrumental Variables Estimation, With Controls				
Twitter Diffusion (t=2014, 2019)		0.00836***	-0.00794***	-0.000685
		(0.00110)	(0.000777)	(0.000751)
Observations	1,079,207	1,079,207	1,079,207	1,079,207
KP F-Stat	24137.80			
Anderson Rubin Wald Test	56.93			
State FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Weights	YES	YES	YES	YES
Controls	YES	YES	YES	YES

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 respectively. Survey weights are applied. IHS transformation applied to all dependent variables and the Twitter variable. Anderson Rubin Wald test statistic from (2).

3.4 Probing the Instrument

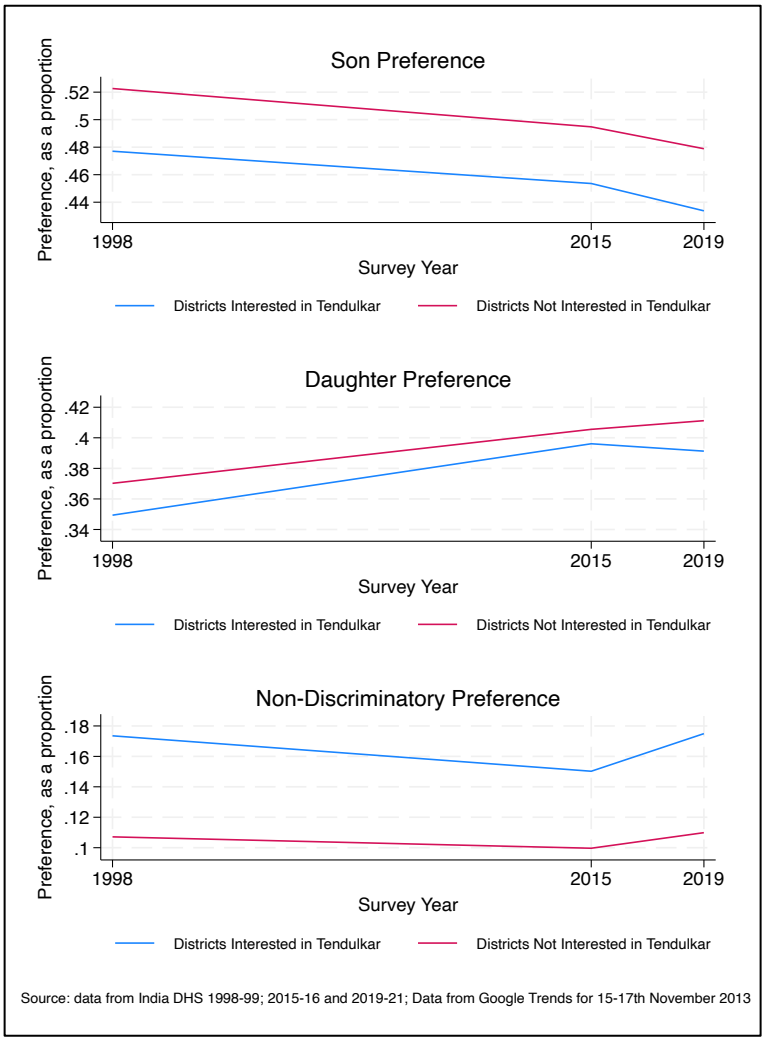
To confirm this result empirically, we test the robustness of our baseline IV specification. First, we run a balancing test to check if areas with more or less interest in the search term 'Sachin Tendulkar' are the same in their baseline characteristics including proportion of households in district which are Hindu, scheduled caste, scheduled tribe, proportion of literate females in the district, and average years of education of the husband from the 1999 India DHS and 2001 Census of India. We find insignificant results for all covariates with the exception of female literacy (Table 5, Column 1). We control for these set of covariates to strengthen our independence assumption in our main analysis.

Replacing female literacy with average years of education for females at baseline yields insignificant results (Table 5, Column 2). We include this result as education, particularly since higher and secondary education is associated with weaker son preference (Pande & Astone, 2007).

Second, we split the districts based on our IV as those with reported interest in Sachin Tendulkar and those without and track preferences for the sex of the child over time using

district level data from the DHS data from 1998-99, 2015-2016, and 2019-2021[20]. Figure 4 shows parallel time trends until 2015 for both categories of cells with respect to our three outcomes of interest. Last, we conduct a placebo regression and replace our main dependent variable with aggregate son preference in a district in 1998-99, i.e. prior to any social media, on interest in Sachin for all districts. Our results are not significant which further affirms the exclusion restriction (see Table 4, column 1).

Figure 4: Preferences for the sex of the child and interest in Sachin Tendulkar over time



[20] We exclude the DHS data from 2005-2006 as it anonymises the district level data.

Table 4: Regressing Son Preference (District Aggregate) in DHS 1998-99 on Interest in Sachin

	Aggregate Son Preference in District (t=1998-199)
Panel A: OLS Estimations with Controls and Weights for Ever Married Women	
Interest in Tendulkar (t=2013)	-0.0114
	(0.0139)
Observations	404
R-squared	0.432
State FE	YES
Controls	YES

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 respectively. Weights for total ever married women in district applied from 2001 census in Panel A.

Table 5: Balancing Test

	(1) Interest in Tendulkar t=2013	(2) Interest in Tendulkar t=2013
Hindu	-0.0986	-0.0485
	(0.182)	(0.192)
Proportion of Scheduled Caste (Female) (t=2001)	-0.473	-0.685
	(0.548)	(0.591)
Proportion of Scheduled Tribe (Female) (t=2001)	0.00715	-0.278
	(0.250)	(0.258)
Average Years of Husband's Education (t=1998-99)	0.0106	0.0156
	(0.0227)	(0.0358)
Proportion of Literate Women (t=2001)	1.628***	
	(0.383)	
Average Years of Women's Education (t=1998-99)		0.0545
		(0.0377)
Constant	-0.258	0.265
	(0.217)	(0.194)
Observations	404	404
R-squared	0.263	0.222
State FE	YES	YES

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 respectively. State fixed effects. Weights for total ever married women in district applied from 2001 census.

4. Mechanisms and Heterogeneity

4.1 Online Narratives about Children

So far, the evidence supports that technology access and exposure decrease son preference for the Indian context. But what are the factors behind this influence? One possible reason is that individuals engage in and are exposed to content outside of their typical worldview. For instance, Twitter Trends India suggests that there has been an increase in the recognition of women's contribution to the workforce on the platform (2022). Yet, this offers a limited account of the content that is produced in India, while new cultural models could challenge discrimination, virtual spaces host a vast variety of narratives. We explore how children's issues are viewed online through a deep learning text classification fine-tuned transformers model followed by a thematic analysis of Tweets.

4.1.1 Classification

English language Tweets are collected from two additional corpora for 2015 and 2018[21]. The Tweets for the 2015 corpus were composed between February and April. The 2018 dataset contains Tweets from September to December. We identify and label Tweets which concern topics which distinctly mention children and adolescents or are related to children and their experiences, such as school and family relations. Tweets were annotated by one author and used to validate the model. A small minority (approximately 3.3%) of Tweets in the sample of annotated Tweets concern children.

The strategy to identify these two themes among a larger corpus of Tweets involves fine-tuning a sequence classifier to identify these themes on Twitter. XLM-RoBERTa, a multilingual transformers model is used to classify the Tweets, which has been pre-trained on 2.5 TB of data and 100 languages. The choice of classifier is guided by two additional factors. Firstly, the abundance of languages supported by Indian Twitter, as the model can predict Tweets for multiple languages with an English language input. Secondly, the size of the data and consequently the computationally intensive nature of classification which are simplified by using a pre-trained architecture. Two of the chief advantages of using XLM-RoBERTa are the possibilities of using data which are trained solely in one language, and its success in outperforming other comparable model's such as BERT and architectures, such as GPT-3.5, including for sexism and toxicity classifiers (Azadi et al., 2024; Conneau et al., 2020).

[21] Collected using the erstwhile Twitter Academic Track API in 2023, with a query used to pull all georeferenced Tweets in the respective time periods

From 2800 self-annotated Tweets, a total of 407 were used to fine-tune the model[22]. Predictions were made on a random sample of 100,000 Tweets each from the 2015 (February, March and April) and 2018 corpora (July to December)[23].

Table 6: Predicted Labels from Fine-Tuned Model

Year	Label	Freq.	Proportion
2015	Unrelated to children	95,364	.9702
	Related to children	2,928	.02979
	Total	98,292	1
2018	Unrelated to children	93,105	.9682
	Related to children	3,056	.03178
	Total	96,161	1

To examine the salience of language on Twitter, we include a sample of de-identified English language outputs from the model. To maintain anonymity Twitter users have been truncated to '@', names have been replaced with ''. Major spelling errors in Tweets have also been corrected. Samples of code mixed in Indic language Tweets with translations are provided in the Appendix (Table A2.2).

Table 7: Model Outputs for a Chosen Sample of 2018 Tweets in English

Sample of Tweets (2018)
@ homes I would go for an Indoor Kids Play Zone. It's very important that my kids leave the gadgets and involve themselves in some physical play activities. #'' @
After request from many audience '' glasses for kids has been ordered and available from 2day. Kindly ask for the kids glass it will be small in size and compact which in turn will give a great viewing and enjoyable experience for the kids. #
Often Children are major decision influencers in a household. Using them as strong '' Agents. "Amma-Naana! Won't you go vote? I'm a kid, so I can't vote. Are you a kid as well??" So! Don't b a Kid in your kid's eyes. Come Vote! PS: My poster for TelanganaElection2018
When you finish ... All that remains are the memories ... The kids made it colorful .. as part of the '' #'' we sang .. we played... We danced.. stories did happen :)…
@ @ In [Children's Organisation], Children get benefited. Here, one such child is sharing his personal experience 👇 #BuildingMastersOfFuture
fellow @F after her Opening Workshop with her girls today at Delhi leadership #womenmatters #younggirls #womenempowerment #womenleaders #girlpower #genderequality #shecan #IWillGoOut
@ #GharLaoTyohaar I am gonna make my Navrartri best this year by Light an Akhand Jyot in front of Goddess Durga & having fun in Durga puja,Garba dance with my friends and give gifts like sweets and traditional food to young girls. join @ @ @

[22] To treat the unbalanced data, we downsample the majority class and upsample the minority class. The model is trained Over 6 epochs – until a good fit is achieved, see Appendix (A2.1).
[23] Data cleaning involved removing links, newlines and symbols, as a result the final corpus does not contain link only Tweets so the total number of Tweets on which predictions are made is smaller

Students were given Hands-on assignments and encouraged to ask questions to get doubts cleared in different areas of science and promote their effective participation. @ '' Science School Students stem knowledge #TuesdayThoughts

Table 8: Model Outputs for a Chosen Sample of 2015 Tweets in English

Sample of Tweets (2015)
God blessed us with baby boy
Enjoy the delightful squeals as your lil one rides in a Baby Walker Be sure they are Safe by following these Tips
'' in jail due to POCSO Misuse done by an adult girl! Why Legal System can't save innocents? निर्दोष '' न्यायमिले
@ masala chaas on the present '' crisis I see a very weak comments from ''! Think d students of india have a say ??
Two Mumbai girls rap against rape, video goes viral
So, this kid mind you (18 years old) is just too crazy ! We just keep bumping into each other…
@ Lol!! Uncle - I studied in govt schools all my life.. That is not the point. We shouldn't increase the burden
'' prescribed only two text books for class I and II and three text books for students of class III to V.
@ says he'd like his kids to get into films. Read the full interview by @ in '' today

These Tweets indicate that Twitter is indeed a platform where children and their issues are openly and commonly discussed, which include rights and education, as well as appreciation. We also find that between the two time periods, the proportion of Tweets related to children has shown a modest increase (Table 6).

4.1.2 Thematic Analysis

One of the main challenges to analysing Twitter discourse on children is the variety of topics may co-occur alongside narratives on children. We identify and dissect these topics through a thematic analysis of Tweets. To understand whether Twitter can expose individuals to progressive or regressive content related to children, a measure of 'stance' is constructed to categorise Tweets into 'progressive', 'neutral' or 'regressive (See table A2.4 in the appendix for definitions).

A Tweet is considered as a single document in which codes are inductively added and refined until all primary topics embedded in a Tweet are covered[24]. From an analysis of 500 Tweets, 250 for each of the two time periods, we identify 7 themes and 21 categories (Figure A2.3 in the Appendix). One category under the stance theme is uniquely assigned to each Tweet, against which co-occurrences of the remaining categories are observed. Here we explore if the content of Tweets related to children more progressive or regressive.

[24] All documents are coded using AtlasTI, all data are de-identified prior to input, with only data containing the text of the Tweet and year in which the Tweet was composed in the analysis. Indic language Tweets (23%) were translated to English and coded.

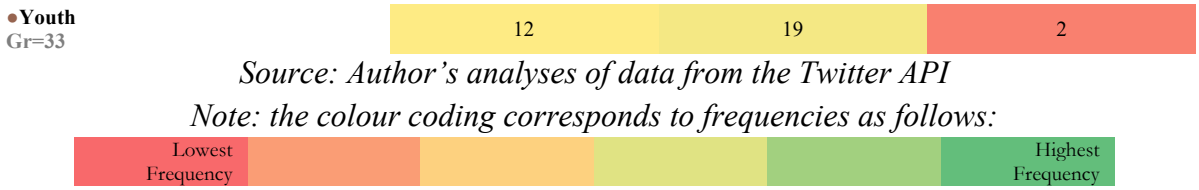
Our main finding relates to stance (Table 9). We analyse categories in 6 of the themes against stance to learn that a vast majority of Tweets are either neutral or progressive. Which suggests that exposure to discourse related to children on Twitter is rarely regressive.

The largest intersection of Tweets are neutral discussions include on education and student life. The second most prevalent category includes politics, elections and political leadership, which are neutral or progressive in their stance. In terms of identity, discussions on youth, girls and young women are largely progressive, whereas those on boys and young men are mainly neutral. Where we find most of the progressive discourse in the thematic area of rights and public policy which includes, activism and campaigns; health, law and public policy; maternal and childbirth; and sexual violence and harassment. Regressive discourse is most commonly associated with religion, disinformation, and sexual violence[25]. Of the total volume of Tweets analysed, we can conclude that users' discussions on twitter related to children are topically varied. Twitter exposes its users to a great variety of ideas regarding children and adolescent, a majority of which express neutral or progressive messages.

Table 9: Category Frequences and Stance

Category	●Neutral Gr=235	●Progressive Gr=152	●Regressive Gr=43
●Activism and Campaigns Gr=32	2	30	0
●Boys and Young Men Gr=25	18	5	2
●Childhood and Nostalgia Gr=14	11	3	0
●Crime, News and Incidents Gr=36	29	6	1
●Discrimination (Incl Sexism) Gr=15	1	8	6
●Disinformation Gr=21	1	1	19
●Education and Student Life Gr=212	123	79	10
●Empowering or Celebratory Gr=68	28	40	0
●Entertainment and Sports Gr=32	21	11	0
●Family and Relationships Gr=55	40	10	5
●Girl Child or Young Woman Gr=55	13	31	11
●Health, Rights or Inclusion Gr=49	14	34	1
●Humorous Gr=18	15	0	3
●Law, Public Policy or Programmes Gr=39	6	20	13
●Maternal and Childbirth Gr=21	7	13	1
●Political Leaders and Electoral Politics Gr=73	33	31	9
●Religion Gr=48	15	11	22
●Sexual Violence and Harassment Gr=43	9	14	20

[25] Although a volume of Tweets related to all three are part of a disinformation campaign by the followers of Asaram, a spiritual leader with an extensive following, convicted of raping a minor girl. We therefore do not draw inferences on disinformation.



Source: Author's analyses of data from the Twitter API
Note: the colour coding corresponds to frequencies as follows:

4.2 Mechanisms: Outcomes Related to the #SelfieWithDaughter Campaign

To capture the mechanisms behind impact, we explore the link between a targeted twitter-driven campaign for girls' rights and women's preferences for the sex of their child. On June 28, 2015, Narendra Modi, Prime Minister of India Tweeted "I urge all of you, share a #SelfieWithDaughter. Also share a tagline that will encourage 'Beti Bachao, Beti Padhao' ['*Save the girl child, educate the girl child*']: PM *@narendramodi*"[26] on the *@PMOIndia* account (Figure 5). The Tweet referred to two distinct interventions. Firstly, the *#SelfieWithDaughter* campaign, a website started by Sunil Jaglan, the village head of Bibipur village in Haryana, a state in India with historically adverse sex ratios (Jha *et al*., 2011; Kishor, 1993; Jejeebhoy *et al.*, 2015). The intention of the campaign was to post a picture with one's daughter in an attempt to celebrate daughters in a state with strong son preference. Secondly, the *@PMOIndia* Twitter handle brought this campaign to Modi's millions of followers one year after his first electoral success, to bring attention to the *Beti Bachao Beti Padhao* scheme for the girl child. This scheme was launched as a joint effort by several Indian ministries in January 2015 in Indian states with adverse sex ratios at birth with a behavioural design comprising largely of campaigns. *#SelfieWithDaughter* was the second most popular hashtag with engagement surpassing 375,000 Tweets, often with fathers clicking selfies with their daughters. Its popularity was second only to *'#IPL'* for the Indian Premier League, a cricket tournament in 2015 (Twitter, 2015). The hashtag and movement have endured. *#SelfieWithDaughter* was the topic of discussion in April 2023 on the official radio channel of the Prime Minister, titled 'Mann ki Baat' ['*Conversations of the Heart, Mind*'], an outreach initiative for Indian public. The success and importance of the campaign was discussed on the show.

Figure 5: Tweet by @PMOIndia encouraging readers to share a #SelfieWithDaughter

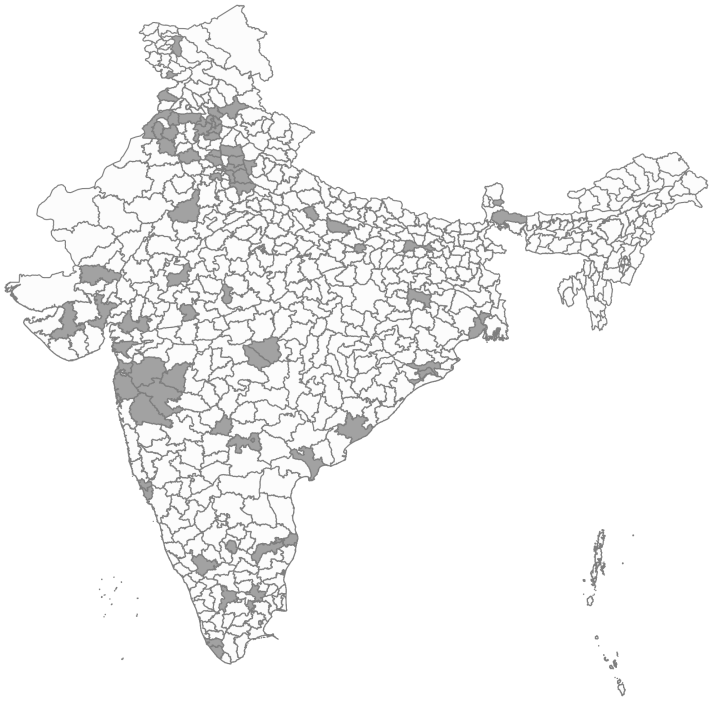


Source: @PMOIndia Twitter Account

[26] Permalink: https://twitter.com/PMOIndia/status/615034960492171264

To capture the footprint of the hashtag, we use our corpus of gendered language georeferenced Tweets over a period of one month starting from the day *@PMOIndia* Tweeted with the hashtag including the month of July. We conduct a basic text analysis and limit the corpus into Tweets mentioning 'beti' or combinations of 'selfie with', to account for mutated narratives, with the final corpus comprising of *#selfiewithdaughter* and associated Tweets. To study the impact of exposure, we contain the sample to districts where the campaign related messages were produced and study the relationship with exposure to Twitter. Figure 6 shows a map of all 76 districts which we include in the analysis where the campaign was considered active based on these Tweets.

Figure 6: 'Selfie with Daughter' Districts



Source: Twitter API, Authors Calculation

Table 11 reports the results of our analysis. With this, we find that the magnitude of our estimate indicates that exposure to twitter is connected with strikingly lower son preference for campaign districts i.e., the effect is magnified for these regions (Table 10, Panel B, Columns 2-4). The results are significant to at the 5% level[27].

[27] OLS results are presented in A5.6 of the appendix

Table 11: Twitter Exposure and Preference for the Sex of the Child: Results for Girl Child Activism

VARIABLES	(1) First Stage (Twitter)	(2) Either	(3) Son	(4) Daughter
Panel A: First Stage Instrumental Variables				
Interest in Tendulkar (t=2013)	0.0439***			
	(0.0150)			
State FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Controls	NO	NO	NO	NO
Panel B: Instrumental Variables Estimation, #SelfieWithDaughter				
Twitter (t=2014)		0.416**	-0.481**	0.0679
		(0.199)	(0.196)	(0.105)
Observations	74,302	74,302	74,302	74,302
KP F-Stat	8.52			
Anderson Rubin Wald Test	8.93			
State FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Controls	YES	YES	YES	YES

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 respectively. Survey weights are applied. IHS transformation applied to all dependent variables and the Twitter variable. Anderson Rubin Wald test statistic from (2).

4.3 Heterogeneity

The main analysis shows that Twitter exposure is associated with women's preferences for the sex of their child. Thus, providing an answer to our central research questions, we find evidence of a systematic link between social media exposure and preferences. Son-biased preferences may vary due to numerous social, economic, geographic and cultural factors. Heterogenous effects for literacy, religion, caste, and region are explored in this section. All tables are reported in section A1 of the Appendix

To perceive the relationship between exposure to Twitter and literacy, we limit the sample to those districts with average baseline female literacy higher than the national median. Data for baseline literacy is taken from the 2001 Census of India Primary Census Abstract. Table A1.4 suggests that the relationship between social media exposure and weaker gender norms regarding non-discriminatory preference is stronger for districts with higher baseline female literate population, however the same cannot be said about son preference as the coefficients in exposure to Twitter are akin to that of the baseline IV specification (Column 3). We should expect a stronger result because higher education in linked to weaker son preference (See Pande & Astone, 2007). However, in the case of Twitter this relationship now pivots to less discrimination in preferences overall (Column 2). For districts with low baseline literacy (Panel D), we find significant results and a relatively weaker coefficient. The effect does not differ

very substantially between low and high baseline literacy districts, against expectations on the role of literacy in Twitter's influence on discrimination.

Other social and regional characteristics which could condition Twitter's influence are examined in further analyses (Appendix, Tables A1.1-A.1.3) and confirm Twitter's role in weakening son preference. By splitting the sample for North Indian states, we can observe the influence of the platform on a region marked by a skewed, male-biased sex ratio (See Agnihotri *et al.*, 2002; Kabeer *et al.,* 2014). For North India, the coefficient for non-discriminatory preferences is significant but lower than in other states (Table A1.3, Panel B and D, Column 2). For this group, the coefficient for son preference and a relatively modest coefficient for daughter preference are also significant and negative (Columns 3 and 4). Other social factors which could potentially mould influence are caste and religion (Jayachandran, 2015; Kabeer 2014). Results for the Hindu majority split suggest that Twitter exposure is associated with less discriminatory preferences and lower son preference to a slightly greater degree than the non-Hindu segment (Table A1.2, Panels B and D, Columns 2-3). For individuals belonging to districts with high baseline female scheduled caste population, the effect of Twitter is also significant and positive for preferences for daughters (Table A 1.1., Panel B, Column 4). While the directions for non-discriminatory and son preference are the same (Columns 2 and 3), the weaker coefficient for both these specifications and the positive coefficient for daughter preference (Column 4) suggest that regions with different caste structure may receive media technologies differently. Among all heterogenous analysis, it is those who do not belong to these districts (Panel D, i.e., low female scheduled caste population), which have the highest magnitude for the estimate on non-discriminatory preferences among all socio-economic and geographical segments. For this group, coefficients for both son and daughter preferences carry negative direction (Columns 3 and 4). All results from the heterogeneity analyses find that the effect of son and non-discriminatory preferences is statistically significant and the direction of impact unchanged albeit with varying magnitudes.

5. Robustness checks

We test the sensitivity of the Twitter variable, using an alternative time period and two unique measures of diffusion, testing the robustness of the main variable of interest.

To validate the 2019 data, a secondary approach is used as the Twitter measure for robustness to check for the different method of aggregation for the 2019 Twitter data. A new measure, namely *Twitter Density* is used as an alternative aggregation technique, an artificial hexagonal fishnet is created to represent a uniform geographical space, and the centroid of the polygon is then associated with a Tweet. Bounding box points smaller than .3 square-degree are included and hexagons of 0.5 degree. An illustration of the 2019 points is presented in Figure A4.1 in the Appendix. The correlation between the two time periods for Twitter is positive at 0.60, despite using a different approach and boundaries for aggregation.

Twitter Distance: An additional check is introduced wherein the effect of distance of the sampled cluster from the nearest Tweet is measured for the 2019 measure[28]. The distinction between the density of Twitter activity and distance (unlike proximity) from Twitter activity will have an inverse relationship.

In the distance estimations, presented in Panel B for Table A5.2 (IV) and Table A5.1 (OLS) in the Appendix, we note that the relationship between distance and son preference is straightforward, the further one is from the physical location of a Tweet in these estimations is associated with more son preference relative to preference for daughters, *i.e.,* relatively more gendered preferences and are more likely to discriminate. This corresponds with the central finding that twitter exposure is linked with less discriminatory preferences and access to social media may be associated with weaker traditional gender norms, as indicated by the negative direction of the son preference variable in Table A5.2, column (3) in Appendix.

Alternative clustering and weighting strategies: we apply different clustering strategies to find that our results hold for state and primary sampling unit (PSU) clusters. A nearest neighbour technique is applied to geo-match clusters between the two DHS rounds. We also test our results against different approaches to weighting and find that an alternative weighting strategy drives significant results for daughter preference not found in the main analyses. Results for son preference and non-discriminatory preferences hold across all these approaches, presented in tables A5.3-5.5 in the Appendix.

Including data from 1998-99 DHS: finally, we test our claim that Twitter leads to lower son preference with the addition of 1998-99 data, a period before the prevalence of social media in the country. A caveat to using the 1998-99 data is the significantly smaller sample size of 90,303. Following the specification of our main model, we readjust the fixed district boundaries to their 1998 units and assign null values to both Twitter and the IV. We do not find any change in the relationship other than weaker magnitudes for both son and non-discriminatory preferences. Results are reported in table A5.7 of the Appendix.

6. Conclusions

We examine the effect of social media platform, Twitter on son-biased fertility preference in India. Findings from over a million Tweets, and survey data for more than one million women consistently confirm that social media exposure reduces son preference, and women are less likely to discriminate on the sex of the child. Until now, the large-scale impact of new media technology on gender norms was unclear. Our research utilises an instrumental variables strategy to examine the treatment effect of exposure to Twitter to discover that social media can influence girls' rights and the value that women ascribe to daughters. This result adds to a literature on media and its impact on fertility and preferences for women and girls, in agreement

[28] While this is a noisy variable as the DHS reports an approximate cluster location (having a maximum random variation from actual sampled location of 5 km for rural clusters with 1% randomly displaced by up to 10 km and 2 km for urban clusters), as well as our strategy of estimating an approximate location. This allows us to create a variable which reflects the distance between the DHS samples cluster and the nearest measured estimate for the location of the Tweet.

with the findings by La Ferrara *et al.,* (2012), Jensen & Oster (2009) and Kearney and Levine (2015), *i.e.,* exposure to new media is linked to the changing beliefs and preferences.

The potential of social media technology in reducing harmful preferences is not immune to pre-existing conditions. All reported results from the heterogeneity analyses confirm that Twitter reduces discrimination irrespective of socioeconomic group. The platform does not have a uniform influence on gender norms. While we find differing degrees of influence, the direction and statistical significance do not change for son and non-discriminatory preferences. We examine the relationship for those who belong to the historically patriarchal North Indian states or districts where the disadvantaged and marginalised scheduled caste population is predominant, the effect of social media is weaker for this group than for its counterparts, respectively. Interestingly, for the scheduled caste segment, we find that instead of substituting sons for a child of either sex, technology exposure increases preferences for daughters as well, a result which is unique for this socioeconomic group. We also analyse literacy, by containing our sample to historically high and low female literacy regions. Our estimates for these two groups are not substantially different. The differences in estimates between the Hindu majority segment and all others is likewise modest.

To capture how technology renders its effect, our paper also puts the spotlight on endorsements as a mechanism for the impact. Through the analysis of a widely adopted social media movement #SelfieWithDaughter which was publicly endorsed by Prime Minister Narendra Modi on the platform, we learn that narratives channel the effect of a new technology in challenging long-held beliefs. Indeed, the magnitude of impact is far greater for districts where the campaign was active. While narratives have the ability to produce social norms (Shiller, 2017), we demonstrate that these norms respond to digitally mediated narratives.

We also examine the nature of content on the platform related to children. Narratives are distilled using a fine-tuned transformers model to capture how children and their concerns appear on Twitter. Content on the platform is overwhelmingly either neutral or progressive. Twitter offers a platform for users to discuss an array of issues associated with children, from public policy, rights and education to family and relationships. Discussions on girls and young women are more often than not progressive. This evidence suggests that new media influences women's preferences through the amplification girls' rights issues via hashtag activism, on a platform where child-centric discussions indeed take place.

This evidence leads us to the conclusion that digital technology may generate a cultural shift in a developing country context, especially through the spread of progressive ideas when championed by public figures and accepted by users. Our research on preferences allows us to capture the promise of a new technology in addressing a nefarious form gender-based discrimination, by showing that new technology has persuaded Indian women to change their minds about favouring sons.

5. References

1) Adena, M., Enikolopov, R., Petrova, M., Santarosa, V., & Zhuravskaya, E. (2015). Radio and the Rise of The Nazis in Prewar Germany*. *The Quarterly Journal of Economics*, *130*(4), 1885–1939. https://doi.org/10.1093/qje/qjv030
2) Agnihotri, S., Palmer-Jones, R., & Parikh, A. (2002). Missing women in Indian districts: A quantitative analysis. *Structural Change and Economic Dynamics*, *13*(3), 285–314. https://doi.org/10.1016/S0954-349X(02)00023-1
3) Ahankari, A. S., Myles, P., Tata, L. J., & Fogarty, A. W. (2015). Banning of fetal sex determination and changes in sex ratio in India. *The Lancet Global Health*, *3*(9), e523–e524. https://doi.org/10.1016/S2214-109X(15)00053-4
4) Ahlawat, N. (2009). Missing brides in rural Haryana: A study of adverse sex ratio, poverty and addiction. *Social Change*, *39*(1), 46–63. https://doi.org/10.1177/004908570903900103
5) Alatas, V., Chandrasekhar, A. G., Mobius, M., Olken, B. A., & Paladines, C. (2019). *When Celebrities Speak: A Nationwide Twitter Experiment Promoting Vaccination In Indonesia* (Working Paper 25589). National Bureau of Economic Research. https://doi.org/10.3386/w25589
6) Ang, D. (2023). The Birth of a Nation: Media and Racial Hate. *American Economic Review*, *113*(6), 1424–1460. https://doi.org/10.1257/aer.20201867
7) Arias, E. (2019). How Does Media Influence Social Norms? Experimental Evidence on the Role of Common Knowledge. *Political Science Research and Methods*, *7*(3), 561–578. https://doi.org/10.1017/psrm.2018.1
8) Azadi, A., Ansari, B., & Zamani, S. (2024). *Bilingual Sexism Classification: Fine-Tuned XLM-RoBERTa and GPT-3.5 Few-Shot Learning* (arXiv:2406.07287). arXiv. https://doi.org/10.48550/arXiv.2406.07287
9) Battisti, M., Kauppinen, I., & Rude, B. (2022). *Twitter and Crime: The Effect of Social Movements on Gender- Based Violence* (ifo Working Paper No. 381). https://www.ifo.de/en/publications/2022/working-paper/twitter-and-crime-effect-social-movements-genderbased-violence
10) Bhalotra, S., & Cochrane, T. (2010). *Where have all the young girls gone? Identification of sex selection in India* [The Centre for Market and Public Organisation]. The Centre for Market and Public Organisation, University of Bristol, UK. https://econpapers.repec.org/paper/bricmpowp/10_2f254.htm
11) Bursztyn, L., Egorov, G., Enikolopov, R., & Petrova, M. (2019). *Social Media and Xenophobia: Evidence from Russia* (Working Paper 26567). National Bureau of Economic Research. https://doi.org/10.3386/w26567
12) Castells, M. (2009). Prologue: The Net and the Self. In *The Rise of the Network Society* (pp. 1–27). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781444319514.ch
13) Cheng, I.-H., & Hsiaw, A. (2022). Reporting Sexual Misconduct in the #MeToo Era. *American Economic Journal: Microeconomics*, *14*(4), 761–803. https://doi.org/10.1257/mic.20200218
14) Chandy, R. K., Johar, G. V., Moorman, C., & Roberts, J. H. (2021). Better Marketing for a Better World. *Journal of Marketing*, *85*(3), 1–9. https://doi.org/10.1177/00222429211003690
15) Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., & Stoyanov, V. (2020). *Unsupervised Cross-lingual Representation Learning at Scale* (arXiv:1911.02116). arXiv. https://doi.org/10.48550/arXiv.1911.02116
16) Davis, B., Grewal, D., & Hamilton, S. (2021). The Future of Marketing Analytics and Public Policy. *Journal of Public Policy & Marketing*, *40*(4), 447–452. https://doi.org/10.1177/07439156211042372
17) Deshpande, A., & Gupta, A. (2019). Nakusha? Son Preference, Resource Concentration and Gender Gaps in Education. Working Papers, Article 13. Ashoka University, Department of Economics. https://ideas.repec.org//p/ash/wpaper/13.html
18) Devan, P. (2012). Cricket and the global Indian identity. *Sport in Society,* 15(10), 1413–1425. https://doi.org/10.1080/17430437.2012.744210
19) Donati, Dante and Orozco-Olvera, Victor H. and Rao, Nandan Mark, Marketing Gender Norms: A Social Media Experiment in India (October 4, 2022). Columbia Business School Research Paper No. 4158065, Available at SSRN: https://ssrn.com/abstract=4158065 or http://dx.doi.org/10.2139/ssrn.4158065
20) Ferrara, E. L. (2016). Mass Media and Social Change: Can We Use Television to Fight Poverty? *Journal of the European Economic Association*, *14*(4), 791–827.
21) Jayachandran, S. (2015). The Roots of Gender Inequality in Developing Countries. *Annual Review of Economics*, *7*(1), 63–88. https://doi.org/10.1146/annurev-economics-080614-115404
22) Jayachandran, S., & Pande, R. (2017). Why Are Indian Children So Short? The Role of Birth Order and Son Preference. *American Economic Review*, *107*(9), 2600–2629. https://doi.org/10.1257/aer.20151282
23) Jejeebhoy, S. J., R. Acharya, S. Basu and A. J. F. Zavier. 2015. Addressing gender-biased sex selection in Haryana, India: Promising approaches. *New Delhi: Population Council*.
24) Jensen, R., & Oster, E. (2009). The Power of TV: Cable Television and Women's Status in India *. *Quarterly Journal of Economics*, *124*(3), 1057–1094. https://doi.org/10.1162/qjec.2009.124.3.1057
25) Jha, P., Kumar, R., Vasa, P., Dhingra, N., Thiruchelvam, D., & Moineddin, R. (2006). Low female-to-male sex ratio of children born in India: National survey of 1.1 million households. *Lancet (London, England)*, *367*(9506), 211–218. https://doi.org/10.1016/S0140-6736(06)67930-0
26) Kabeer, N., Huq, L., & Mahmud, S. (2014). Diverging Stories of "Missing Women" in South Asia: Is Son Preference Weakening in Bangladesh? *Feminist Economics,* 20(4), 138–163. https://doi.org/10.1080/13545701.2013.857423
27) Kidambi, P. (2011). Hero, celebrity and icon: Sachin Tendulkar and Indian public culture. In A. Bateman & J. Hill (Eds.), The Cambridge Companion to Cricket (pp. 187–202). Cambridge University Press. https://doi.org/10.1017/CCOL9780521761291.015
28) Kearney, M. S., & Levine, P. B. (2015). Media Influences on Social Outcomes: The Impact of MTV's 16 and Pregnant on Teen Childbearing. *American Economic Review*, *105*(12), 3597–3632. https://doi.org/10.1257/aer.20140012

29) Khanna, S. (2014, July 31). Twitter's Dick Costolo sees India turning into key business market in five years. *Mint.* https://www.livemint.com/Home-Page/Zrs2pMCKw5jHYezGCfrn2J/Twitters-Dick-Costolo-sees-India-turning-into-key-business.html
30) Kotler, P., & Zaltman, G. (1971). Social Marketing: An Approach to Planned Social Change. *Journal of Marketing*, *35*(3), 3–12. https://doi.org/10.1177/002224297103500302
31) Kishor, S. (1993). 'May God Give Sons to All': Gender and Child Mortality in India. *American Sociological Review*, *58*(2), 247–265. https://doi.org/10.2307/2095969
32) La Ferrara, E., Chong, A., & Duryea, S. (2012). Soap Operas and Fertility: Evidence from Brazil. *American Economic Journal: Applied Economics*, *4*(4), 1–31. https://doi.org/10.1257/app.4.4.1
33) Lapowsky, I. (n.d.). Your Old Tweets Give Away More Location Data Than You Think. *Wired*. Retrieved 21 January 2024, from https://www.wired.com/story/twitter-location-data-gps-privacy/
34) Michalopoulos, S., & Xue, M. M. (2021). Folklore*. *The Quarterly Journal of Economics*, *136*(4), 1993–2046. https://doi.org/10.1093/qje/qjab003
35) Mitter, S. (2015). How Twitter Changed Its Mind on India. *Forbes India*. https://www.forbesindia.com/article/big-bet/how-twitter-changed-its-mind-on-india/39391/1
36) Müller, K., & Schwarz, C. (2023). From Hashtag to Hate Crime: Twitter and Antiminority Sentiment. *American Economic Journal: Applied Economics*, *15*(3), 270–312. https://doi.org/10.1257/app.20210211
37) Ongo Nkoa, B. E., Ondoua Beyene, B., Ngo Nsoa Simb, J. F., & Ngnouwal Eloundou, G. (2023). Does social media improve women's political empowerment in Africa? *Telecommunications Policy*, *47*(9). https://doi.org/10.1016/j.telpol.2023.102624
38) Pande, R. P., & Astone, N. M. (2007). Explaining Son Preference in Rural India: The Independent Role of Structural versus Individual Factors. *Population Research and Policy Review*, *26*(1), 1–29.
39) Patel, A., Badhoniya, N., Mamtani, M., & Kulkarni, H. (2013). Skewed Sex Ratios in India: "Physician, Heal Thyself". *Demography*, *50*(3), 1129–1134.
40) Pearce, K. E., Donohoe, D., Barta, K., & Vitak, J. (2022). The Influence of Social Media Discussion on Son Preference in Azerbaijan: Reinforcing Norms, Bargaining With Patriarchy, Space for Dissent. *International Journal of Communication*, *16*(0), Article 0.
41) Pesando, L. M. (2022). Safer If Connected? Mobile Technology and Intimate Partner Violence. *Demography*, *59*(2), 653–684. https://doi.org/10.1215/00703370-9774978
42) Qian, N. (2008). Missing Women and the Price of Tea in China: The Effect of Sex-Specific Earnings on Sex Imbalance*. *The Quarterly Journal of Economics*, *123*(3), 1251–1285. https://doi.org/10.1162/qjec.2008.123.3.1251
43) Rajkhowa, P., & Qaim, M. (2022). Mobile phones, women's physical mobility, and contraceptive use in India. *Social Science & Medicine (1982)*, *305*, 115074. https://doi.org/10.1016/j.socscimed.2022.115074
44) Ramani, S. V. (2016, November 18). Why it's easier for India to get to Mars than to tackle its toilet challenge. *The Conversation*. http://theconversation.com/why-its-easier-for-india-to-get-to-mars-than-to-tackle-its-toilet-challenge-69015
45) The Pre-Natal Diagnostic Techniques (Regulation and Prevention of Misuse) Amendment Act, (1994).
46) The Pre-Natal Diagnostic Techniques (Regulation and Prevention Of Misuse) Act, (2002).
47) Rogers, E. M. (1995). *Diffusion of Innovation* (4th edition). The Free Press.
48) Sen, A. (1992). Missing Women. *BMJ : British Medical Journal*, *304*(6827), 587–588.
49) Shiller, R. J. (2017). Narrative Economics. *American Economic Review*, *107*(4), 967–1004. https://doi.org/10.1257/aer.107.4.967
50) Tabaie, S. (2017). Stopping female feticide in India: The failure and unintended consequence of ultrasound restriction. *Journal of Global Health*, *7*(1), 010304. https://doi.org/10.7189/jogh.07.010304
51) Times of India. (2014, November 22). Twitter: India among our fastest growing markets. *The Times of India*. https://timesofindia.indiatimes.com/tech-news/twitter-india-among-our-fastest-growing-markets/articleshow/45240322.cms
52) The Conversation: Twitter Trends 2022 India. (n.d.). Retrieved 21 January 2024, from https://marketing.twitter.com/en_apac/insights/the-conversation-twitter-trends-2022-in
53) Twitter. (2015, July 12). *The 2015 #YearOnTwitter in India*. https://blog.twitter.com/en_in/a/2015/the-2015-yearontwitter-in-india-in
54) Yasir, S. (2023, August 4). Where 'No One Wanted Girls,' a Dad in India Takes On the Patriarchy. *The New York Times.* https://www.nytimes.com/2023/08/04/world/asia/sunil-jaglan-india-haryana-womens-rights.html
55) Zhuravskaya, E., Petrova, M., & Enikolopov, R. (2020). Political Effects of the Internet and Social Media. *Annual Review of Economics*, *12*(1), 415–438. https://doi.org/10.1146/annurev-economics-081919-050239

6. Appendix

A1. Heterogeneity

A1.1 Heterogeneity Analysis Results

Table A1.1: Twitter Exposure and Preference for the Sex of the Child: Results from Heterogeneity Analysis for High/Low Scheduled Caste districts

VARIABLES	(1) First Stage (Twitter)	(2) Either	(3) Son	(4) Daughter
Panel A: First Stage Instrumental Variables, High Scheduled Castes (Female Population in District)				
Interest in Tendulkar (t=2013)	0.888***			
	(0.00555)			
Panel B: Instrumental Variables Estimation, High Scheduled Castes (Female Population in District)				
Twitter Exposure (t=2014, 2019)		0.00360***	-0.00568***	0.00236**
		(0.00138)	(0.00102)	(0.000988)
Observations	544,194	544,194	544,194	544,194
KP F-Stat	25633.01			
Anderson Rubin Wald Test	6.82			
Panel C: First Stage Instrumental Variables, Low Scheduled Castes (Female Population in District)				
Interest in Tendulkar (t=2013)	1.103***			
	(0.00954)			
Panel D: Instrumental Variables Estimation, Low Scheduled Castes (Female Population in District)				
Twitter Exposure (t=2014, 2019)		0.0139***	-0.0109***	-0.00411***
		(0.00162)	(0.00111)	(0.00107)
Observations	535,013	535,013	535,013	535,013
KP F-Stat	13387.47			
Anderson Rubin Wald Test	73.15			
State FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Weights	YES	YES	YES	YES
Controls	YES	YES	YES	YES

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 respectively. Survey weights are applied. IHS transformation applied to Twitter variables. All dependent variables are as IHS adjusted ratios. Anderson Rubin Wald test statistic from (2). Data for Scheduled Castes from Census 2001. High scheduled castes are those districts where baseline female scheduled caste population is greater than the country median.

Table A1.2: Twitter Exposure and Preference for the Sex of the Child: Results from Heterogeneity Analysis for North India

VARIABLES	(1) First Stage (Twitter)	(2) Either	(3) Son	(4) Daughter
Panel A: First Stage Instrumental Variables, North Indian				
Interest in Tendulkar (t=2013)	1.483*** (0.00514)			
Panel B: Instrumental Variables Estimation, North Indian				
Twitter Exposure (t=2014, 2019)		0.00678***	-0.00576***	-0.00157***
		(0.000844)	(0.000572)	(0.000536)
Observations	386,231	386,231	386,231	386,231
KP F-Stat	83281.45			
Anderson Rubin Wald Test	64.76			
Panel C: First Stage Instrumental Variables, Other State				
Interest in Tendulkar (t=2013)	0.735*** (0.00785)			
Panel D: Instrumental Variables Estimation, Other State				
Twitter Exposure (t=2014, 2019)		0.00981***	-0.00923***	-0.000590
		(0.00189)	(0.00134)	(0.00130)
Observations	692,976	692,976	692,976	692,976
KP F-Stat	8778.674			
Anderson Rubin Wald Test	26.84			
State FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Weights	YES	YES	YES	YES
Controls	YES	YES	YES	YES

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 respectively. Survey weights are applied. IHS transformation applied to Twitter variables. All dependent variables are as IHS adjusted ratios. Anderson Rubin Wald test statistic from (2). Data for North Indian measured at the state/union territory level, includes Chandigarh, Delhi, Haryana, Himachal Pradesh, Jammu and Kashmir, Ladakh, Punjab, Rajasthan, Uttar Pradesh, Uttarakhand

Table A1.3: Twitter Exposure and Preference for the Sex of the Child: Results from Heterogeneity Analysis for Religion

VARIABLES	(1) First Stage (Twitter)	(2) Either	(3) Son	(4) Daughter
Panel A: First Stage Instrumental Variables, Hindu				
Interest in Tendulkar (t=2013)	0.995*** (0.00683)			
Panel B: Instrumental Variables Estimation, Hindu				
Twitter Exposure (t=2014, 2019)		0.00867***	-0.00800***	-0.000886
		(0.00117)	(0.000830)	(0.000802)

Observations	826,478	826,478	826,478	826,478
KP F-Stat	3086.15			
Anderson Rubin Wald Test	5.01			

Panel C: First Stage Instrumental Variables, Non-Hindu

Interest in Tendulkar (t=2013)	0.757***			
	(0.0136)			

Panel D: Instrumental Variables Estimation, Non-Hindu

Twitter Exposure (t=2014, 2019)		0.00728**	-0.00781***	.000099
		(0.00325)	(0.00220)	(0.00212)
Observations	252,729	252,729	252,729	252,729
KP F-Stat	21196.02			
Anderson Rubin Wald Test	54.80			
State FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Weights	YES	YES	YES	YES
Controls	YES	YES	YES	YES

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 respectively. Survey weights are applied. IHS transformation applied to Twitter variables. All dependent variables are as IHS adjusted ratios. Anderson Rubin Wald test statistic from (2). Data for Hindu from DHS measured at the Individual level, Non-Hindu includes Muslim, Christian, Sikh, Buddhist/Neo-Buddhist, Jain, Jewish, Parsi/Zoroastrian, No Religion, Other.

Table A1.4: Twitter Exposure and Preference for the Sex of the Child: Results from Heterogeneity Analysis for Literacy

VARIABLES	(1) First Stage (Twitter)	(2) Either	(3) Son	(4) Daughter

Panel A: First Stage Instrumental Variables, High Baseline Literacy

Interest in Tendulkar (t=2013)	0.972***			
	(0.00934)			

Panel B: Instrumental Variables Estimation, High Baseline Literacy Districts

Twitter Exposure (t=2014, 2019)		0.00830***	-0.00729***	-0.00108
		(0.00168)	(0.00119)	(0.00116)
Observations	497,409	497,409	497,409	497,409
KP F-Stat	10814			
Anderson Rubin Wald Test	24.14			
State FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Weights	YES	YES	YES	YES
Controls	YES	YES	YES	YES

Panel C: First Stage Instrumental Variables, Low Baseline Literacy Districts

Interest in Tendulkar (t=2013)	1.044***

		(0.00453)		
Panel D: Instrumental Variables Estimation, Low Baseline Literacy Districts				
Twitter Exposure (t=2014, 2019)		0.00670***	-0.00644***	-0.000797
		(0.00110)	(0.000773)	(0.000730)
Observations	581,798	581,798	581,798	581,798
KP F-Stat	53201.44			
Anderson Rubin Wald Test	36.84			
State FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Weights	YES	YES	YES	YES
Controls	YES	YES	YES	YES

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 respectively. Survey weights are applied. IHS transformation applied to all dependent variables. All dependent variables are as IHS adjusted ratios. Anderson Rubin Wald test statistic from (2). Literacy retrieved from Census 2001, splitting the sample by median for low and high baseline literacy districts

Appendix A2: Fine-Tuned Model Outputs and Training Statistics

Table A2.1: Training and Validation Loss

Epoch	Training Loss	Validation Loss	Validation Accuracy	Training Time	Validation Time
1	0.627324	0.653364	0.643229	00:00:13	00:00:01
2	0.635627	0.599525	0.723958	00:00:13	00:00:01
3	0.595604	0.561084	0.723958	00:00:13	00:00:01
4	0.519925	0.498822	0.770833	00:00:13	00:00:01
5	0.359484	0.285268	0.882812	00:00:13	00:00:01
6	0.279843	0.292419	0.880208	00:00:13	00:00:01

Notes: The model for trained over 6 epochs with a learning rate of 5e-5 and a batch size of 32. An 80-20 split is used for training and validation respectively.

Table A2.2 Nature of Tweets – Indic Languages

2015		
Language	**Original Text**	**Translation [Approximate]**
Hindi	‘’ बेटी को जिसने मरवा दिया था पत्नी की कोख में, मोहल्ले में लड़किया ढूँढ रहा है नवरात्रे के कन्या भोज में।।। -बेटी बचाओ बेटी पढाओ	The one who killed his daughter in his wife's womb, is looking for girls in the neighbourhood in the Kanya Bhojan of Navratri.. - Save daughters, educate daughters
Hindi	हमे इसे देखना चाहिये ये ये आज भी कुछ लोगो की सोच ऐसी है हमें इसे बदलना होगा #IndiasDaughter	We should see this, even today some people think like this, we have to change it #IndiasDaughter
Telugu	పిల్లలే దేశానికి ఆస్తి బడులలో ఇంటర్నెట్ మే నాటికి టీచర్ల ఖాళీల భర్తీ --ప్రతిభా పురస్కార సభలో "	Children are the country's asset. Internet in schools, teacher vacancies to be filled by May - Pratibha Puraskar Sabha ‘’

Tamil	தேசிய விருது பெற்ற சிறுவர்களுக்கு " தங்க சங்கிலி #children #kakkamuttai	National Award-winning children receive 'gold chain' ‘’ #children #kakkamuttai
Telugu	సాముదాయ బడులపై చర్చిస్తాం విద్యాశాఖ కార్యదర్శి" హామీ!	We will discuss community schools, promises Education Secretary ‘’!
	2018	
Language	**Original Text**	**Translation [google]**
Tamil	பள்ளியில் தனக்கு இடம் தராவிட்டால் கமலாலய குளத்தில் விழுந்து இறந்து விடுவேன் என்று போராடி இடம் பெற்றவர் இறந்த பின் தனக்கான இடத்தை பெறவும் போராடி கொண்டிருக்கிறார அரசு மரியாதையும், தேசிய கொடியும் தேவையில்லை. கழக கொடி போர்த்தி அறிவாலயத்தில் உறங்கு தலைவா	If he is not given a place in the school, he will fall into the Kamalalaya pond and die. He who got a place after fighting and is fighting to get his place after his death. He does not need the respect of the government and the national flag. Sleep in the Arivalayam, leader, wrapped in the party flag.
Mixed script and code-mixed Hindi-English	Aur wo modern girls kaha hai? सुकन्याओं की भाषा पर जरूर गौर करें।	And where are those modern girls? Do pay attention to the language of the girls.
Gujarati	જે મા-બાપે પોતાના બાળકોને MR રસી નથી અપાવી બીકના મારે કે બીજા કોઇ કારણે તેમણે આ વાંચીને વહેલી તકે રસી અપાવવી જોઇએ. આવો સાથે મળી રુબેલા અને મીઝલ્સને ભારતથી વીદાય કરીએ અને આપણા બાળકોને સુરક્ષીત બનાવીએ. #‘’ #‘’ #‘’ #‘’ #‘’ #‘’	Parents who have not vaccinated their children with MR vaccine, whether due to fear or any other reason, should read this and get them vaccinated as soon as possible. Let's get rid of Rubella and Measles from India together and make our children safe. #‘’ #‘’ #‘’ #‘’ #‘’ #‘’
Mixed script and code-mixed Hindi-English	#DayofTheGirl बेटियां देश का भविष्य हैं। आइये हम सब यह प्राण लें कि देश का भविष्य हम अपने हाथों से संवारेंगे। अंतरराष्ट्रीय बालिका दिवस की सभी बेटियों को हार्दिक शुभकामनाऐं।	#DayofTheGirl Daughters are the future of the country. Let us all take a pledge that we will shape the future of the country with our own hands. Heartiest greetings to all the daughters on International Girl Child Day.
Mixed script and code-mixed Telugu-English	‘’ ‘’ 10th క్లాస్ exam pass results వచ్చిన పిల్లల్లా enjoy చేస్తున్నారు!!	‘’ ‘’ They are enjoying themselves like children who have passed their 10th class exam results!!

Figure A2.3 Themes and their Categories

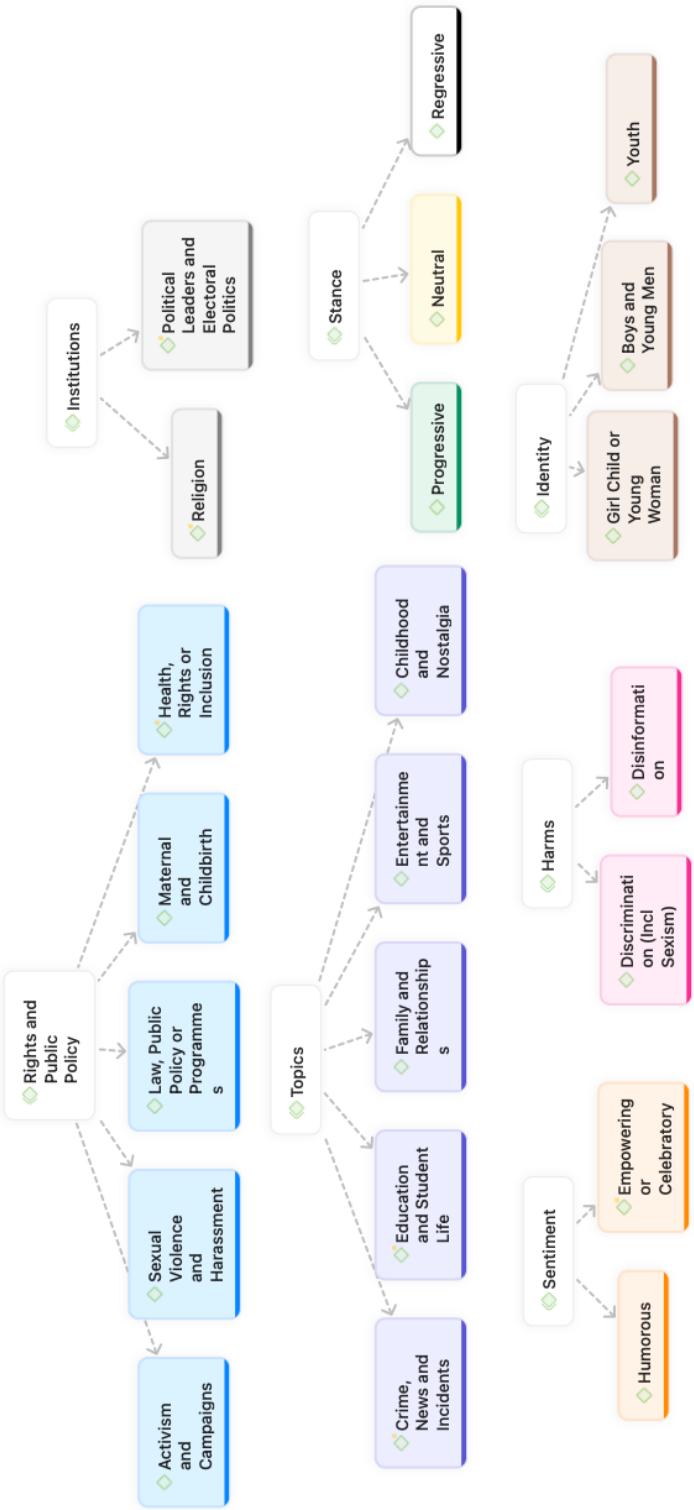
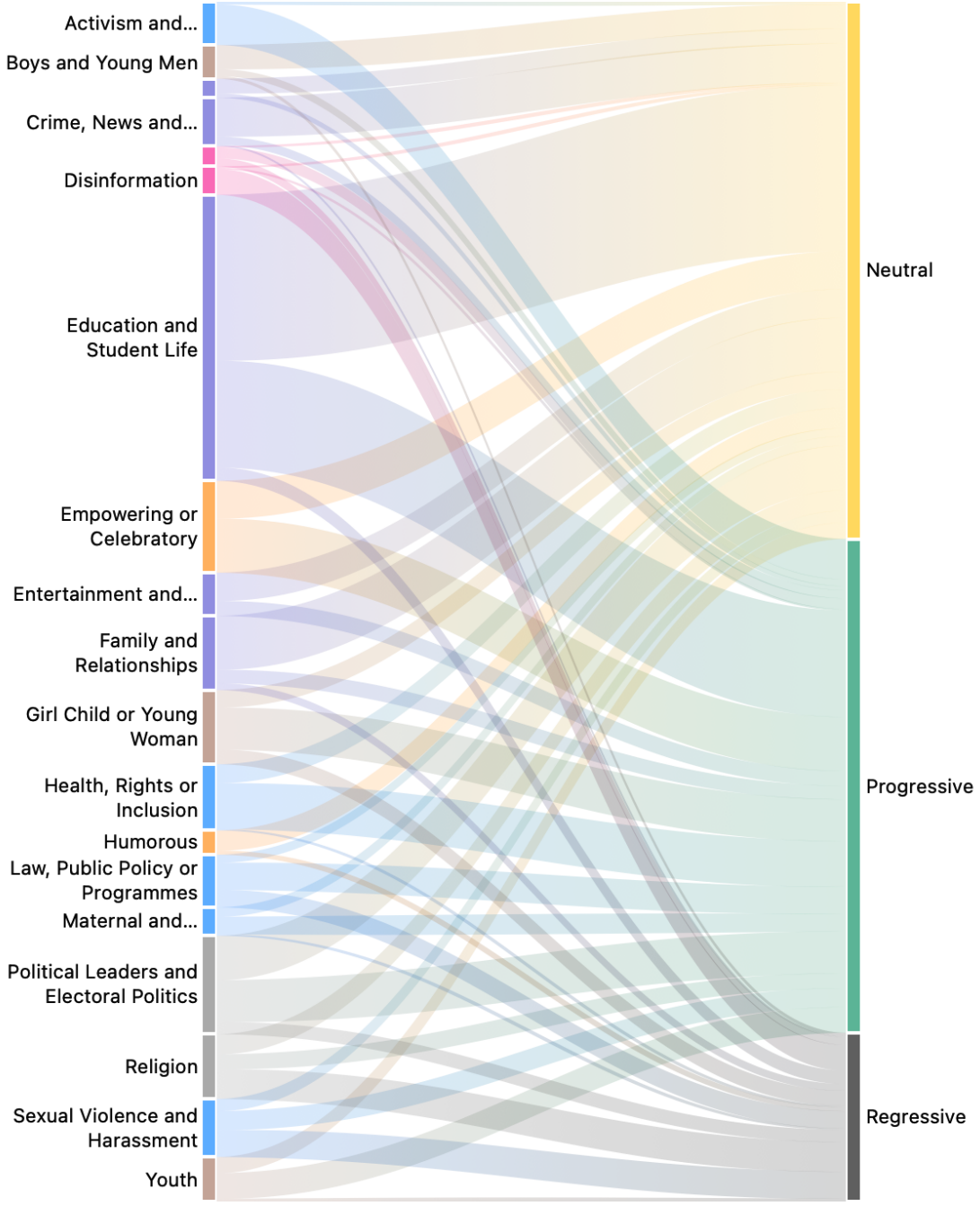


Figure A2.4 Sankey Diagram



Source: Co-occurrences of categories against stance using Twitter data, Author's elaboration via AtlasTI

Table A2.5: Definitions for Progressive, Regressive, and Neutral

Category	Definition
Unrelated	The Tweet is unrelated to children and has been erroneously included (e.g., the use of the term 'boys' to denote a sports team or using 'child' metaphorically)
Regressive	The document contains messages which cause harm, whether intentionally or unintentionally, reflect an opinion which may be harmful, critical of social or economic development, promote or perpetuate stereotypes, insult or cause injury to an individual or group, oppose progressive ideas.
Progressive	The document contains messages that seek to promote ideas or social change which benefits an individual, group, institution, promotes or discusses social change, opposes regressive ideas.
Neutral	Related to children and neither progressive nor regressive

A2.6: Further Analyses

The content of Tweets has also marginally changed over time (Table A2.6). Tweets in the 2018 corpus are categorically richer, i.e., they contain a greater frequency of categories per Tweet. In terms of stance. We find that Tweets are more neutral and less regressive, with a negligible increase (+0.03 p.p.) in progressivity. Relatively fewer discussions related to family and relationships (3-04 p.p.) and the 2018 corpus contains fewer instances of statements related to sexual violence and harassment (-2.96 p.p.) than the 2015 corpus. On the other hand, education and student life (+5.76 p.p.), and politics (+3.49 p.p) are two topics which have gained prominence in discourse related to children on Twitter. However, it is vital to note here that given the short time frames in which these Tweets were collected and the relatively larger character length of the 2018 Tweets[29], several of the topics presented here may be driven by themes relevant to current events at the time they were collected. Nonetheless, we find that most of the themes retain their significance in the discourse over the two time periods.

Table A2.6: Category Frequencies across Two time periods

	2015 N=250		2018 N=250		Totals N=500		Change	
Category	**Freq.**	**Perc.**	**Freq.**	**Perc.**	**Freq.**	**Perc.**	**Freq.**	**Perc. Pt.**
Activism and Campaigns	13	1.95%	19	2.64%	32	2.30%	6	0.69%
Boys and Young Men	13	1.95%	12	1.66%	25	1.80%	-1	-0.28%
Childhood and Nostalgia	4	0.60%	10	1.39%	14	1.01%	6	0.79%
Crime, News and Incidents	20	2.99%	16	2.22%	36	2.59%	-4	-0.77%
Discrimination (Incl Sexism)	6	0.90%	9	1.25%	15	1.08%	3	0.35%
Disinformation	17	2.54%	4	0.55%	21	1.51%	-13	-1.99%
Education and Student Life	82	12.28%	130	18.03%	212	15.26%	48	5.76%
Empowering or Celebratory	26	3.89%	42	5.83%	68	4.90%	16	1.93%
Entertainment and Sports	16	2.40%	16	2.22%	32	2.30%	0	-0.18%
Family and Relationships	37	5.54%	18	2.50%	55	3.96%	-19	-3.04%
Girl Child or Young Woman	30	4.49%	25	3.47%	55	3.96%	-5	-1.02%

[29] Twitter increased its character limit from 140 to 280 in 2017

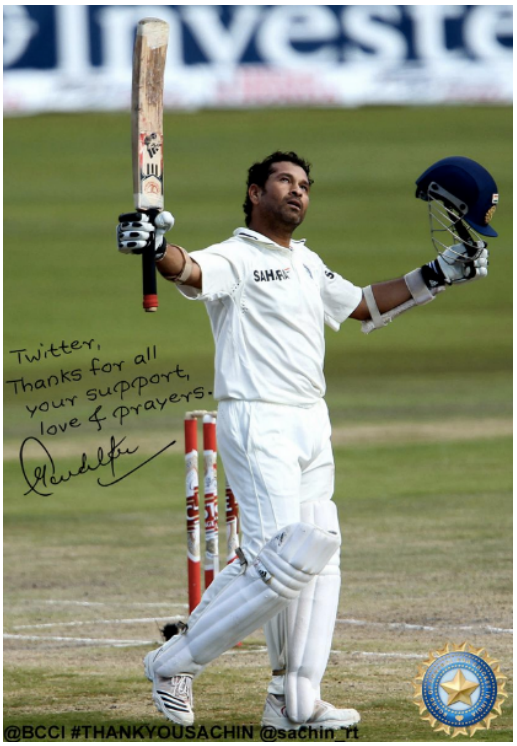
● **Health, Rights or Inclusion**	24	3.59%	25	3.47%	49	3.53%	1	-0.13%
● **Humorous**	8	1.20%	10	1.39%	18	1.30%	2	0.19%
● **Law, Public Policy or Programmes**	20	2.99%	19	2.64%	39	2.81%	-1	-0.36%
● **Maternal and Childbirth**	7	1.05%	14	1.94%	21	1.51%	7	0.89%
● **Neutral**	110	16.47%	125	17.34%	235	16.92%	15	0.87%
● **Political Leaders and Electoral Politics**	23	3.44%	50	6.93%	73	5.26%	27	3.49%
● **Progressive**	73	10.93%	79	10.96%	152	10.94%	6	0.03%
● **Regressive**	29	4.34%	14	1.94%	43	3.10%	-15	-2.40%
● **Religion**	29	4.34%	19	2.64%	48	3.46%	-10	-1.71%
● **Sexual Violence and Harassment**	30	4.49%	13	1.80%	43	3.10%	-17	-2.69%
○ **Unrelated**[30]	38	5.69%	32	4.44%	70	5.04%	-6	-1.25%
● **Youth**	13	1.95%	20	2.77%	33	2.38%	7	0.83%
Totals	668	100.00%	721	100.00%	1389	100.00%	53	

Note: the colour coding corresponds to frequencies and percentage point change as follows:

Lowest Frequency/ - P. change | Highest Frequency/+ P change

A3. Facts and Figures related to #ThankYouSachin

Figure A3.1: Automated "Digi-Graph" from the #ThankYouSachin Campaign on Twitter

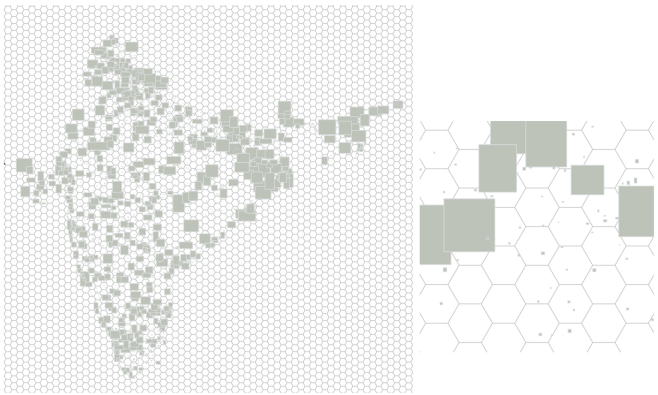


Source: Twitter India Trends Report, 2013

[30] See definitions in Appendix (Table A2.5)

Appendix A4: Geospatial Characteristics

Figure A4.1: Robustness: Map of Twitter in 2019 - Illustration of Bounding Boxes and Fishnet Grids



Source: The map plots the geographic spread of Twitter in 2019, Authors' calculation using Twitter API metadata

A4.2 Districts

For regressions with controls, the following districts are included; for regressions without controls we consider all districts as per 2015 DHS

N	District Names	N	District Names	N	District Names	N	District Names	N	District Names
1	Adilabad	116	Deoghar	231	Kanpur Nagar	346	Palamu	461	Tawang
2	Agra	117	Deoria	232	Kanshiram Nagar	347	Pali	462	Tehri Garhwal
3	Ahmadabad	118	Dewas	233	Kapurthala	348	Palwal	463	Thane
4	Ahmadnagar	119	Dhalai	234	Karauli	349	Panchmahal	464	Thanjavur
5	Aizawl	120	Dhanbad	235	Karbi Anglong	350	Panipat	465	The Dangs
6	Ajmer	121	Dhar	236	Karimganj	351	Parbhani	466	Theni
7	Akola	122	Dharwad	237	Karimnagar	352	Paschim Medinipur	467	Thiruvallur
8	Alappuzha	123	Dhaulpur	238	Karnal	353	Pashchim Champaran	468	Thiruvananthapuram
9	Aligarh	124	Dhemaji	239	Karur	354	Pashchimi Singhbhum	469	Thiruvarur
10	Alirajpur	125	Dhenkanal	240	Kasaragod	355	Patan	470	Thoothukkudi
11	Allahabad	126	Dhubri	241	Kathua	356	Pathanamthitta	471	Thoubal
12	Almora	127	Dhule	242	Katihar	357	Patiala	472	Thrissur
13	Alwar	128	Dibang Valley	243	Katni	358	Patna	473	Tikamgarh
14	Ambala	129	Dibrugarh	244	Kaushambi	359	Perambalur	474	Tinsukia
15	Ambedkar Nagar	130	Dimapur	245	Kendrapara	360	Peren	475	Tirap
16	Amravati	131	Dindori	246	Kendujhar	361	Phek	476	Tiruchirappalli
17	Amreli	132	Dohad	247	Khagaria	362	Pilibhit	477	Tirunelveli
18	Amritsar	133	Dumka	248	Khammam	363	Pithoragarh	478	Tiruppur
19	Anantapur	134	Dungarpur	249	Khandwa (East Nimar)	364	Porbandar	479	Tiruvannamalai
20	Anantnag	135	East District	250	Khargone (West Nimar)	365	Prakasam	480	Tonk

21	Anugul	136	East Godavari	251	Kheri	366	Pratapgarh	481	Tuensang
22	Anuppur	137	East Kameng	252	Khunti	367	Pudukkottai	482	Tumkur
23	Araria	138	East Khasi Hills	253	Kishanganj	368	Pune	483	Udaipur
24	Ariyalur	139	East Siang	254	Koch Bihar	369	Purba Champaran	484	Udalguri
25	Arwal	140	Ernakulam	255	Kodagu	370	Purba Medinipur	485	Udham Singh Nagar
26	Ashoknagar	141	Erode	256	Kodarma	371	Purbi Singhbhum	486	Udhampur
27	Auraiya	142	Etah	257	Kohima	372	Puri	487	Udupi
28	Aurangabad	143	Etawah	258	Kolar	373	Purnia	488	Ujjain
29	Azamgarh	144	Faizabad	259	Kolhapur	374	Puruliya	489	Ukhrul
30	Badgam	145	Faridabad	260	Kolkata	375	Rae Bareli	490	Umaria
31	Bagalkot	146	Faridkot	261	Kollam	376	Raichur	491	Una
32	Bageshwar	147	Farrukhabad	262	Koppal	377	Raisen	492	Unnao
33	Baghpat	148	Fatehabad	263	Koraput	378	Rajgarh	493	Upper Subansiri
34	Bahraich	149	Fatehpur	264	Kota	379	Rajkot	494	Uttar Dinajpur
35	Balaghat	150	Firozabad	265	Kottayam	380	Rajsamand	495	Uttara Kannada
36	Balangir	151	Firozpur	266	Kozhikode	381	Ramanathapuram	496	Vadodara
37	Baleshwar	152	Gadchiroli	267	Krishna	382	Ramgarh	497	Vaishali
38	Ballia	153	Gajapati	268	Kulgam	383	Rampur	498	Valsad
39	Balrampur	154	Ganderbal	269	Kullu	384	Ranchi	499	Varanasi
40	Banaskantha	155	Gandhinagar	270	Kupwara	385	Rangareddy	500	Vellore
41	Banda	156	Ganganagar	271	Kurnool	386	Ratlam	501	Vidisha
42	Bandipore	157	Ganjam	272	Kurukshetra	387	Ratnagiri	502	Viluppuram
43	Bangalore	158	Garhwa	273	Kushinagar	388	Rayagada	503	Virudhunagar
44	Banka	159	Garhwal	274	Lahul and Spiti	389	Reasi	504	Visakhapatnam
45	Bankura	160	Gaya	275	Lakhimpur	390	Rewa	505	Vizianagaram
46	Banswara	161	Ghaziabad	276	Lakhisarai	391	Rewari	506	Warangal
47	Bara Banki	162	Ghazipur	277	Latehar	392	Rohtak	507	Wardha
48	Baramula	163	Giridih	278	Latur	393	Rohtas	508	Washim
49	Baran	164	Goalpara	279	Lawngtlai	394	Rudraprayag	509	Wayanad
50	Barddhaman	165	Godda	280	Lohit	395	Rupnagar	510	West District
51	Bareilly	166	Golaghat	281	Longleng	396	Sabarkantha	511	West Garo Hills
52	Bargarh	167	Gonda	282	Lower Subansiri	397	Sagar	512	West Godavari
53	Barmer	168	Gondiya	283	Lucknow	398	Saharanpur	513	West Kameng
54	Barnala	169	Gopalganj	284	Ludhiana	399	Saharsa	514	West Khasi Hills
55	Basti	170	Gorakhpur	285	Lunglei	400	Sahibganj	515	West Siang
56	Bathinda	171	Gulbarga	286	Madhepura	401	Saiha	516	West Tripura
57	Begusarai	172	Gumla	287	Madhubani	402	Salem	517	Wokha
58	Belgaum	173	Guna	288	Madurai	403	Samastipur	518	Y.S.R.
59	Bellary	174	Guntur	289	Mahbubnagar	404	Sambalpur	519	Yadgir
60	Betul	175	Gurdaspur	290	Mahendragarh	405	Sangli	520	Yamunanagar
61	Bhadrak	176	Gurgaon	291	Mahesana	406	Sangrur	521	Yavatmal
62	Bhagalpur	177	Gwalior	292	Mahrajganj	407	Sant Ravidas Nagar (Bhadohi)	522	Zunheboto
63	Bhandara	178	Hamirpur	293	Mainpuri	408	Saraikela Kharsawan		
64	Bharatpur	179	Hanumangarh	294	Malappuram	409	Saran		
65	Bharuch	180	Haora	295	Maldah	410	Satara		
66	Bhavnagar	181	Harda	296	Malkangiri	411	Satna		
67	Bhilwara	182	Hardoi	297	Mandi	412	Sawai Madhopur		
68	Bhind	183	Hardwar	298	Mandla	413	Sehore		
69	Bhiwani	184	Hassan	299	Mandya	414	Senapati (Excluding 3 Sub-Divisions)		
70	Bhojpur	185	Hazaribagh	300	Mansa	415	Seoni		
71	Bhopal	186	Hingoli	301	Mathura	416	Shahdol		
72	Bid	187	Hisar	302	Mau	417	Shahjahanpur		
73	Bidar	188	Hoshangabad	303	Mayurbhanj	418	Shajapur		
74	Bijapur	189	Hoshiarpur	304	Medak	419	Sheikhpura		
75	Bijnor	190	Hugli	305	Meerut	420	Sheopur		

76	Bikaner	191	Hyderabad	306	Mirzapur	421	Shimla
77	Birbhum	192	Idukki	307	Moga	422	Shimoga
78	Bishnupur	193	Indore	308	Mokokchung	423	Shivpuri
79	Bokaro	194	Jabalpur	309	Mon	424	Shrawasti
80	Budaun	195	Jagatsinghapur	310	Moradabad	425	Siddharth Nagar
81	Bulandshahr	196	Jaintia Hills	311	Morena	426	Sidhi
82	Buldana	197	Jaipur	312	Morigaon	427	Sikar
83	Bundi	198	Jaisalmer	313	Muktsar	428	Simdega
84	Burhanpur	199	Jajapur	314	Mumbai	429	Sindhudurg
85	Buxar	200	Jalandhar	315	Mumbai Suburban	430	Singrauli
86	Cachar	201	Jalaun	316	Munger	431	Sirmaur
87	Chamarajanagar	202	Jalgaon	317	Murshidabad	432	Sirohi
88	Chamba	203	Jalna	318	Muzaffarnagar	433	Sirsa
89	Chamoli	204	Jalor	319	Muzaffarpur	434	Sitamarhi
90	Chandauli	205	Jalpaiguri	320	Mysore	435	Sitapur
91	Chandel	206	Jamnagar	321	Nabarangapur	436	Sivaganga
92	Chandrapur	207	Jamtara	322	Nadia	437	Sivasagar
93	Changlang	208	Jamui	323	Nagaon	438	Siwan
94	Chhatarpur	209	Jaunpur	324	Nagapattinam	439	Solan
95	Chhindwara	210	Jehanabad	325	Nagaur	440	Solapur
96	Chikkaballapura	211	Jhabua	326	Nagpur	441	Sonbhadra
97	Chikmagalur	212	Jhalawar	327	Nainital	442	Sonipat
98	Chitradurga	213	Jhansi	328	Nalanda	443	Sonitpur
99	Chitrakoot	214	Jharsuguda	329	Nalgonda	444	South
100	Chittaurgarh	215	Jhunjhunun	330	Namakkal	445	South District
101	Chittoor	216	Jind	331	Nanded	446	South Goa
102	Churachandpur	217	Jodhpur	332	Nandurbar	447	South Tripura
103	Churu	218	Jorhat	333	Narmada	448	South Twenty Four Parganas
104	Coimbatore	219	Junagadh	334	Narsimhapur	449	Sri Potti Sriramulu Nellore
105	Cuddalore	220	Jyotiba Phule Nagar	335	Nashik	450	Srikakulam
106	Cuttack	221	Kachchh	336	Navsari	451	Srinagar
107	Dakshin Dinajpur	222	Kaimur (Bhabua)	337	Nawada	452	Subarnapur
108	Dakshina Kannada	223	Kaithal	338	Nayagarh	453	Sultanpur
109	Damoh	224	Kalahandi	339	Nizamabad	454	Sundargarh
110	Darbhanga	225	Kancheepuram	340	North Goa	455	Supaul
111	Darjiling	226	Kangra	341	North Tripura	456	Surat
112	Darrang	227	Kannauj	342	North Twenty Four Parganas	457	Surendranagar
113	Datia	228	Kanniyakumari	343	Nuapada	458	Tamenglong
114	Dausa	229	Kannur	344	Pakur	459	Tapi
115	Dehradun	230	Kanpur Dehat	345	Palakkad	460	Tarn Taran

A4.3 Construction of Dataset

All changes in district boundaries and names were identified through multiple data sources, such as census data, district webpages and other official sources where available. For identifying controls at baseline, we match baseline statistics to districts which were later bifurcated e.g., Midnapore or West Dinajpur based on their original district boundaries when the data was collected. E.g., for literacy Uttar and Dakshin Dinajpur would have the same value if the data was collected when these old boundaries were in place.

West Bengal			
New Name	**Old Name**	**Year**	**Action**
Uttar dinajpur	West Dinajpur	1992	Bifurcated and Different Name
Dakshin dinajpur	West Dinajpur	1992	Bifurcated and Different Name
Kolkata	Calcutta	2001	Name Changed
Paschim medinipur	Midnapore	2002	Bifurcated and Different Name
Purba medinipur	Midnapore	2002	Bifurcated and Different Name

Table A 4.2: Reorganisation of example districts in the state of West Bengal

Due to extensive reorganisation, several parts of the Bodoland region in Assam and Chhattisgarh could not be included in regressions with controls.

Appendix A5: Robustness

Table A5.1: Twitter Density, Distance and Preference for the Sex of the Child: Results from Sensitivity Analysis, OLS with Controls.

VARIABLES	(1) Either	(2) Son	(3) Daughter
Panel A: OLS Estimations with Controls – Density			
Twitter Density (t=2019)	0.00450*** (0.000233)	-0.00289*** (0.000162)	-0.00215*** (0.000157)
Observations	1,079,207	1,079,207	1,079,207
R-squared	0.039	0.048	0.013
Panel B: OLS Estimations with Controls – Distance			
Twitter Distance (t=2019)	-0.0126*** (0.000491)	0.00916*** (0.000340)	0.00494*** (0.000329)
Observations	1,079,207	1,079,207	1,079,207
R-squared	0.040	0.048	0.013
State FE	YES	YES	YES
Year FE	YES	YES	YES

Notes: Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1 respectively. Survey weights are applied. IHS transformation applied to all dependent variables. Twitter Distance is a transformed measure of distance in kilometres from nearest hub. Density refers to the exposure to Twitter for those belonging in an artificial fishnet grid. Anderson Rubin Wald test statistic from (2).

Table A5.2: Twitter Density, Distance and Preference for the Sex of the Child: Results from Sensitivity Analysis, IV

VARIABLES	(1) First Stage (Twitter)	(2) Either	(3) Son	(4) Daughter

Panel A: First Stage Instrumental Variables for Density				
Interest in Sachin Tendulkar (t=2013)	0.517***			
	(0.00914)			
Panel B: Instrumental Variables Estimation				
Twitter Density in Grid (t=2019)		0.0154***	-0.0146***	-0.00126
		(0.00203)	(0.00144)	(0.00138)
Observations	1,079,207	1,079,207	1,079,207	1,079,207
KP F-Stat	3203.56			
Anderson Rubin Wald Test	56.93			
Panel C: First Stage Instrumental Variables for Distance				
Interest in Sachin Tendulkar (t=2013)	-0.277***			
	(0.00286)			
Panel D: Instrumental Variables Estimation for Distance				
Twitter Distance, in km (t=2019)		-0.113***	0.107***	0.00922
		(0.0154)	(0.0112)	(0.0101)
Observations	1,079,207	1,079,207	1,079,207	1,079,207
KP F-Stat	550.36			
Anderson Rubin Wald Test	56.93			
State FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Weights	YES	YES	YES	YES
Controls	YES	YES	YES	YES

Notes: Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1 respectively. Survey weights are applied. IHS transformation applied to all dependent variables. Twitter Distance is a transformed measure of distance in kilometres from nearest hub. Density refers to the exposure to Twitter for those belonging in an artificial fishnet grid. Anderson Rubin Wald test statistic from (2).

Table A5.3: IV regressions linking Twitter and Preferences for the Sex of the Child, Different Clustering Strategies

	(1)	(2)	(3)	(4)
VARIABLES	First Stage (Twitter)	Either	Son	Daughter
Panel A: First Stage Instrumental Variables, PSU Cluster				
Interest in Tendulkar (t=2013)	0.950***			
	(0.0391)			
Panel B: Instrumental Variables Estimation, PSU Cluster				
Twitter Exposure (t=2014,2019)		0.00836***	-0.00794***	-0.000685
		(0.00227)	(0.00150)	(0.00127)
Observations	1,079,207	1,079,207	1,079,207	1,079,207

KP F-Stat	590.510			
Anderson Rubin Wald Test	13.24			
Panel C: First Stage Instrumental Variables, State Cluster				
Interest in Tendulkar (t=2013)	0.950***			
	(0.172)			
Panel D: Instrumental Variables Estimation, State Cluster				
Twitter (t=2014,2019)		0.00836***	-0.00794***	-0.000685
		(0.00296)	(0.00184)	(0.00165)
Observations	1,079,207	1,079,207	1,079,207	1,079,207
KP F-Stat	30.49			
Anderson Rubin Wald Test	7.01			
State FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Weights	YES	YES	YES	YES
Controls	YES	YES	YES	YES

Notes: PSU (Panels A-B) and State (Panels C-D) standard errors in parentheses. PSU weights are geomatched across DHS rounds i.e., round 1to nearest neighbours to round 2. *** p<0.01, ** p<0.05, * p<0.1 respectively. Survey weights are applied. IHS transformation applied to all dependent variables. All dependent variables are as IHS adjusted ratios. Anderson Rubin Wald test statistic from (2).

Table A5.4: IV regressions linking Twitter and Preferences for the Sex of the Child, Different Weight Strategies

	(1)	(2)	(3)	(4)
VARIABLES	First Stage (Twitter)	Either	Son	Daughter
Panel A: First Stage Instrumental Variables, No Weight				
Interest in Tendulkar (t=2013)	1.214***			
	(0.00336)			
Panel B: Instrumental Variables Estimation, No Weight				
Twitter Exposure (t=2014, 2019)		0.00833***	-0.00695***	-0.00191***
		(0.000586)	(0.000406)	(0.000388)
Observations	1,079,207	1,079,207	1,079,207	1,079,207
KP F-Stat	130810.86			
Anderson Rubin Wald Test	202.31			
Panel C: First Stage Instrumental Variables, Married Women Weight				
Interest in Tendulkar (t=2013)	1.267***			
	(0.00425)			
Panel D: Instrumental Variables Estimation, Married Women Weight				
Twitter (t=2014, 2019)		0.00870***	-0.00736***	-0.00181***
		(0.000650)	(0.000462)	(0.000442)
Observations	1,079,207	1,079,207	1,079,207	1,079,207
KP F-Stat	88856.20			
Anderson Rubin Wald Test	179.25			
State FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Weights	YES	YES	YES	YES
Controls	YES	YES	YES	YES

Notes: Robust) standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 respectively. No weights are applied for Panels A-B; weights in panels C-D are retrieved from the 2001 census of India as the count of ever marred women in the district from Tables F-6 . IHS transformation applied to all dependent variables. All dependent variables are as IHS adjusted ratios. Anderson Rubin Wald test statistic from (2).

Table A5.5. IV Results without Controls

	(1) First Stage (Twitter)	(2) Either	(3) Son	(4) Daughter
Panel A: First Stage Instrumental Variables, No Controls				
Interest in Sachin Tendulkar (t=2013)	1.524***			
	(0.00596)			
Panel A: Instrumental Variables Estimation, No Controls				
Twitter Diffusion (t=2014. 2019)		0.0149***	-0.0117***	-0.00448***
		(0.000587)	(0.000411)	(0.000397)
Observations	1,291,869	1,291,869	1,291,869	1,291,869
KP F-Stat	65379.95			
Anderson Rubin Wald test	639.50			
State FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Weights	YES	YES	YES	YES
Controls	NO	NO	NO	NO

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 respectively. Survey weights are applied. IHS transformation applied to all dependent variables. All dependent variables are as IHS adjusted ratios. Anderson Rubin Wald test statistic from (2). The discrepancy between the sample sizes is a result of matching loss for districts in 1998-99 DHS and 2001 Census from which district level control variables are sought.

Table A5.6. IV regressions without Controls – Mechanisms (Selfie with Daughter)

VARIABLES	(1) First Stage (Twitter)	(2) Either	(3) Son	(4) Daughter
Panel A: First Stage Instrumental Variables without Controls				
Interest in Tendulkar (t=2013)	0.831***			
	(0.0281)			
Panel B: Instrumental Variables Estimation, #SelfieWithDaughter without controls				
Twitter Exposure (t=2014)		0.0373***	-0.0351***	-0.00342
		(0.00555)	(0.00427)	(0.00400)
Observations	84,068	84,068	84,068	84,068
KP F-Stat	872.931			
Anderson Rubin Wald Test	46.06			
Panel A: First Stage Instrumental Variables				

Interest in Tendulkar (t=2013)	0.0439***			
	(0.0150)			
State FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Controls	NO	NO	NO	NO

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 respectively. Survey weights are applied. IHS transformation applied to Twitter variables. All dependent variables are as IHS adjusted ratios. Anderson Rubin Wald test statistic from (2). Data for Selfie With Tweets from 2015 Twitter API

Table A5.7: IV regressions with Three Rounds of the DHS (1998-1999, 2015-2016, 2019-2021)

	(1)	(2)	(3)	(4)
VARIABLES	First Stage (Twitter)	Either	Son	Daughter
Panel A: First Stage Instrumental Variables				
Interest in Tendulkar (t=2013)	1.007***			
	(0.00616)			
State FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Controls	NO	NO	NO	NO
Panel B: Instrumental Variables Estimation, #SelfieWithDaughter				
Twitter (t=2014)		0.00659***	-0.00639***	-0.000404
		(0.00103)	(0.000723)	(0.000698)
Observations	1,145,226	1,145,226	1,145,226	1,145,226
KP F-Stat	67876.44			
Anderson Rubin Wald Test	761.03			
State FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Controls	YES	YES	YES	YES

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1 respectively. Survey weights are applied. IHS transformation applied to Twitter variables. All dependent variables are as IHS adjusted ratios. Anderson Rubin Wald test statistic from (2). Data for Selfie With Tweets from 2015 Twitter API