Online News in India: Appraising the Digital News Consumption Landscape in the
World's Largest Democracy (2014-2018)

Subhayan Mukerjee*

National University of Singapore

*Corresponding author: mukerjee@nus.edu.sg

Abstract

How do people in the world's largest democracy consume news online? In this paper, I aim to answer this question by conducting a quantitative assessment of the online news consumption behavior of a large sample of Indian internet users ($N \approx 50,000$) over a period of 45 months. In doing so, I contribute to theoretical debates about global news media use, by systematically appraising the prominence and trends in audience share of different types of news sources, thereby shedding light on the digital news consumption landscape of a crucial, but understudied context. Theoretically, I engage with the displacement-complementarity hypothesis and find no evidence that digital-born media have contested the hegemony of legacy media in India online. Next, I investigate the regional-national media divide and find that regional, vernacular media have suffered significant declines in their audience shares over time. This begs the question whether the notion of "polycentrism" – the idea that the Indian media environment is comprised of national and regional media of equivalent weight – is at all applicable online as it is offline. These findings also run counter to claims of "internet vernacularization" that have been touted in the past. Finally, I propose the concept of audience mobility, and use it to identify qualitatively distinct dynamics in how vernacular audiences in India have migrated to national vis-à-vis international outlets. The findings and their implications are discussed in light of contemporary changes in Indian society that is characterized by rapid digitization and increasing literacy.

Keywords: news consumption, online news, audience studies, digital media, India

Online News in India: Appraising the Digital News Consumption Landscape in the
World's Largest Democracy (2014-2018)

The advent of the Internet and the rise of digital platforms have had a profound impact on how people consume news around the world. Consequently a large body of literature emerged at the turn of the 21st century that sought to understand the nature of this transformation (for a review, see Mitchelstein & Boczkowski, 2010). Scholars pondered the extent to which the rise of online news would displace traditional media consumption habits. Others argued how online news would instead complement existing news reading behavior. Buoyed in part by optimistic speculation about the supposed democratizing potential of the internet, many scholars celebrated how non-traditional media could flourish, and challenge the hegemony of legacy media (Croteau, 2006; Deuze, 2003; Meraz, 2009). Yet others cautioned how a high-choice information environment like the internet could give rise to hyper-partisan media outlets, and exacerbate societal problems such as polarization and information inequalities (Hindman, 2009; Sunstein, 2009).

Yet, even after two decades of research, academic consensus on online news consumption rests largely on evidence gathered from western democratic contexts such as the US and western Europe. The global south, which is home to some of the largest online markets in the world, remains understudied, even as theories of media use have evolved with the state of the art of news consumption research. Such a preponderance of empirical evidence from western contexts not only undermines the more global aspirations of the field of media studies (Thussu, 2009), but also paints a parochial picture of online news consumption, and more generally, the new media environment, as a socio-political phenomenon restricted to certain countries only.

This paper tries to take a step in contributing towards the “dewesternization” (Gunaratne, 2010; Iwabuchi, 2014; Waisbord & Mellado, 2014) of the field by focusing on India, a country that remains quantitatively understudied in the literature. From an academic perspective, this dearth of evidence about India is concerning: India is on the one hand, the largest democracy in the world, and news has

often been regarded as the “currency of citizenship” (Delli Carpini & Keeter, 1996, p. 8), as important to democratic politics, as “money is to economics”. The lack of empirical evidence from India and other democracies in the global south thus implies that theory building exercises about how people navigate the new media landscape have largely relied on the experiences of *certain* democracies only. This undermines not just scholarship about the so-called “new” media environment, but also media studies more generally. Not only does India’s complex media system “not fit neatly into the framework developed by Hallin and Mancini (2004)” (Neyazi & Schroeder, 2021, pg. 88) but, as Chakravartty and Roy (2013) argue, the theoretical insights obtained from even comparative media research are often “of limited utility when it comes to explaining the political influence of news beyond what some might identify as “exceptionalism” or “provincialism” of Eurocentric models of comparison” (pg. 350).

This paper aims to contribute to theoretical debates about digital media use in the global South, by shedding empirical light on the news consumption landscape of Indian internet users. Thus, the “dewesternization” that this paper attempts to engage in, is not directly theoretical but empirical. However, the evidence speaks to the long line of theoretical reasoning that has characterized the literature on online news consumption, and helps triangulate the phenomenon as a socio-political phenomenon that is holistic and more global. I draw on the slim available literature on India’s news consumption to motivate the need to understand how online news consumption patterns in the country have changed over time and pose a set of hypotheses. I then introduce the dataset I use to test these hypotheses. In subsequent sections, I describe my findings: contrary to how Chakravartty and Roy (2013) described the television media landscape in India, as one having “multiple (national and regional) media systems of equivalent weight”, I find systematic evidence regarding the disproportionate prominence of national English-language media outlets, compared to regional vernacular media outlets, in so far as online audience reach is concerned. Longitudinal analysis shows that counter Neyazi’s notion of “internet vernacularization” (2019), audiences of online vernacular outlets have gradually diminished over time as well.

Further unpacking their dwindling audiences reveals insights into how Indian audiences are navigating the legacy/digital-born divide online in migrating to national and international news websites from vernacular news websites. I propose the concept of “audience mobility” – the progressive migration over time of parts of an audience from one type of media to another – to better interpret these migrations. I conclude by discussing the implications of an increasingly nationalized and “unified” news consumption landscape, with a few gravity-centers commanding most of the online audience attention. This research endeavor thus constitutes the largest systematic study of online news use in India – the second largest online population in the world, and reveals novel evidence of how $\approx 15\%$ of the global online audience consumes news.

Online News Consumption in India

A survey of early research on online news consumption reveals several paradoxical findings about the relationship of online news media with so-called “legacy” news media that predated the internet. A prominent debate in the early literature relates to whether “digital-born” media have complemented or displaced “legacy” media thereby potentially re-configuring hierarchies of power in the media environment – and whether the manner in which people consume online news reflects these changes (Chan & Leung, 2005; Chyi & Lasorsa, 2002; Croteau, 2006; Deuze, 2003; Gentzkow, 2007; Kayany & Yelsma, 2000; Kaye & Johnson, 2003; Lin, Salwen, Garrison, & Driscoll, 2005; Meraz, 2009; Nguyen & Western, 2007). This debate about displacement or complementarity has engendered various theoretical quandaries regarding the “fragmentation” of news media diets, many of which have been subjected to empirical validation (for e.g. Mukerjee, Majó-Vázquez, & González-Bailón, 2018; Yang, Majó-Vázquez, Nielsen, & González-Bailón, 2020) under monikers such as “selective exposure” (Spohr, 2017) and “echo-chambers” (Dubois & Blank, 2018; Guess, Nyhan, Lyons, & Reifler, 2018; Sunstein, 2017). However, most of this literature is western centric, even as the debate about the power-balance between legacy and digital-born media has important ramifications about the future of media in the global south, including in India, as well.

Displacement-Complementarity in Online Indian News Media

Testing the displacement-complementarity hypothesis in the context of online Indian media is important because Indian media's transition from the print and the broadcast to the digital started later than it did in the west, and happened at a slower rate (see Figure 1). While India has since witnessed a significant surge in the number of online news outlets, it has primarily been driven by legacy media outlets (newspapers and television networks) embracing digital platforms (such as *The Hindu* in 1995 and *The Times of India* in 1999) ¹.

Following the “IT boom” (Kapur, 2002) in India in the early 21st century, a large number of digital-born outlets were founded, as more and more of a rapidly digitizing population turned online for their news. Whether this transformation was accompanied by any substantial changes in the power-structures of the Indian media system still remains an open question that, while being the subject of speculation (Brookings India, 2015; Chaudhury, 2016; Wang, 2016), has never been investigated empirically.

Figure 1. India's growth in internet penetration compared with western countries.

Source: Internet Telecommunication Union of the United Nations

¹ though some exceptions exist, such as *Rediff* an early technology company, founded in 1996 that transitioned to a news portal

An important point of distinction that sets Indian media apart from western media models is the continued growth in the newspaper and broadcast space, even as online news has steadily become prominent (*Press in India 2019-20*, 2020). Notably, this shift to online news consumption has largely been driven by the youth (under-35) population who report consuming news online (56%) significantly more than offline (print 16%, television, 24%) (Aneez, Neyazi, Kalogeropoulos, & Nielsen, 2019). However, while online news has certainly become more popular in India, it is specifically the online offerings of *legacy brands* that people seem to prefer: even though a few digital-born outlets feature in respondents' self-reports, they are outnumbered by legacy outlets like *NDTV* and *The Times of India*. Another Reuters report sought to explore how online outlets competed over the provision of news during the 2019 General Election and found that digital-born outlets were not only more active in posting news stories, but that they also enjoyed greater engagement (relative to their follower base) than legacy outlets on social media (Majó-Vázquez, Mukerjee, Neyazi, & Nielsen, 2019). Beyond these reports, there exists very little scholarly work that tests the displacement-complementarity hypothesis in the context of Indian media, and those remain limited to comparing online and offline media, not legacy and digital-born outlets within the online space: Neyazi and colleagues (2019), for example, found evidence of traditional media use predicting social media use during the 2014 Indian elections. Elsewhere, Neyazi and Schroeder (2021) identified a more complex dynamic where they show how traditional media helped give visibility to digital media during the 2019 Indian elections.

The evidence about the prominence of digital-born outlets in India therefore, reflected in their popularity or *audience reach*, is still far from conclusive. Questions remain regarding whether digital-born outlets have at all challenged the hegemony of traditional (legacy) media in the Indian online space. This allows me to formalize the first two hypotheses related to the dynamic between legacy and digital-born outlets.

H1a: Legacy outlets have significantly higher audience reach online compared to digital-born outlets in India.

H1b: The audience reach of digital-born outlets has significantly increased over time.

Regional Media, National Media and the Vernacularization of the Indian Internet

The digital media landscape in India was not born in a vacuum, but was shaped by powerful contextual forces that characterized the pre-digital media environment in the country (Athique, 2012). This resulted in a landscape that is, at least structurally, considerably different from media landscapes in other (usually western democratic) nation states. One of these stark differences is rooted in the linguistic diversity of India: India has 23 official languages. The linguistic diversity, which reflects a deeper cultural diversity in the country, is also mirrored in the political organization of the country around 29 states - the vast majority of which are ethno-linguistically defined. The existence of these linguistically segregated regional cultures for several centuries, gave rise to a variegated media environment, in both, print and broadcast, with different regional media industries (publishing in the local vernaculars) existing in parallel with the “national” media. Chakravartty & Roy (2013) argue that while, in Euro-American countries the media environment has grown to be spatially nested under an apex, all-encompassing national media, the cultural diversity in India has given rise to a more poly-centric and decentralized media structure. Even when it comes to ownership patterns, the west has seen more consolidation and integration at a national level, whereas in India, the integration has happened in different regions across the country. As Kumar (2014) argues, “the term regional media is traditionally used to describe an intranational or subnational category of media produced in the many regional languages spoken within India. [...] Since most states within India are linguistic states, regional media in India are usually defined in relation to the geographic boundaries of the states and the dominant languages within those states”. This historical cleavage between “national” media and “regional” media that has long characterized India’s offline media environment, has persisted in the online space as well and represents one of the most striking points of departure from the western model. In fact, the poly-centric structure

of the Indian media system (Chakravartty & Roy, 2013) is further reflected in the overwhelming popularity that vernacular regional media enjoy compared to their English-language national counterparts, in both, the print and broadcast spaces (BARC India, 2017; Neyazi, 2018; *Press in India 2019-20*, 2020) as well as the existence of linguistically distinct “news reading publics” (Mukerjee, 2022) online. This demand for regional media has led to many vernacular outlets transitioning to online news, while many digital-first news outlets have also begun catering to regional online audiences in the local vernacular. This leads me to intuit that while national media would still enjoy significantly greater online audience reach than regional media, regional outlets would have become increasingly more popular over time. This would be driven in part by more and more Indians from the rural areas getting Internet access and transitioning from offline regional outlets to their online counterparts.

In parallel to national and regional media, international (i.e. foreign) media have also emerged as prominent sources of news consumption among Indian news consumers. Respondents to a Reuters survey in India mentioned BBC News and CNN among their most used news sources, both offline as well as online (Newman, Fletcher, Robertson, Eddy, & Nielsen, 2022). Thus, while I do not expect international media to be more popular online than national or regional media, I would still expect them to have become significantly more popular over time.

I thus propose the next set of hypotheses:

H2a: National outlets have significantly higher audience reach online compared to regional and international outlets in India.

H2b: The online audience reach of regional outlets has increased significantly over time.

H2c: The online audience reach of international outlets has increased significantly over time.

A related dimension for classifying media outlets in India is the language that they publish in. Since regional media largely publish in the local vernacular (barring a

few exceptions that publish in English), and national media and international media in India publish in English, I expect the regional/national/international cleavage to correlate closely with the vernacular/English-language cleavage. This phenomenon described by Neyazi (2019) as “internet vernacularization” has been thought to parallel the arc of evolution of offline media in the country. Neyazi opines that just as with print media and television media in the past which were originally dominated by English-language media but then gradually lost out to vernacular media, online media is likely headed in the same direction. This allows me to pose the following two hypotheses that speak to the trend of “internet vernacularization” in the country.

H3a: English-language outlets have significantly higher audience reach online compared to vernacular outlets in India.

H3b: The online audience reach of vernacular outlets has increased significantly over time.

The above discussion demonstrates that in addition to the legacy vis-à-vis digital-born media divide that most countries around the world have witnessed in one form or the other, the regional vis-à-vis national media divide that is unique to India provides yet another axis of analysis to understand the evolution of India’s media landscape. The manner in which these two axes intersect has profound implications for how a culturally diverse audience experiences media as well as the future of the media industry in the country. As prior work has shown, the power-imbalance between national media and vernacular media in India is vastly different online compared to offline: the legacy national media constitute the main gravity centers for online attention, unlike the vernacular regional media that dominates the offline space (Mukerjee, 2022). The final set of analyses is therefore exploratory in nature and will examine how the audiences transitioning from national/regional/international media (based on the results of testing H2b and H2c) navigate the legacy/digital-born divide in the Indian online space.

In all the longitudinal analyses, I use *audience reach* as the key variable because,

while not perfect, it is a good indicator of the popular demand for the media outlet. Since the online Indian media landscape is highly privatized and all media outlets effectively compete for market share to survive in the business, they rely largely on advertising revenue and to a lesser extent, subscriptions. Since both advertising revenue and subscriptions are directly correlated with the traffic websites receive, audience reach and its change over time can help understand the health of the outlet as reflected in its public demand in the market.

Data

The dataset used in this study was obtained from ComScore, a media measurement company that maintains representative panels of online audiences in over forty countries around the world. This dataset pertains to their panel of Indian *desktop*² users ($N \approx 50,000$), and is drawn monthly, over a period of 45 months (October 2014 - June 2018). The data are gathered by means of a proprietary technique that integrates the desktop browsing behavior of the panelists (captured using a tracking software) with server-side data (captured using tags in the website pages), followed by de-duplication, and controlling for potential bot activity. The dataset has two statistics that are central to the subsequent analysis: *audience reach* and *cross-visits*³. Audience reach captures the number of unique visitors to an outlet; cross-visits captures the number of unique visitors *shared* by two outlets. Together these statistics allow us to appraise exposure and co-exposure levels to online news. Both statistics are available monthly for the entire 45-month period. The websites that appear every month, however, vary, as ComScore sets a minimum threshold (0.1%) for online traffic for a website to be included in their dataset each month. For the sake of longitudinal comparability, and to ensure that all subsequent statistical estimates are unbiased, only the websites that occur in all the 45 months are retained. This yields a set of 174

² The lack of mobile audience data is elaborated upon in the Discussion section

³ a third statistic, average time spent per visitor, was also available. Analysis of this is reported in the Appendix

outlets, each of which was classified along three dimensions: legacy (the media outlet predated the internet) or digital-born; regional (with audiences localized to a few states of India) or national or international; and whether it publishes in the English language, a vernacular language or in both (See Figure 2).

Figure 2. Bar-charts showing the relative frequency of the 174 media outlets by type.

Findings

Overall Prominence and Centrality of Outlets by Type

One measure of the prominence of the media outlets during the analysis period is their average monthly audience reach as a percentage of the ComScore panel size. A second measure of prominence is the centrality of the media outlet in the consumption network. In order to capture this, an audience overlap network is constructed with nodes as the outlets, and the edges capturing the average number of shared visitors (per month) between every pair two outlets. Following prior research (Mukerjee et al., 2018), statistically non-significant edges were removed. The more number of outlets, each outlet shares audience with, the more central it is in the consumption network. This intuition is captured by degree centrality. Figure 3 shows a scatterplot of the percent reach and the degree centrality of the 174 media outlets that occur in every month in the dataset. The color coding reveals that while the most prominent and central outlets are legacy media, and in the English language, their overall distribution is not

significantly different from that of their digital-born (H1a), and vernacular (H3a) counterparts respectively. The only significant difference is between National outlets and regional outlets (H2a) with national outlets having significantly greater reach than the regional outlets (Mann-Whitney's $U = 1931$, $p = 0.00679$).

Thus, of the cross-sectional hypotheses, only H2a is supported, but not H1a and H3a. The results of all non-parametric tests are reported in Table A1 in the Appendix. Degree centrality and the percent reach values of the outlets are also significantly correlated with a very high rank correlation (Kendall's $\tau = 0.719$ ($p < 2.2 \times 10^{-16}$)).

These cross-sectional statistics from the data aggregated over the 45 month period already point to a landscape that is dominated by national media outlets, both in terms of their number in the dataset as well as their relative prominence. While it is entirely possible that a large number of vernacular or international outlets are not included in the analysis owing to their not occurring every month, it is highly improbable that they were extremely prominent in some of the months (so as to change the qualitative findings), and not prominent enough to register the minimum threshold that ComScore requires, for other months. As such, this study does not account for the long tail of outlets that gather very little monthly audiences. The subsequent analyses that look at the longitudinal trends, therefore, focus only on these 174 outlets.

Longitudinal Trends of Media Audiences by Type

To investigate the longitudinal trends for different types of media outlets (and test H1b, H2b, H2c, and H3b), the *monthly median audience share* (i.e. the audience reach of the outlet as a percentage of ComScore's online population) was estimated for every type of outlet, and a linear model was fit with respect to time. The results are depicted in Figure 4, while regression tables are provided in the Appendix. Panel A depicts this comparison between legacy outlets and digital-born outlets: the median audience shares of legacy outlets have increased over time and that of digital-born outlets decreased. However, neither slope is statistically significant. In other words, there is no evidence to suggest that digital-born outlets have managed to contest the hegemony of legacy media

Figure 3. Scatterplots showing the high correlation between percentage reach and degree centrality, by type.

in the Indian online media landscape by drawing audiences away from the latter. H1b therefore is not supported (see Table A3 in the Appendix for slope estimates).

Panel B compares these trends between regional (Indian), national (Indian), and international (i.e. foreign media outlets). We see that the median audience size of regional outlets has declined significantly (contradicting H2b completely), while the median audience size of international outlets has significantly increased (H2c is supported). Table A4 in the Appendix shows the OLS slope estimates. The trend for national outlets is also shown for comparison. These outlets have sustained no significant increase or decrease in median audience size over time. While I do not test for the cause driving these trends, it could be owing to a potential turn away from localism to globalism in online Indians' news seeking behavior.

Panel C looks at the trends in audience reach based on language of the outlet (English/vernacular). Vernacular outlets suffered a *significant decline* in median audience reach over time (thereby refuting H3b), while for comparison, English language outlets witnessed a significant rise (see Table A5 in the Appendix for slope estimates). If we remove foreign outlets from this group of English language outlets, and concentrate only on Indian English-language outlets, the estimate of the slope of the upward trend as well as the level of significance decreases (see Table A6 in the Appendix). In other words, the increasing median audience reach of English language outlets turns out to be largely driven by the increasing popularity of foreign media outlets in the country.

These analyses provide two main insights. First, despite the growth in the number of digital-born outlets in India, temporal changes in their median audience reach suggests no evidence that they have managed to divert audience attention away from legacy media outlets. The second, and more concerning finding is that contrary to speculation regarding "internet vernacularization" (Neyazi, 2019) the audiences of regional media in vernacular languages have *diminished* in size, while that of English language national media and international media have grown. These findings help provide some background about the evolution of the online Indian news consumption landscape in general, allowing us to next query the nature of these shifts.

Figure 4. Longitudinal trends of the median reach of outlets by type. Blue lines indicate that the slopes are significantly greater than 0, while red indicates the slopes being significantly lower than 0. Black lines have slopes that are not significantly different from 0. All significance levels were tested at the $p < 0.05$ level.

Audience Mobility

To unpack the nature of the migration of audiences from vernacular, regional media to English language national and international media, I propose the concept of audience mobility. *Audience mobility* is defined as the progressive migration over time of an audience (or segments of an audience) from one type of media to another. While this could encompass the migration across media formats in a broader sense (for e.g. from print to online), it could also, crucially in the context of online news, imply the migration from certain types of media outlets to other types of media outlets within the same media format.

While there is some audience mobility that always happens owing to factors not directly related to changes in the media environment (for example, demographic changes or increasing literacy in the country), audience mobility can be amplified following certain infrastructural changes in the media environment. An example of the former is the increasing number of vernacular newspaper readers in rural India since the turn of the 21st century owing to a steadily increasing literacy rate: in this case, audiences can be considered to have “migrated” from media that did not require them to be literate (for e.g. the television and the radio) to one that did (viz. newspapers). An example of the latter is the Indian government’s Digital India campaign, an umbrella term encompassing several concerted efforts to create digital infrastructure to provide digital services to citizens (Neyazi, 2019, pg. 101). While both these forms of audience mobility can be distinguished theoretically, they can often act in tandem, and reinforce each other.

I operationalize audience mobility between two types of media outlets by using a variable called *Percent Overlap*. Percent overlap between two outlets is calculated as the ratio of their cross-visits to the reach of one of the outlets. It thus captures the fraction of the audience that visits one outlet, that also visits the other outlet: a percent overlap of 0.2 between outlets A and B implies that 20% of the audience of outlet A visits outlet B. For calculating audience mobility between two types of media outlets (say, T1 and T2) all pair-wise percent overlap values between outlets of type T1

Figure 5. Trend of increasing literacy in India (1911-2011). Source: Census data

and type T2 are calculated. The *median percent overlap* between types T1 and T2 for any month is the median of all these values. This gives us the median of the audience fractions that visit outlets of type T1 that also visit outlets of type T2 in that month. Analyzing the temporal trend of the median percent overlap can shed light on audience mobility between these two types of outlets. A significant increase in the median percent overlap implies that significantly larger fractions of audiences to outlets of type T1 are also visiting outlets of type T2 and vice versa. If the *median audience reach* of outlets of type T1 is decreasing at the same time, the significant increase in the median percent overlap between T1 and T2 indicates that more people are migrating away from outlets of type T1 to those of type T2.

The results so far indicate that the audience shares of regional outlets have significantly reduced over time. Therefore, the mobility of audiences *from regional outlets to national and international outlets* respectively are analyzed. Specifically, I look at audience mobility through the lens of the legacy/digital-born media divide, and

find qualitative differences in migration patterns.

Panel A in Figure 6 shows the audience mobility from the vernacular outlets to national media outlets over the 45-month period. Audience mobility in this case is measured by the temporal changes in the median percent overlap (between every vernacular outlet and every national outlet in the dataset), as a percentage of the absolute audience reach of the former. These trends are further broken down by the type (legacy/digital-born) of the national outlets. Visual examination of the regression lines suggest that the slope for audience mobility to legacy national outlets is greater than for audience mobility to digital-born national outlets. This is further corroborated by the highly significant regression coefficient for the interaction variable (See Table A7 in the Appendix).

Next, we turn to audience mobility from regional outlets to international outlets. Panel B in Figure 6 shows the trends in median percent overlap between vernacular outlets and international outlets, broken down by the type of the international outlets. This analysis yields a surprising finding: unlike the move to national media, vernacular audiences moving to international media are significantly more attracted to digital-born international outlets than to legacy international outlets (See Table A7 in the Appendix). This could indicate that people migrating to international outlets may either be driven by different motivations, or that structural features of the new media environment are responsible for these qualitatively different patterns.

Discussion

This paper sought to characterize the online news consumption landscape of India – a country, that, despite being the largest democracy in the world, has remained quantitatively under-investigated. The case of India “is not just intrinsically important, but also of broader relevance as an example of a market in the Global South characterised by rapid growth in internet access and a significant number of new digital-born news media competing with legacy media for audience attention and engagement” (Majó-Vázquez et al., 2019). To that end, the contributions of this paper

Figure 6. Audience mobility from vernacular to national (Panel A) and international media outlets (Panel B)

are two fold. First, the findings are of rich empirical value. The analyses reveals novel insights into the online news consumption behavior of Indian internet users, that can serve as valuable evidence for informing theoretical frameworks of news use and media exposure that are more global. Second, the findings speak to several theoretical debates about the power-imbalance between legacy media and digital-born media in the global South. They also make us rethink several conceptual assumptions about the poly-centricism of the media landscape and the trend of “internet vernacularization” in the country and whether they continue to be valid today. Empirically and theoretically therefore, this paper responds to calls to “dewesternize” the field of communication and media studies (Gunaratne, 2010; Iwabuchi, 2014; Waisbord & Mellado, 2014) and bring it more in line with a “multipolar world” (Thussu, 2018).

I find that while legacy outlets outnumber digital-born media in the ComScore

dataset, they do not necessarily register a significantly higher median audience reach. However, there is also no evidence to suggest that digital-born media have succeeded in contesting the hegemony of legacy media as the median audience reach of digital-born outlets hasn't grown significantly over the four-year period of analysis. This undermines many optimistic claims about the rise of digital-born media outlets in India (Brookings India, 2015; Chaudhury, 2016; Wang, 2016), echoing similar findings from existing western-centric studies (e.g. Mukerjee et al., 2018), that online platforms do not necessarily “flatten and democratize hierarchies of information control and power” (Entman & Usher, 2018).

The next set of findings are even more illuminating: I find that not only do regional media receive significantly fewer unique visitors than national media (an artefact likely of the smaller regional markets they cater to), their median audience has (contrary to expectation) also significantly *reduced* over time. In comparison, the median audience size of international outlets has witnessed a significant increase over the four year period of analysis. These findings are also echoed in the trend lines comparing English language media outlets with vernacular outlets. While this study cannot comment on the cause behind these trends, it can propose a plausible explanation, rooted in the the demographic and socio-economic changes that have taken place in contemporary Indian society. Two socio-economic and demographic factors that determine an individual's use of online news are internet penetration, and literacy in English or vernacular languages. Two other factors are personal interest in different types of news as well as the reputation (or awareness among audiences) of different news outlets. We would expect English outlets to become more popular than vernacular outlets if, all else remaining constant, the adult literacy rate in English was increasing at a faster rate than the default adult literacy rate (in vernacular languages) of the country. This is unlikely as the increase in the country's literacy rate is almost entirely driven by an increase in literacy rate in the rural areas of the country. Since primary education in these areas is mostly in the vernacular language of the region, the notion that English language outlets are becoming popular owing to people getting

English-literate faster is probably incorrect. While the differences in adult literacy across different languages are not captured by Census data, the stark disparities in literacy in English and vernacular languages at different grades of primary school are reflected in the Annual Status of Education Report surveys that are administered every year in villages across the country⁴. This then, lends more credibility to the idea that audiences who are increasingly moving to English language outlets are not doing so because they are suddenly more literate in English, but because they are potentially less interested in regional, vernacular news, at least on the internet. If true, this would be an interesting development and would mark a point of distinction between the online and the offline (where vernacular media continue to be extremely popular) in India (see Figures A3 and A4 in the Appendix). From a theoretical perspective, this finding also runs counter to Chakravartty and Roy's (2013) assertion that the media environment in India is "polycentric" comprising regional and national media of "equivalent weight" (pg. 352). Of course, Chakravartty and Roy's focus was largely on the television media of a limited number of states in India, and was also conducted nearly a decade ago. However, in the absence of any comparable study with online media in India and the similar sociopolitical milieu that both television media and online media occupy today, their theorizing does provide a useful starting point to describe the online media landscape in the country. I find that at least in so far as audience shares are concerned, the "weight" is contingent on the format. Certainly on the internet, the national media have far greater "weight" than do the regional media. The increasing inequality in audience shares points to an increasingly centralized media landscape that brings it structurally more in line with the "nested" Euro-American model with the "national media system as apex" (Ibid.). The significant decline in vernacular media audiences also runs counter to Neyazi's idea of "internet vernacularization" (2019). This echoes similar declines in local news readers and the resulting "crisis in journalism" in

⁴ see <http://img.asercentre.org/docs/ASER2018/ReleaseMaterial/aserreport2018.pdf>, findings summarized in <https://www.livemint.com/Opinion/1km3nIazOPD0JgDwqKjz8J/The-state-of-education-in-rural-India.html>

developed nations such as in the United States (Hayes & Lawless, 2018), which some scholars have attributed to the widespread availability of online news (McChesney & Nichols, 2010). This in turn would reinforce claims made by past scholars regarding the general negative effects that the internet has had on the profitability of local news.

A related finding is the increasing prominence of the international media in the online news consumption landscape of the country, as reflected in their significant increase in median audience reach over time. While this is not easily explained by demographic factors, it could mean that online Indian news seekers are increasingly interested in global affairs, and that more people are actively seeking out international news. This could be the direct result of a number of important events that happened globally during the study period (such as the Brexit referendum and the 2016 US election). Alternatively, this could also be a symptom of the rise and growing power of “neoliberal newspeak” (Bourdieu & Wacquant, 2001; Chakravartty & Schiller, 2011) in global financial news coverage which has become embedded as a form of cultural imperialism in the global media landscape. Yet another alternative explanation behind the significant increase in audience size of international outlets could simply lie in greater incidental news exposure (Boczkowski, Mitchelstein, & Matassi, 2018), which could be the by-product of more people in India going online, and inadvertently encountering popular international news outlets.

This is where the findings from the audience mobility analysis can potentially help shed more light: I find that among those moving from regional news to national news, there is greater preference for legacy media than for digital-born media. However, for those moving from regional outlets to international outlets (such as BBC, CNN, The New York Times) this preference is reversed. One plausible reason behind this finding could be the different levels of reputation of outlets — which in turn is a function of the level of awareness of the outlets among online Indian news consumers. In other words, when news seekers are aware of the reputation of media outlets (like they possibly are, with national outlets), one can rationalize their preference for the better known legacy outlets of the country. However, when they are not aware of the reputation of outlets

(like they possibly are not, with international outlets), other factors, such as incidental exposure to viral content, could play a bigger role. As the latter could characterize exposure to many digital-born international outlets such as BuzzFeed, online Indian news consumers transitioning from vernacular to international media, may simply be more exposed to BuzzFeed owing to the viral content they produce (potentially exacerbated by social media use), than visit the New York Times for its repute. While purely speculative, this example can provide a plausible explanation behind the observed trends that future researchers can explore further and test more rigorously.

Limitations are intrinsic to any research endeavor and this study is no exception. The primary limitation of this work relates to the nature of the dataset I use: they are only available in monthly aggregated formats. It is therefore impossible to analyze the data at a more fine-grained or individual level and distinguish between the choices people make and the structural aspects of news exposure. This also makes it impossible to distinguish between users based on their visit frequency or duration, as well as raises concerns about transparency. After all, ComScore uses a proprietary algorithm to generate its monthly aggregates. This limitation notwithstanding, the analysis of trace data for understanding audience behavior, has several documented advantages when compared to self-reports of news and information exposure (González-Bailón & Xenos, 2020; Prior, 2009). While the former is not free of its own systemic biases, the latter suffer from other kinds of (survey) errors, including but not limited to incorrect recall. The use of trace data, to that end, offers a relatively more accurate picture of how people consume news and how those patterns change over time.

Next, the dataset I use only captures desktop browsing activity and ignores the penetration of mobile devices in rural India, where vernacular media are likely to be more popular than English media. The reason behind this omission is the unavailability of ComScore data for non-desktop devices in India before 2017. Even after 2017, when ComScore did introduce multi-platform (desktop and mobile) data, it was available only sporadically. Figure A5 in the Appendix compares the reach of vernacular media and English media on desktop with mobile for the eight months in 2017 when mobile data

were available. Interestingly, I find that the dynamics are similar to what I find with the desktop data: that is, while mobile audiences are certainly larger in size than desktop audiences, English outlets remain overwhelmingly more popular compared to vernacular media even on mobile. The reason behind this could lie in the different ways in which people in rural areas use mobile phones, as compared to the people in urban areas, as a 2018 survey of 6752 adults in rural Karnataka highlights. The study found that only 49% of the respondents owned a mobile device, and fewer than 14% owned a smartphone or a touch phone (Vaijayanti, 2018). Moreover, of the smartphone users, “the majority (85%) were unfamiliar with internet communication channels including email and Skype, while only 11% were familiar with Facebook, WhatsApp and YouTube, 4% with gaming, and less than 1% with online shopping”.

Finally, this study measures exposure as visits to news websites. I recognize that a lot of online news exposure happen outsides of news websites, for example, while scrolling through social media feeds or on mobile applications of different media outlets, that my dataset does not capture. However, while this does prevent us from getting the complete picture of online news consumption in India, it is important to view the ComScore data from the perspective of existing research endeavour: given the scant empirical evidence that is available about news consumption in India today, ComScore’s audience estimates remain the most comprehensive yet accessible source of reliable web-browsing data in the country. Future studies can seek to explore other data sources to capture these alternative modes of online news exposure. Research endeavours in these directions, specifically with more granular audience data would help paint a more holistic picture of the overall news consumption landscape.

References

- Aneez, Z., Neyazi, T. A., Kalogeropoulos, A., & Nielsen, R. K. (2019). *India Digital News Report* (Tech. Rep.). Oxford, UK: Reuters Institute for the Study of Journalism.
- Athique, A. (2012). *Indian Media Global Approaches*. Polity Press.
- BARC India. (2017). *TV audience measurement data collection - BARC India*. Retrieved from <http://www.barcindia.co.in/statistic.aspx>
- Boczkowski, P. J., Mitchelstein, E., & Matassi, M. (2018, October). "News comes across when I'm in a moment of leisure": Understanding the practices of incidental news consumption on social media. *New Media & Society*, 20(10), 3523–3539.
- Bourdieu, P., & Wacquant, L. (2001). Neoliberal newspeak: notes on the new planetary vulgate. *Radical Philosophy*, 105(Jan).
- Brookings India. (2015). *Roundtable Discussion / The Rise of Digital Media and its implications on Journalism in India* (Tech. Rep.). Brookings Institute. Retrieved from <https://www.brookings.edu/events/roundtable-discussion-the-rise-of-digital-media-and-its-implications-on-journalism-in-india/>
- Chakravartty, P., & Roy, S. (2013). Media Pluralism Redux: Towards New Frameworks of Comparative Media Studies "Beyond the West". *Political Communication*, 30(3), 349–370.
- Chakravartty, P., & Schiller, D. (2011). Global financial crisis| neoliberal newspeak and digital capitalism in crisis. *International Journal of Communication*, 4, 23.
- Chan, J. K.-C., & Leung, L. (2005, June). Lifestyles, reliance on traditional news media and online news adoption. *New Media & Society*, 7(3), 357–382.
- Chaudhury, L. (2016). *Can the digital revolution save Indian journalism?* Retrieved from https://www.cjr.org/special_report/india_digital_revolution_startups_scoopwhoop_wire_times.php
- Chyi, H. I., & Lasorsa, D. L. (2002, April). An Explorative Study on the Market Relation Between Online and Print Newspapers. *Journal of Media Economics*, 15(2), 91–106.

- Croteau, D. (2006). The Growth of Self-Produced Media Content and the Challenge to Media Studies. *Critical Studies in Media Communication*, 23(4), 340–344.
- Delli Carpini, M. X., & Keeter, S. (1996). *What Americans know about politics and why it matters*. Yale University Press.
- Deuze, M. (2003). The web and its journalism: Considering the consequences of different types of newsmedia online. *New Media and Society*.
- Dubois, E., & Blank, G. (2018). The echo chamber is overstated: the moderating effect of political interest and diverse media. *Information, communication & society*, 21(5), 729–745.
- Entman, R. M., & Usher, N. (2018, April). Framing in a Fractured Democracy: Impacts of Digital Technology on Ideology, Power and Cascading Network Activation. *Journal of Communication*, 68(2), 298–308.
- Gentzkow, M. (2007). Valuing New Goods in a Model with Complementarity: Online Newspapers. *The American Economic Review*, 97(3), 713–744.
- González-Bailón, S., & Xenos, M. (2020, February). Surveys Underestimate Online News Exposure: A Comparison of Self-Reported and Observational Data in Nine Countries. *SSRN Electronic Journal*. doi: 10.2139/ssrn.3522774
- Guess, A., Nyhan, B., Lyons, B., & Reifler, J. (2018). Avoiding the echo chamber about echo chambers. *Knight Foundation*, 2, 1–25.
- Gunaratne, S. A. (2010). De-Westernizing communication/social science research: opportunities and limitations. *Media, Culture & Society*, 32(3), 473–500.
- Hallin, D. C., & Mancini, P. (2004). *Comparing media systems: Three models of media and politics*. Cambridge university press.
- Hayes, D., & Lawless, J. L. (2018). The decline of local news and its effects: New evidence from longitudinal data. *Journal of Politics*, 80(1), 332–336.
- Hindman, M. S. (2009). *The myth of digital democracy*. Princeton University Press.
- Iwabuchi, K. (2014). De-westernisation, inter-Asian referencing and beyond. *European Journal of Cultural Studies*, 17(1), 44–57.
- Kapur, D. (2002). The causes and consequences of india's it boom. *India review*, 1(2),

91–110.

- Kayany, J. M., & Yelsma, P. (2000, June). Displacement Effects of Online Media in the Socio-Technical Contexts of Households. *Journal of Broadcasting & Electronic Media*, 44(2), 215–229. doi: 10.1207/s15506878jobem4402_4
- Kaye, B. K., & Johnson, T. J. (2003). From here to obscurity?: Media substitution theory and traditional media in an on-line world. *Journal of the American Society for Information Science and Technology*, 54(3), 260–273.
- Kumar, S. (2014). Media Industries in India: An Emerging Regional Framework. *Media Industries*, 2.
- Lin, C., Salwen, M. B., Garrison, B., & Driscoll, P. D. (2005). Online News as a Functional Substitute for Offline News. In M. B. Salwen, B. Garrison, & P. D. Driscoll (Eds.), *Online News and the Public* (pp. 237–256). Mahwah, New Jersey: Lawrence Erlbaum Associates, London.
- Majó-Vázquez, S., Mukerjee, S., Neyazi, T. A., & Nielsen, R. K. (2019). *Online Audience Engagement with Legacy and Digital-Born News Media in the 2019 Indian Elections* (Tech. Rep. No. May). Oxford, UK: Reuters Institute for the Study of Journalism.
- McChesney, R. W., & Nichols, J. (2010). *The death and life of American journalism : the media revolution that will begin the world again*. Nation Books.
- Meraz, S. (2009). Is There an Elite Hold? Traditional Media to Social Media Agenda Setting Influence in Blog Networks. *Journal of Computer-Mediated Communication*, 14(3), 682–707.
- Mitchelstein, E., & Boczkowski, P. J. (2010). Online news consumption research: An assessment of past work and an agenda for the future. *New Media & Society*, 12(7), 1085–1102.
- Mukerjee, S. (2022). Rethinking audience fragmentation using a theory of news reading publics: Online india as a case study. *International Journal of Press/Politics*, 1–22.
- Mukerjee, S., Majó-Vázquez, S., & González-Bailón, S. (2018, February). Networks of

- Audience Overlap in the Consumption of Digital News. *Journal of Communication*, 68(1), 26–50.
- Newman, N., Fletcher, R., Robertson, C., Eddy, K., & Nielsen, R. (2022). Reuters institute digital news report 2022.
- Neyazi, T. A. (2018). *Political communication and mobilisation: The hindi media in india*. Cambridge University Press.
- Neyazi, T. A. (2019). Internet Vernacularization, Mobilization, and Journalism. In S. Rao (Ed.), *Indian Journalism in a New Era* (pp. 95–114). Oxford University Press.
- Neyazi, T. A., Kumar, A., & Dutta, M. J. (2019). Channel complementarity or displacement? theory and evidence from a non-western election context. *Journal of Broadcasting & Electronic Media*, 63(4), 656–676.
- Neyazi, T. A., & Schroeder, R. (2021). Was the 2019 indian election won by digital media? *The Communication Review*, 24(2), 87–106.
- Nguyen, A., & Western, M. (2007). Socio-structural correlates of online news and information adoption/use. *Journal of Sociology*, 43(2), 167–185.
- Press in India 2019-20* (Tech. Rep.). (2020). Registrar of Newspapers for India.
- Prior, M. (2009). The Immensely Inflated News Audience: Assessing Bias in Self-Reported News Exposure. *Public Opinion Quarterly*, 73(1), 130–143.
- Spohr, D. (2017). Fake news and ideological polarization: Filter bubbles and selective exposure on social media. *Business information review*, 34(3), 150–160.
- Sunstein, C. R. (2009). *Republic.com 2.0*. Princeton University Press.
- Sunstein, C. R. (2017). *#Republic : divided democracy in the age of social media*. Princeton University Press.
- Thussu, D. K. (Ed.). (2009). *Internationalizing media studies*. Routledge.
- Thussu, D. K. (2018). A new global communication order for a multipolar world. *Communication Research and Practice*, 4(1), 52–66.
- Vaijayanti, K. (2018). *Penetration of smart phones in rural India and implications for education*. Akshara Foundation. Retrieved from

<https://akshara.org.in/wp-content/uploads/article-V.pdf>

Waisbord, S., & Mellado, C. (2014). De-westernizing Communication Studies: A Reassessment. *Communication Theory*, 24(4), 361–372.

Wang, S. (2016). *In India, a boom of digital media startups find large audiences, but face familiar business challenges » Nieman Journalism Lab*. Retrieved from <https://www.niemanlab.org/2016/05/in-india-a-boom-of-digital-media-startups-find-large-audiences-but-face-familiar-business-challenges/>

Yang, T., Majó-Vázquez, S., Nielsen, R. K., & González-Bailón, S. (2020, November). Exposure to news grows less fragmented with an increase in mobile access. *Proceedings of the National Academy of Sciences*, 117(46), 28678–28683.