

THE UNIVERSITY OF CHICAGO

DECODING INDIAN MATRIMONIAL ADVERTISEMENTS: INDIVIDUAL  
PREFERENCES ENTRENCHED IN SOCIO CULTURAL BIASES

BY  
PRANATHI IYER

05.01.2023

A PAPER SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS  
FOR THE MASTER OF ARTS DEGREE IN THE MASTER OF ARTS IN  
COMPUTATIONAL SOCIAL SCIENCE

FACULTY ADVISOR: JON CLINDANIEL

PRECEPTOR: JEAN CLIPPERTON



## ABSTRACT

Matchmaking is an age-old practice in the Indian context. What started out as a tradition of seeking partners in close-knit communities, has today expanded to newspapers and online websites. These advertisements however continue to be extremely explicit, content-heavy, and rich sources of cultural and socio-economic individual preferences. In this study, I attempt to use this text, along with computational methods, to understand the most important preferences of individuals seeking partners through newspaper advertisements, and how their preferences vary with time. Moreover, I attempt to qualitatively understand how people employ print versus digital media to facilitate the matchmaking process, and how this impacts the relevance of newspaper matrimonial advertisements. I use robust methods of natural language processing and computational linguistics such as part-of-speech tagging and word embeddings to conduct a longitudinal study across years. I argue that individual preferences of brides and grooms are rooted in gender and cultural biases, and remain constant with time. Using newspaper matrimonial advertisements of a north Indian newspaper, The Tribune, from 2001-2014, I find that there exists a bias for “tall”, “slim”, “fair”, and “beautiful” women, and “handsome”, “tall”, and “intelligent” working men, from the Sikh or Jat community. This underscores the fact that these advertisements are embedded in stereotypical gendered roles and a strong emphasis on intra-faith alliances. The results of this research also highlight the increasing preference of individuals towards using online platforms for matrimonial advertisements. In doing so, this study discovers nuanced biases in the social fabric of Indian matchmaking traditions and raises questions about the relevance of newspaper matrimonial advertisements and their implications on the matchmaking ecosystem.

## Introduction

Arranged marriage is a tradition that originated in ancient India, and continues into the present day (Jaiswal, 2014). It emerged in the historical Vedic religion and consisted of families spreading the word about prospective brides and grooms amongst communities staying in close proximities or social circles, hoping to find a suitable match. The institution has been fairly robust in adapting itself to changing social circumstances (Ramchandani, 2000). Thus, as times changed and these communities mobilized and moved farther from their circles, this communication amongst them migrated to Indian newspapers in the form of matrimonial advertisements (Anand, 1965). These advertisements were then used by families, friends, and matchmakers to find potential partners for individuals, with the intention of marriage. Today, these advertisements also find their place in digital media as several major Indian websites have created platforms to facilitate the matchmaking process for prospective brides and grooms. This shift to newspapers and websites has been accompanied by continued preference towards arranged marriages, by an overwhelming majority of the Indian subcontinent. In fact, approximately 90% of marriages in India today are reported to be arranged marriages<sup>1</sup>. Therefore, not only have these matrimonial advertisements found newer channels of dissemination through newspapers and websites, but they also continue to remain influential facilitators of a cardinal Indian practice. In times when technology has made communication and the search for partners and relationships more accessible, why do people continue to seek partners through traditional matchmaking? This persistence of the matchmaking system raises questions about the foundations of arranged marriages and how matrimonial advertisements are able to aid this social process. Thus, this paper hopes to probe two facets associated with Indian matchmaking. First, the sociological aspects of marriages and relationships formed via the matchmaking system. Second, the role of media – digital versus analog – in facili-

---

1. <https://harvardpolitics.com/harmful-representation-arranged-marriage-in-netflixs-indian-matchmaking/>

tating these relationships. In order to explore these perspectives, it is first important to understand what these matrimonial advertisements and digital platforms look like.

### *Matrimonial advertisements*

Most major Indian newspapers have a dedicated section for matrimonial advertisements. These advertisements are short and explicit texts that intend to convey individual preferences for prospective partners within limited words (Figure 1).

**Looking for a professionally qualified match for beautiful, well settled 1992, 5'-2", Canadian PR, Ravidasia girl. Done Masters from York University, Toronto. Presently working as Assistant Manager in Bank (Toronto). Family well settled in Mohali. Send**

Figure 1: Matrimonial advertisement in an Indian newspaper, The Tribune, 2011

These advertisements have distinct sentence structures and contain mentions of instrumental cultural and socioeconomic preferences such as caste, physical appearance, religion, income, education, and occupation (Ramakrishnan, 2012). This rich cultural information renders matrimonial advertisements as interesting lexical sources to understand preferences, perceptions, and ideologies of people seeking lifelong partners.

With the increasing pace of technology, today such advertisements appear on several matrimonial websites with almost similar structures and forms. However, depending on website design, digital advertisements might be longer than newspaper advertisements. Prospective brides and grooms also have the option of adding photos and personal contact details to their profiles on such platforms. Given that advertisements on these digital platforms are subject to creating a profile or registering, I do not present an example here. However, some of the most popular websites where this can be found are Bharat-

matrimony.com <sup>2</sup>, Jeevansathi.com <sup>3</sup>, and Shaadi.com <sup>4</sup>. These websites hold no specific significance for this research apart from giving readers a sense of digital matrimonial advertisements.

It is important to note that despite the emergence of matrimonial websites, newspaper advertisements continue to be heavily used in India. In fact, newspaper advertising in India is projected to reach 3.02 billion dollars by 2027 <sup>5</sup>. I further emphasize the eminence of widespread relevance of analog advertisements for this study later on. I use analog and digitized, and online and born-digital interchangeably respectively throughout the rest of this paper where digitized refers to digitized analog advertisements.

### *Longitudinal nature of arranged marriages*

While talking about matrimony and matchmaking, it is essential to account for the frame of time in consideration. According to a survey conducted by the Family Planning Association of India in 2000, 60% of urban India between the ages of 15 and 29 preferred arranged over love marriages. As of 2018, 93% of respondents of another survey conducted by Lok Survey Foundation claimed to have an arranged marriage (S., 2018). This speaks to the continued relevance of arranged marriages across generations of the population with changing times. However, marriage as an institution itself is not static (Hunter, 2007). The foundations of, and patterns in marriages and relationships change over time even if there is a prevailing inclination towards arranged marriages. For instance, scholars have used longitudinal surveys of labor economics to show that the importance of earnings or income for marriage formation grew in the late 90s (Sweeney, 2002). Thus,

---

2. <https://www.bharatmatrimony.com/>

3. <https://www.jeevansathi.com/>

4. <https://www.shaadi.com/>

5. <https://www.statista.com/outlook/amo/advertising/print-advertising/newspaper-advertising/india>

it then becomes important to understand if and how people’s partner-seeking behaviour has changed over the years whilst they continue to use matrimonial advertisements. One way to do this would be to perform a longitudinal analysis of matrimonial advertisements over the years. A longitudinal study of these texts also allows us to understand cultural transformations in matchmaking preferences, which can be slow to transpire and too slow to be observed in studies within a short timeline. For example, it can be hard to capture shifts in individual preferences at a year-wise temporal resolution since there might not be visible change between two years. However, analyzing advertisements over several years allows us to capture macro-level cultural shifts in these preferences.

As mentioned above, accessing born-digital data can be hard and raises potential questions of privacy since most matrimonial websites restrict access to advertisements only to registered users. Therefore, I adhere to using digitized matrimonial advertisements for the purpose of longitudinal content analysis of advertisements. In the realm of digital media, decoding these advertisements also enables us to understand the relevance of newspaper advertisements despite the existence of online platforms which offer faster channels of communication.

While manual analysis poses a challenge to large-scale longitudinal analysis, methods of computation allow us to carry out such content analysis across a wide span of time. Therefore, the objective of this research is to employ computational methods to answer the question as to what the most important individual preferences of men and women seeking matrimony through newspaper advertisements are and how these preferences vary with time. It also aims to understand the relevance and success of newspaper and online advertisements in facilitating the matchmaking process.

To answer the first part of my question, I use data collected for the years 2001 to 2014 from the digital archives of The Tribune, an English newspaper widely distributed in

northern India. I then employ methods of computational linguistics such as part of speech tagging, and statistical measures such as Kullback-Leibler Divergence (KL divergence), to understand structural and semantic similarities and differences in both, bride and groom seeking advertisements across time. Finally, in order to understand any observed patterns from a lens of gendered differences and societal norms outside the Indian context, I train word embedding models on bride and groom seeking advertisements across all years. I use these models to construct a dimension of gender on which I project exogenous lists of adjectives and occupations to capture any gendered biases in the embeddings and validate findings observed from the previous parts. In order to understand the relevance of newspaper and online advertisements, I then use Amazon’s Mechanical Turk to undertake a survey amongst a small sample of people who have been married through the Indian matchmaking system between 2005 and 2022.

The results highlight that preferences of men and women seeking partners through newspaper matrimonial advertisements are broadly captured by three categories of religion, physical appearance and qualities, and occupation. They also emphasize intra-faith alliances across both genders. I also observe that while several occupational preferences appear in association with men, only a few appear in association with women. This shows that individual preferences are also rooted in cultural biases and stereotypes of gendered roles — wherein women are associated with homemaking, while men are encountered by expectations of occupational roles and financial independence. Moreover, the results show how these patterns are consistent across time. This is validated by a comparison of KL divergence between years from 2001 to 2014. Furthermore, the results of projections on embedding models are commensurate with those of part-of-speech tagging. They underscore the cultural bias and stereotypes present in these matrimonial advertisements. Lastly, additional qualitative analysis shows an increasing preference for online advertisements compared to newspaper advertisements, but with gendered and



geographical differences among those who successfully find partners through born-digital advertisements. Thus indicating a potential shift from print to digital media in the Indian matchmaking ecosystem for certain groups of the population, but a continued relevance of analog advertisements for others. This raises questions about the implications of analog advertisements for prospective brides and grooms, newspaper media houses, and online matchmaking platforms.

The consistency of cultural patterns across years raises sociological questions about factors influencing preferences of individuals and why they do not vary conspicuously with time. This study is the first of its kind to my knowledge that undertakes a large scale longitudinal and quantitative analysis of Indian matrimonial newspaper advertisements and attempts to understand their relevance compared to online matrimonial advertisements. In doing so, it makes several contributions. First, it attempts to add to sociological literature around the foundations of the phenomenon of marriage and relationships in the Indian context. Second, it hopes to foster conversations around the consistent nature of individual preferences against the backdrop of a diachronic institution such as marriage. Third, it initiates discussions around understanding and exposing cultural and gender biases associated with the matchmaking process. This in turn constitutes the first step towards rethinking platform designs to improve gender equality and mitigate potentially oppressing stereotypes. Fourth, this paper lays the stepping stone for future research that must delve deeper into sociological and cultural reasons that influence preferences of individuals in the arranged marriage system in the first place, and their possible implications on conceptions of gender roles and relationships in India with changing times. Fifth, it probes a more thorough investigation around the relevance of newspaper and online platforms in their ability to facilitate matchmaking through matrimonial advertisements. Lastly, it attempts to lay the foundations for broader sociological research that can investigate if and how preferential seeking patterns of individuals vary with time

across other widely prevalent channels such as dating platforms. The rest of this paper is organized in a manner that aims to inform readers about the background literature, data, methodology, and findings of this research in a comprehensive way.

## **Background literature**

I think of relevant works in this area as having three moving parts. First, sociological theory and research that has probed the foundations of marriage, relationships, and partner-seeking preferences. Second, studies that have looked into matchmaking and matrimonial advertisements specifically in the Indian context. Last, works that examine how different media could potentially contribute to different phenomena and norms.

### *Marriage and mate selection*

Sociologists have long studied mating preferences because of how marriages affect different sociological trends (Buss, 1985). As far back as 1950, Hollingshead talked about this idea of a narrow theory of cultural determinism where single males and females select mates based on controls imposed on them by their culture (Hollingshead, 1950). He found culture, specifically American culture to place restrictions on an individual's mate choice with respect to race, religion, age, ethnic origin, and education. Families in the Indian context seem to exert similar control in the mate selection process and institutionalization of arranged marriages in India (Gupta, 1976). This speaks to the specificity and explicit nature of matrimonial advertisements along these cultural dimensions, given that these advertisements are a channel of mate selection. Similarly, in other studies, scholars have found propinquity, age, race, religious faith, ethnic background, education, and socioeconomic status to be key determinants of how a majority of individuals decide who to mate with (Buss, 1985). Moreover, theories of socialization elaborate further to say that selective contacts along these variables should be reflective in high correlation with

a variety of personality, attitudinal, and social practice items since individuals from the same demographic are more or less exposed to similar patterns of socialization (Bolton, 1961). This could potentially align with the fact that most matrimonial advertisements often seek partners from the same community as the prospective bride or groom since that might imply exposure to similar social practice items. However, researchers have also found gendered differences in such preferences. Sociological research on marriage has also built on theoretical frameworks such as the Parental Investment Model which claims that women are more likely than men to seek a mate who possesses nonphysical characteristics that maximize the survival or reproductive prospects of their offspring. Based on an empirical study in the US, scholars verified this model and that women accorded more weight than men to socioeconomic status ambitiousness, character, and intelligence (Feingold, 1992).

Such research could offer sociological explanations of why individuals continue to participate in the Indian matchmaking system. Matrimonial advertisements are vehicles of mate selection preferences that have been found to be essential determinants of marriage in different cultures. However, it is apparent that most sociological research around mate selection, and foundations of individual preferences towards marriages thus far is ethnocentric in the West, more specifically the United States. This makes it hard to embed research questions and analysis of Indian matchmaking in existing theoretical frameworks which encompass the Indian subcontinent. Given that arranged marriage and matchmaking are tightly knit with Indian communities and culture, it would be meaningful to extend current sociological frameworks and research to account for broader contexts. Previously, Indian scholars have emphasized the dearth of ethnographic research on rural and urban families and marriage in India (Hill, 1958). This research hopes to engage in content analysis of matrimonial advertisements to address this lacking and contribute towards developing broader theories of marriage and matchmaking.

## *Indian matchmaking and matrimony*

While there is a paucity of sociological frameworks around marriage in the Indian context, there have been few attempts at studying the cultural aspects of Indian matrimonial advertisements. They trace as far back as 1969 when researchers observed that individuals who seek partners through this institution tend to conform to the prevalent value system of their status group in doing so (Stuers, 1969). Several studies prior to the 2000s have engaged in qualitative and small-scale quantitative analyses of these advertisements. They have identified cultural preferences for brides and grooms, specifically along caste (Titzmann, 2011, Upreti, 1967). More recent research has used large scale quantitative methods to analyze online matrimonial advertisements (Rajadesingan et al., 2019). While the authors of this study use a large corpus of online advertisements to examine openness to inter-caste marriages in India, they collect and scrape data by creating a profile on a major matchmaking online platform (Rajadesingan et al., 2019). This approach could raise potential questions of ethics and respect for an individual’s privacy. Moreover, it only analyses online advertisements and not newspaper advertisements. Thus, there exists no major study that looks at large scale analysis of newspaper advertisements - possibly because of the challenge of scraping data. It can be difficult to form broad theories or extend current ones based on small-scale qualitative research. Hence, large scale analysis of advertisements also allows for a meaningful extension of existing sociological research as discussed in the above section.

As previously mentioned, marriage is not a static institution. Therefore, in addition to large scale analyses, a longitudinal investigation of these advertisements also captures temporal shifts in preferences for mate selection for marriage. However, existing research in the Indian context has not yet looked at longitudinal approaches that analyze how preferences of individuals seeking matrimony have or have not changed with time. This study aims to bridge this gap and further the existing repository of research around Indian

matchmaking.

### *Analog vs born-digital matrimonial advertisements*

With the boom of the internet and the increase in digital spaces and born-digital data, researchers have explored the place these different media hold today and their influence on society. Previously, scholars have posited that while newspapers are experiencing declining circulation figures, they offer a niche for regional and local newspapers that cannot be substituted by online advertisements (Lindstädt and Budzinski, 2011). This is extremely relevant for a country like India which has been reported to speak over 121 languages<sup>6</sup>. On the contrary, there is a visible growing importance of internet advertisements for consumer purchasing in India, and the potential of online advertisements as an efficient medium cannot be overlooked (Dash and Belgaonkar, 2012). However, most literature in understanding this space for born-digital advertisements surrounds consumer behaviour. There have been only a few studies that have looked at born-digital data in the context of matchmaking and have found that the use of matrimonial websites has implications for family disintermediation, cultural convergence, continuous information flows, ease of disengagement, virtual dating and reduced stigma in arranged marriages in India (Seth, 2011). A more small-scale analysis of online advertisements conveys that younger women i.e. would-be brides posted more advertisements in online media, compared to older family members (most websites allow users to describe who they are posting for when creating a profile). The analysis also finds the continued prevalence of physical attributes such as fair skin in online advertisements (Chattopadhyay and Chattopadhyay, 2019). This however does not tell us if there is a potential shift in the preferences of users to use a certain medium, and any distinction between populations that use newspaper and online advertisements respectively.

---

6. <https://censusindia.gov.in/nada/index.php/catalog/42561>

The newspaper industry is still thriving in India with a population of more than 1 billion people and varying internet penetration across geographies (Auletta, 2013). Thus, analyzing the content of newspaper advertisements seems of utmost relevance. However, there is also an apparent growth of the born-digital ecosystem in India. Therefore, there seems to be this friction between the explosion of the internet and online platforms, and the persisting need for print and newspaper content especially in the Indian context. This could imply a difference in the populations and hence the strata of society that use analog and born-digital advertisements. While it is hard to understand and capture if there is a shift in individuals using different forms of media, it can be attempted through qualitative and survey approaches. Current literature has not attempted to qualitatively understand the relevance of these analog and born-digital matrimonial advertisements in the matchmaking process in today's times. This study hopes to fill this gap in research by undertaking a qualitative approach to understand people's demographics and preferences towards certain media of matrimonial advertisements, and potential interaction between the two.

To encapsulate, there seem to be three broad overarching gaps in current scholarship with respect to Indian matchmaking, matrimonial advertisements, and arranged marriages. First, a lack of sociological literature that establishes the foundations that explain arranged marriages. Second, limited research when it comes to longitudinal and large scale analysis of analog Indian matrimonial advertisements. Lastly, a dearth of qualitative understanding of differences in individual preferences of people towards various media (analog vs born digital) of matrimonial advertisements. This study hopes to work towards narrowing these gaps.

## Data

I use two sets of data to carry out my research. I first use data from the digital archives of one Indian newspaper, The Tribune. The Tribune is one of the most widely distributed papers in Northern India. It covers five states, namely, Punjab, Haryana, the Union Territory of Chandigarh, Jammu and Kashmir, and Himachal Pradesh. It is one of the few Indian newspapers which has systematically archived matrimonial advertisements in textual formats from the year 2001 to 2014. There are typically two categories of these advertisements — bride-seeking and groom-seeking. My sample has 179,649 and 351,101 bride-seeking and groom-seeking advertisements respectively. While this is not representative of the entire country, owing to constraints of feasibility, I adhere to scraping and collecting data across these years from a single newspaper. It is also important to note here that data collected can vary in sparseness and volume across years. However, I use text analysis methods on individual years and then compare my results to maintain year-wise temporal resolution, but also understand change across years. Thus avoiding any interdependencies on frequency and amount of data across years to some extent.

As mentioned earlier, matchmaking in today's times in India has shifted to several online websites with large consumer bases. I use Amazon's Mechanical Turk (MTurk) in order to actually capture the relevance and success of these different media — print and digital — in enabling people to find their prospective partners. My sample of MTurk workers consisted of 60 people who got married via Indian matrimonial advertisements between 2005 and 2022. Among those that participated 50% of individuals identified as male, 48% identified as female, and the rest preferred to not disclose their gender. While MTurk might not be the best way to reach out to a representative Indian diaspora, it is one of the few ways that allows me to easily reach out to individuals, without any prior bias about their potential behaviour or responses.

## Methods

### *Part of speech tagging*

Once having pre-processed, tokenized, and normalized the data for all years, I use part-of-speech tagging using SpaCy, to understand the change in the semantics of these advertisements across time (Voutilainen, 2003). Part-of-speech tagging is a method of computational linguistics, which identifies various parts of speech based on a piece of text (Schubert, 2020).

**QUALIFIED PROFESSIONAL BRAHMIN OR NON-BRAHMIN VERY FAIR BRIDE,**  
164 CM MINIMUM, FROM **WELL-CULTURED FAMILY** FOR **PUNJABI BOY** 27 YEARS, 178  
CMS, HIGH RANK **IT PROFESSIONAL** WITH **MNC (USA),**

Figure 2: Part of speech tagging of a matrimonial advertisement

For instance, given the advertisement in Figure 2, SpaCy identifies three primary parts of speech.

- Words highlighted in yellow — Brahmin, Non-Brahmin, Punjabi, and MNC(USA) — are categorized as proper nouns. While Brahmin and Non-Brahmin are preferences of caste in the Hindu religion, Punjabi and MNC(USA) emphasise geographical and broad religious preferences.
- Words highlighted in red — bride, family, boy, and IT professional — are classified as nouns. This showcases an occupational preference of an individual with respect to a bride or groom.
- Words highlighted in pink — fair and well-cultured — are categorized as adjectives. These are key qualities and apparent physical attributes.

This preliminary analysis shows that key preferences of physical appearance, caste, religion, occupation, and education — essentially content that varies between advertise-



ments — are primarily captured by proper nouns, adjectives, and nouns. Therefore, part of speech tagging can be an extremely potent method to capture preferential changes of brides and grooms in these advertisements. Thus, owing to the short sentence like and homogenous lexical nature of these advertisements, I specifically look at changes in adjectives, proper nouns and nouns in these texts across time using part-of-speech tagging.

In the above example, I do not talk about “qualified” and “professional” since these appear in many advertisements across categories and do not offer meaningful insights. Moreover, I also exclude attributes of height from the scope of analysis of this paper since it can get extremely challenging to extract attributes at such a granular level because they are often mentioned in different formats. However, future research can potentially look into analysing this information to capture patterns across height — an important dimension of physical appearance.

### *Kullback-Leiber Divergence*

After analyzing how proper nouns, adjectives, and nouns in these advertisements have changed over time, I then use Kullback-Leibler Divergence (KL divergence) to further validate and measure changes in the content of these advertisements across years. KL divergence is a statistical distance that is used to extrapolate how a distribution P is different from distribution Q (Kullback and Leibler, 1951). In other words, it can also be interpreted as the amount of information that would be lost if one used distribution Q to extrapolate information about distribution P. P and Q in this project refer to distributions of advertisements of two different years, within a given category.

### *Word Embeddings*

Next, I use word embedding models to validate my findings by using part-of-speech tagging at a macro level. This approach aims to understand the cultural trends observed via using part-of-speech tagging, from a perspective of adjectives and occupations associated with

different genders outside the Indian context. This can not only help us better understand what it means for these biases to be reflective of real-world stereotypes but also help us outline any associations that were especially characteristic of the Indian context.

Word embeddings are rich vectorized condensations of textual data (Jurafsky and Martin, 2009), and hence ideal for matrimonial advertisements which are short, explicit, and dense in their purely textual descriptions of what people seek in their partners. They have previously been able to capture biases and cultural patterns in text (Garg et al., 2017), and exploit vectorization to measure cultural phenomenon against constructed dimensions of morality, affluence and so on (Kozlowski et al., 2019). I thus believe that word embeddings are well suited to capture changes in the preferences of individuals and their cultural aspects.

I use Gensim’s Word2Vec (skip-gram model, window size = 5, min\_count = 5) to train two separate embedding models for bride and groom seeking advertisements for all years. I choose to train one model for all years since annual embedding models would have an insufficient amount of advertisements to generate meaningful results. This also accounts for the sparseness of data scraped across years. I then create a dimension of gender by subtracting the normalized vector sum of ‘groom’, ‘boy’, ‘man’, and ‘male’ from that of ‘bride’, ‘girl’, ‘woman’, and ‘female’. I finally project exogenous word lists of adjectives and occupations on this constructed dimension of gender to capture any cultural biases in these advertisements. These word lists were previously used by Garg et al to capture gender stereotypes in 100 years of text (Garg et al., 2017), and were carefully curated and crowd-sourced to cover broad notions of gender roles.

### *Shortcomings of Word2Vec*

Word2Vec’s neural network is able to capture text semantics and is a popular way to compute dense embeddings. However, one of the limitations of word embeddings is that

it collapses words with different meanings into a single representation. Yet, this is not a major problem for this study and I tackle this by removing words such as “professional” which could either mean bride or groom, or a profession. Another shortcoming of training word embedding models from scratch is the dependency of the vocabulary of these models on the data that they have been trained on. This can raise two challenges. First, the construction of a gendered dimension is limited by words related to ‘female’ and ‘male’ appearing in the advertisements. However, I ensure that every possible word associated with these opposite genders is included in the dimension to thoroughly capture any gendered differences that appear when other constructs or words are projected on this dimension. Second, the set of exogenous words of occupations and adjectives to be projected on this dimension is also limited by the model’s vocabulary. This implies that any words that appear out of vocabulary have to be excluded from projections, thereby constraining the complete adaptability of an exhaustive word list that was developed for analysis for a model that was trained on a larger data set.

One possible solution to the second limitation is to use pre-trained embeddings or fine-tune pre-trained embedding models trained on a large corpus. Yet, these models are not suitable for capturing any bias in Indian matrimonial advertisements because they would not contain esoteric terminologies of caste, religion, and Indian English which are characteristic of Indian culture. Therefore, I attempt to account for this challenge by training one embedding model for all years, thus including all occupations and adjectives occurring across time between 2001-2014.

### *Qualitative analysis*

Lastly, I qualitatively analyze survey data collected from MTurk to understand the relevance and success of print and online media to help individuals find their lifelong partners through matrimonial advertisements.

While I stick to print media for the scope of this paper, a pivotal point in the history of matrimonial advertisements has been their transformation from print to digital media in the form of websites, and how this has affected their relevance and success. Thus, further research must extend this pipeline to online advertisements at a larger scale, to draw a more relevant and meaningful comparison of content between these different media.

## Results

This section synthesises some of my key findings about cultural patterns observed in matrimonial advertisements using several measurements and methods. It is divided into three parts – semantic investigation using part of speech tagging, validating cultural patterns using word embeddings, and findings from Mturk.

### *Semantic investigation using part of speech tagging*

In this section, I use part of speech tagging to analyze adjectives, nouns, and proper nouns in bride and groom seeking advertisements, and the insights they provide us with respect to individual preferences. Firstly, to understand how these advertisements articulate preferences, I use the adjectives in these texts. Adjectives are integral parts of language that can be very informative about the characteristics of a person. I specifically analyze adjectives associated with the word “girl” in bride-seeking advertisements, and those with the word “boy” in groom-seeking advertisements to understand preferences for partners<sup>7</sup>. Interestingly, the top ten adjectives in both categories of advertisements do not change drastically across years. Therefore, I only present the results of three years here. While these years are distinctly spaced out between 2001 and 2014, the results for all years can be produced by using the code in the GitHub repository. The results for the years 2002, 2007, and 2014 can be seen in the tables below.

---

7. The results of looking at adjectives associated with girl and boy in groom-seeking and bride-seeking ads were not very different and can be generated using the code in the GitHub repository.

2002	2007	2014
<b>professional</b>	<b>handsome</b>	<b>handsome</b>
<b>qualified</b>	<b>professional</b>	<b>qualified</b>
<b>handsome</b>	<b>qualified</b>	vegetarian
progressive	<b>tall</b>	unmarried
<b>cleanshaven</b>	<b>cleanshaven</b>	<b>suitable</b>
vegetarian	young	<b>cleanshaven</b>
cultured	Indian	bachelor
natured	local	<b>tall</b>
equivalent	skilled	smart
qual	<b>suitable</b>	Canadian

Table 1: Top ten adjectives associated with "boy" in groom-seeking ads

2002	2007	2014
<b>beautiful</b>	<b>beautiful</b>	<b>beautiful</b>
<b>qualified</b>	<b>qualified</b>	<b>qualified</b>
<b>professional</b>	<b>homely</b>	decent
<b>slim</b>	<b>professional</b>	sober
<b>homely</b>	<b>slim</b>	<b>slim</b>
<b>smart</b>	<b>smart</b>	<b>tall</b>
intelligent	<b>tall</b>	<b>fair</b>
<b>fair</b>	cultured	Indian
<b>tall</b>	attractive	<b>homely</b>
<b>pretty</b>	<b>pretty</b>	<b>smart</b>

Table 2: Top ten adjectives associated with "girl" in bride-seeking ads

There is a clear preference for handsome men and beautiful women. Additionally, while physical attributes such as tall and clean-shaven appear for men, those such as fair, and slim appear for women. Beyond mere physical appearance, qualities such as cultured and qualified appear across both genders, which is fairly common in the Indian matchmaking ecosystem. Moreover, women are also desired to be homely, over time. Homely in the Indian context broadly refers to well-cultured individuals. More specifically in the context of homely brides, these advertisements could potentially be referring to women who are well-cultured, respectful of the groom’s family, and willing to take care of their in-laws. These findings substantiate previous findings which look at matrimonial advertisements

on online websites (Rajadesingan et al., 2019). These observations are not surprising since these were discussions that took place as the matchmaking tradition originated in close-knit communities. However, individuals using matrimonial advertisements today, use these as media to not only be explicit about desires of physical attributes but also other behavioural qualities.

I then use part-of-speech tagging to analyze nouns associated with the word "boy" and "girl" in groom and bride-seeking advertisements respectively. I observe that nouns in these advertisements — besides words such as bride, groom, and family — can broadly be categorized as educational or occupational preferences. Educational preferences did not particularly vary between both categories, however, occupational preferences show an interesting pattern. While words such as engineer, lawyer, doctor, industrialist, teacher, lecturer, and professional appear in the lists of nouns associated with "boy" in groom-seeking advertisements— although with varying counts —, very few words such as nurse, and secretary appear in association with "girl" in bride-seeking advertisements, and occur with very low counts even if they do. Moreover, this pattern is consistent across years and does not exceptionally vary with time. This emphasizes the lack of occupational preferences for women and the possible underlying notion that men are considered to be breadwinners of the family, while women are expected to be homemakers. Therefore, occupational preferences appear to adhere to stereotypes of gender roles across 14 years from 2001 to 2014.

Lastly, I use part of speech tagging to extract proper nouns from the data. For texts as short as these advertisements, analyzing proper nouns is a simple and efficient method to capture esoteric words such as religion, caste, and other cultural nuances in the text. The approach indeed captures key preferences, and results primarily constitute preferences of religion or caste. Akin to the previous two results, these findings did not vary particularly across years. Moreover, they were very similar across both categories of advertisements.

This aligns with the Indian norm that caste and religion are essential factors for marriage irrespective of the gender of the individual. The results for 2002, 2007, and 2014 can be seen in the table below. While these years are distinctly spaced out between 2001 and 2014, the results for all years can be produced by using the code in the GitHub repository.

2002	2007	2014
<b>Sikh</b>	<b>Sikh</b>	<b>Sikh</b>
<b>Jat</b>	<b>Jat</b>	<b>USA</b>
<b>MBA</b>	Canadian	<b>Jat</b>
<b>Rajput</b>	Arora	<b>Khatri</b>
<b>Khatri</b>	<b>Manglik</b>	<b>Canada</b>
Arora	<b>BSC</b>	Hindu
<b>Canada</b>	Agrawal	Saini
<b>USA</b>	<b>MBA</b>	<b>Manglik</b>
MNC	<b>USA</b>	<b>Rajput</b>
India	<b>MSC</b>	Gursikh

Table 3: Top ten proper nouns associated in groom seeking ads <sup>8</sup>

Several words that appear in the results — Sikh, Jat, Hindu, Rajput, Manglik, Khatri, Arora and Gursikh — are some of the major religions and communities within these religions in northern India. While Sikh and Hindu are religions found in the majority of most states that the newspaper is distributed in, the remaining words are communities within these religions. A deeper analysis shows that communities such as “Muslims” and “Christians” or any other communities not found in northern India rarely occur in the corpus, thus highlighting a preference for intra-faith alliances. While one might question the value of this result because of the lack of prevalence of other communities in the region, Figure 1 below showcases the proportion of bride-seeking advertisements that mention preferences for individuals from the Sikh religion or the Jat community (of the Hindu and Sikh religion).

8. Results do not remarkably vary in bride-seeking ads across years

9. Results do not drastically vary in groom-seeking ads across years. Ads from 2001-2004 were omitted from this figure since they were lesser in number compared to other years and seemed almost zero against the scale of the y-axis.

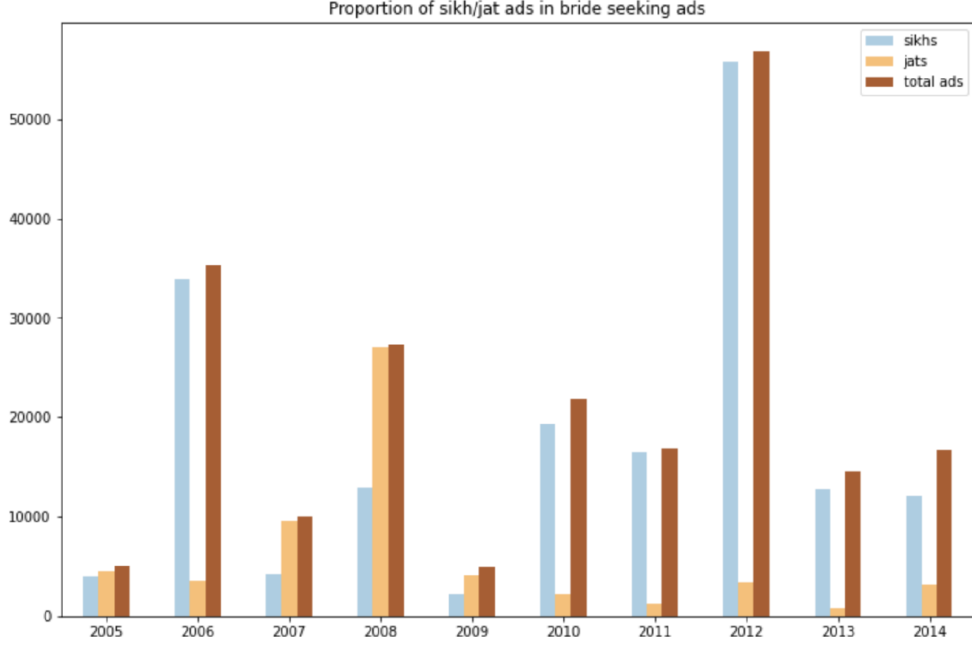


Figure 3: Proportion of sikh/jat ads in bride seeking ads <sup>9</sup>

It is apparent that almost every advertisement emphasizes the preference of religion or community across years. It is also important to note that while the counts for advertisements with preferences for "Sikh" or "Jat" vary across years, their general importance is consistent over the years since at least one of the two is commensurate with the total number of advertisements across time. This highlights the sheer importance of religion and communal factors in the process of seeking matrimony. Moreover, the fact that this trend is uniform across both categories of advertisements indicates the weight of intra-faith alliances in the matchmaking process.

An important structure that appears across all parts of the part-of-speech tagging analysis above is the difference in variation of words between bride-seeking and groom-seeking advertisements. For instance, in Table 1, 50% of the adjectives are common across two or all three years over time in groom-seeking advertisements. On the contrary, in Table 2, 80% of the adjectives are constant over two or all three years over time in bride-



seeking advertisements <sup>10</sup>. This raises questions about the static nature of gender roles, characteristics, and traits associated specifically with women and potential shifts in those associated with men. For instance, in groom-seeking advertisements, the word ‘cultured’ appears in 2002, but not in 2007 or 2014. However, in bride-seeking advertisements, the word ‘homely’ (which is closely associated with ‘cultured’ as was mentioned above), appears in 2002, 2007, and 2014. Moreover, amongst words that repeat in both categories of advertisements, only 50% words — ‘handsome’, ‘cleanshaven’, and ‘tall’ seem to be directly associated with physical attributes in groom-seeking advertisements. Whereas 55% of the words — ‘beautiful’, ‘slim’, ‘fair’, ‘tall’, and ‘pretty’ seem to be directly associated with physical attributes in bride-seeking advertisements. This underscores the potential persistent focus on the physical objectification of women compared to a reduced focus on that of men. We can also see that 70% of the proper nouns remain constant in Table 3. While ‘Sikh’, ‘Jat’ and references to other communities remain constant across all three years, there appear no words associated with education (‘MBA’, ‘BSC’, ‘MSC’) in 2014. However, it is challenging to claim that education was no longer an important desire in partner-seeking behaviour without further content analysis of the advertisements.

Another key observation across all findings thus far has been the consistency of patterns across the years. I, therefore, use KL divergence to measure changes amongst advertisements and quantitatively validate these findings. The results can be seen in Figure 2. It is clear that apart from the initial years with lesser data, the divergence is under .5 for most years in both categories of advertisements. A lower divergence implies a lesser loss of information if the distribution of advertisements in one year was used to extrapolate the distribution of advertisements in another year. Hence, the divergence is high for years with lesser data. A lower divergence can also be interpreted as a lack of change in the content of advertisements thereby validating the observation of consistency of cultural

---

10. Words that repeat appear in bold in all tables for ease of observation

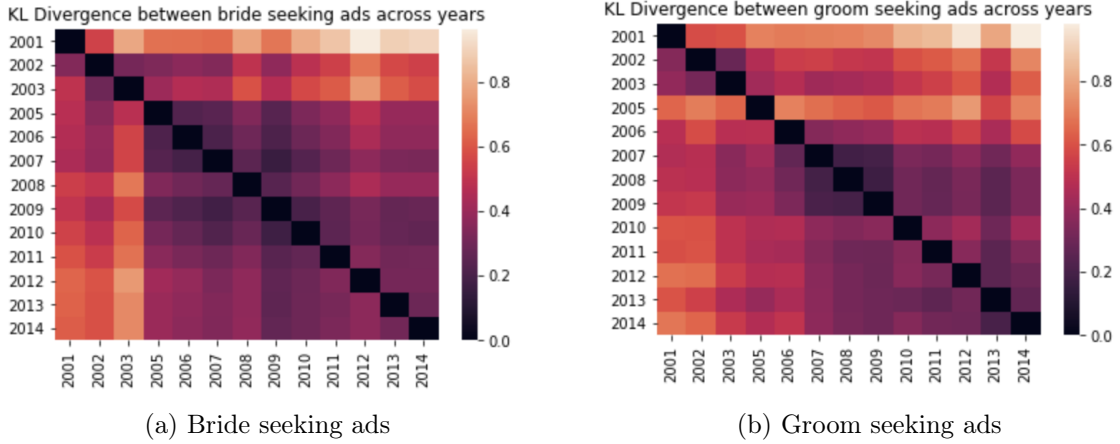


Figure 4: KL divergence between ads across years

patterns across years. It is also important to note that the given figures represent the loss that would occur if a year along the column was used to interpret the distribution of data for a year along the row. KL divergence is not symmetrical along directions and hence, the loss of using 2001 advertisements to extrapolate information about advertisements of 2012 is not the same as the loss computed the other way around.

### *Validating cultural patterns using word embeddings*

In this section, I utilize word embeddings to recognize cultural bias that was identified using part of speech tagging earlier, using adjectives and occupations from external settings. In order to validate the presence of gender bias and stereotypes of gender roles as observed in the above section, I use word embedding models trained on both categories of advertisements separately. I project an exogenous collection of adjectives and occupations on a dimension of gender to analyze the underlying associations of these words with the male and female gender in matrimonial advertisements. The results of the projection can be seen in the figures below.

It is apparent that even with an extrinsic projection of adjectives and occupational preferences, the embedding model underscores the cultural bias present in these adver-

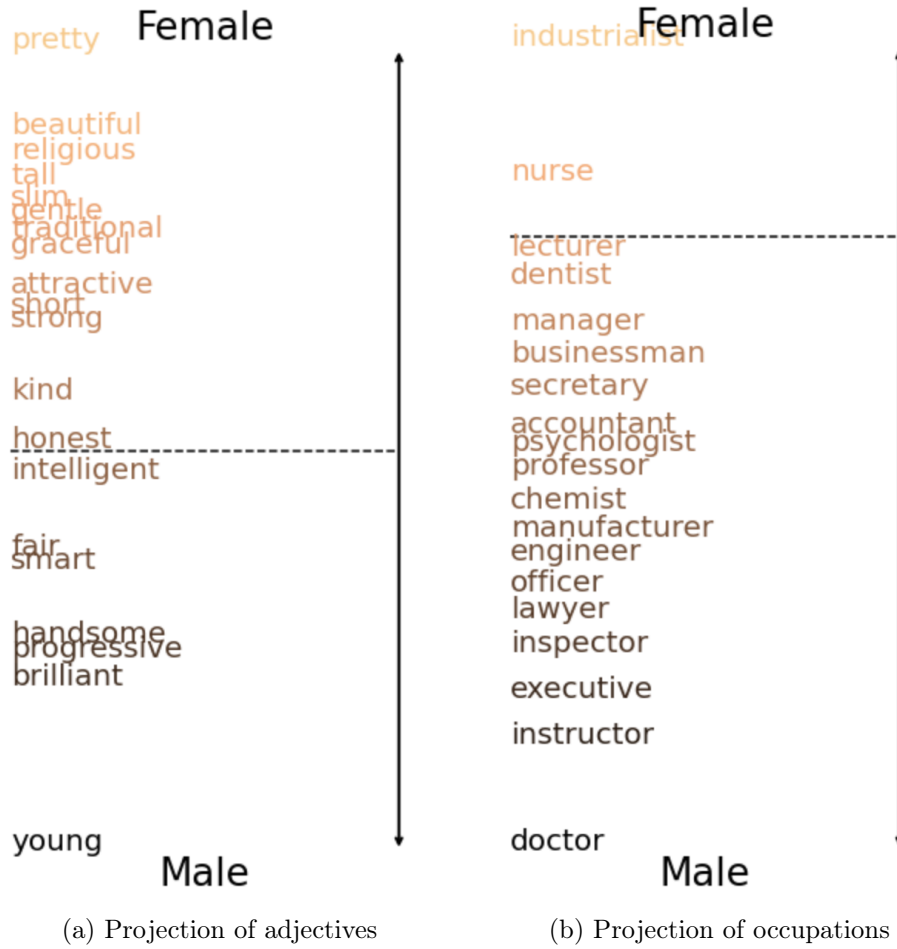
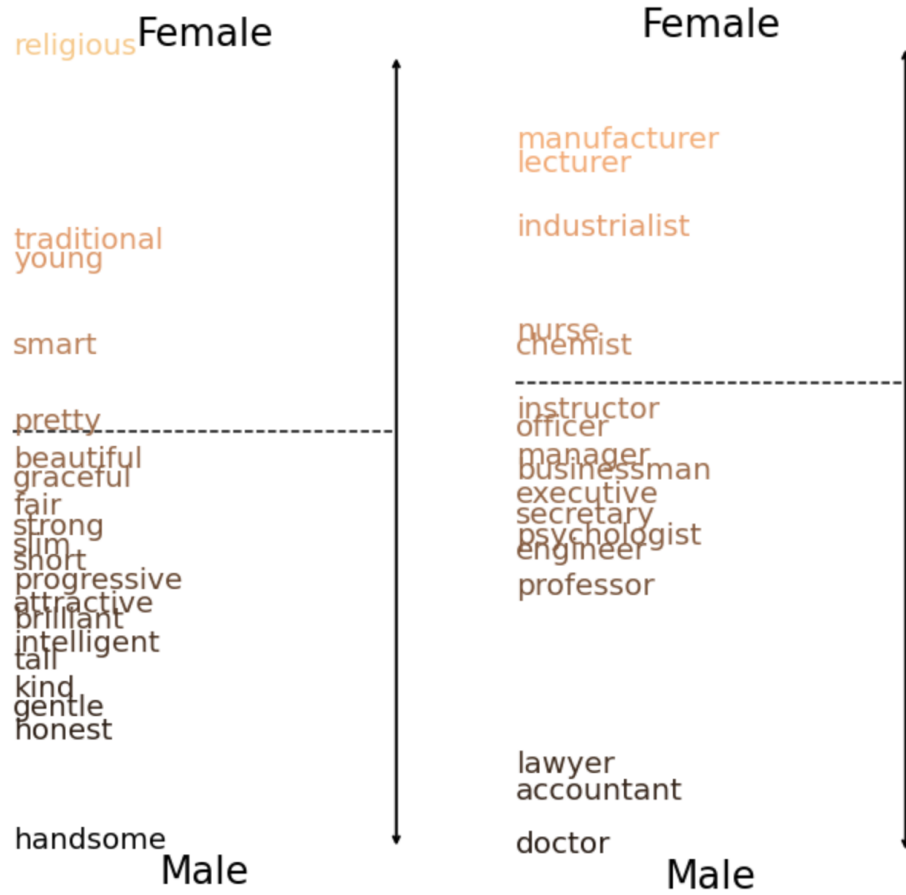


Figure 5: Projection of words on a dimension of gender in bride-seeking ads: The black dashed line represents the value of a gender-neutral projection

tisements (Figure 5). For bride-seeking advertisements, the findings align with those of part-of-speech tagging. Adjectives such as "pretty", "beautiful", "religious", and "traditional" are associated with the female aspect of the dimension, whereas those such as "handsome" and "intelligent" are associated with the male aspect. Similarly, we only see two occupations associated with the female aspect, whereas most of the other occupations cluster towards the male side of the gendered dimension. This reiterates the fact that these advertisements adhere to stereotypical gender roles in their preferences. More specifically, they also align with broader perceptions of gender roles, outside the Indian context.



(a) Projection of adjectives

(b) Projection of occupations

Figure 6: Projection of words on a dimension of gender in groom-seeking ads: The black dashed line represents the value of a gender-neutral projection

In the case of groom-seeking advertisements, while adjectives such as "traditional", "religious", and "pretty" cluster around the female aspect, "beautiful", and "graceful" slightly shift towards the male side (Figure 6). This is potentially because of the short nature of these advertisements and how they are structured. For instance "beautiful girl seeks groom", can be interpreted as beautiful being in close proximity of groom, which drives it towards the male aspect. Similar reasoning can be extended to the occurrence of an increase in the number of occupational preferences associated with the female side. This observation is bolstered by a qualitative analysis of these advertisements, wherein groom-seeking advertisements are occasionally more prescriptive of desired preferences of a man when talking about the bride. This could be a possible reason for any discrepancy between results of the two embedding models.

### *Findings from MTurk*

Finally, I use MTurk<sup>11</sup> to qualitatively understand the relevance and success of newspaper advertisements compared to that of online advertisements. I present details of my MTurk survey and its results here. The entire survey can be found in the Appendix. The sample of respondents on MTurk consisted of 60 individuals who successfully found their partners through Indian matrimonial advertisements between 2005 and 2022. Among those that participated, 50% of individuals identified as male, 48% identified as female, and the rest preferred to not disclose their gender. Moreover, 80% of these individuals used both newspaper and online advertisements to seek matrimony, 15% of them used only newspaper advertisements, and 5% of them used only online advertisements. This highlights the continued use of newspaper matrimonial advertisements. Approximately, 70% of the people who used both newspaper and online advertisements successfully found their matches only through online advertisements. Among these, 40% of the individu-

---

11. This study was determined to be exempt from review by the Institutional Review Board (IRB) of the University of Chicago

als identified as female and 60% identified as male. Therefore, there might be gendered differences among those that successfully find matches using born-digital advertisements. However, it is difficult to make such a claim based on such a small sample. Additionally, those who used only online advertisements or matched with partners through them, found these platforms to be easy to navigate, easy to use, and permitting customized search. Those who employed only newspaper advertisements did not explicitly elaborate on why advertisements in newspapers were a successful medium. Therefore, while newspaper advertisements continue to be used, their relevance might be decreasing when compared to that of online platforms. However, this finding could also mean that newspaper advertisements continue to be used, and remain relevant amongst the Indian population that does not have access to internet and cannot be reached via MTurk.

About 17% of the respondents who used both modalities claimed to be from Tamil Nadu, 6% from Delhi, 4% from Kerala, 12% claimed to be married in India but currently living in the United States, and the rest spanned across states of Karnataka, Rajasthan, and Maharashtra in India. All of the mentioned Indian states barring Delhi and Rajasthan are South Indian states. This differs from the predominantly northern states studied using the analysis of newspaper advertisements. Therefore, there could be a potential difference in how media is used and facilitates matchmaking in India across geographies. Yet, this needs to be probed to be able to make concrete conclusions.

## Discussion

This research paper started out as an attempt to understand key preferences of men and women seeking partners through newspaper matrimonial advertisements, and how these preferences have changed over time. The results highlight that individual preferences of people seeking matrimony can be bucketed into three categories of physical appearance, religion or caste, and occupational preference. They show that matrimonial advertise-

ments are reflective of real word cultural biases and stereotypes of gender roles. They also underscore how preferences for both men and women can be objectifying when it comes to physical attributes, but extend to specific qualities for women, which can be subjective, and individualistic choices (e.g. *homely*). We also see a constant emphasis on physical appearance specifically for women over the years. Additionally, this research sheds light on the lack of or meagre occupational preferences for women, as opposed to those of men which range from businessman to doctor. Previously, Ramasubramanian and Jain used around 1500 newspaper advertisements to find support for the social exchange of men’s financial stability for women’s physical attractiveness, and gender polarization in ideal spousal occupations (Ramasubramanian and Jain, 2009). More importantly, this observation ties in with this paper’s finding that draws attention to preferences entrenched in real-life biases in Indian society. It also underscores the larger Indian mindset which assigns different weights to a woman’s physical appearance and values and a man’s occupational designation. Moreover, as mentioned in the background literature earlier, sociological research has shown that propinquity, age, race, religious faith, ethnic background, education, and socioeconomic status are key determinants of mate selection and that women accord more weight to socioeconomic status — which can be associated with occupation — than men (Buss, 1985; Feingold, 1992). Therefore, the results of this research also align with broader theories of mate selection.

To build on the importance of religion, the extensive use of religion and religious communities in a sizeable proportion of matrimonial advertisements across years is indicative of the instrumental nature of religion and intra-faith alliances in the matchmaking process. This research, more importantly, showcases how these cultural patterns and preferences remain conservative and have not remarkably changed over a span of 14 years between 2001 and 2014. This also ties in with relevant works in the sociology of mate selection in the United States which showcase that religion remains relevant in sorting partners for

many young adults (McClendon, 2016). Thus, these cultural patterns seem to transcend cultures.

Lastly, this study also provides insights about the potential decrease in the success and relevance of newspaper advertisements amongst populations that have access to internet and are using online platforms for advertisements. It sheds light on the potential gendered disparity between users who successfully find mates using born-digital advertisements, and the role that geographical distribution of individuals might play in determining what modality of advertisements enables people to find partners successfully. This then calls for further investigation into if born-digital matrimonial advertisements facilitate cultural biases in matchmaking, and if and how different forms of media of matrimonial advertisements help different people differently.

One important avenue for extension of such analysis of born-digital advertisements in today's realm is dating websites and applications. While these websites might not have evolved to digital spaces from analog advertisements, they facilitate mate selection amongst individuals in ways similar to matrimonial websites. There is an immense body of research in the context of online dating websites that finds similar patterns across romantic relationships and factors determining preferences amongst partners. Previous studies have found that men and women on dating websites exhibit strong same-race preferences when deciding to contact a potential partner (Hitsch et al., 2010). This preference also does not vary much across age, income, and education levels. Additionally, women have a stronger preference for income than men for income over physical attributes. This aligns with the findings of this study where men do not demonstrate specific occupational preferences for potential brides while women seek grooms with stereotypical occupations which could be concomitant with income. Past scholars have also shown that racial hierarchies emerge amongst heterosexual internet daters on a major online dating website in the United States. (Lin and Lundquist, 2013). Similarly, racial homophily has



been discovered in online dating across European countries (Potârca and Mills, 2015). This corroborates observations from the current project where both, bride-seeking and groom-seeking advertisements depict the desires of potential partners from certain major communities and castes. While these patterns are potentially indicative of the fact that digital spaces enable cultural biases akin to newspaper advertisements, most research in the space of dating websites too, is ethnocentric in the West. This then calls for future research that further tries to probe the role of media in facilitating mate selection and social norms. Thus, this work builds on findings and raises questions about their implications on gender, race, physicality, and other societal constructs, that transcend the Indian matchmaking context. It also probes the generalizability of sociological theory, and why individuals adhere to such societal stereotypes when seeking partners in different cultures whether they intend to marry or date them.

In the context of methods, this research successfully reproduced some of the earlier findings in literature and discovered new ones on a large scale dataset of newspaper advertisements. Moreover, unlike previous usage of statistical and analytical methods, this study uses potent methods of computational linguistics such as part-of-speech tagging to identify cultural biases and stereotypes in newspaper advertisements across the years. It also employs robust methods such as word embedding models to validate these findings and understand some of the layers of bias in the underlying advertisements. This project adds to the literature of social sciences that attempts to understand how computational methods can be used to capture and discuss societal stereotypes and biases. It exploits the short nature of matrimonial advertisements to use methods such as part-of-speech tagging to extract larger meanings for the social sciences from associations of adjectives, nouns, and proper nouns. Furthermore, it uses dimension reduction methods such as word embeddings on these short texts and is able to replicate cultural and human bias found in the Indian matchmaking process. It, therefore, extends the ability of word embeddings to

capture bias to shorter and more nuanced text —matrimonial newspaper advertisements in Indian English.

Yet, the contributions of this study go beyond theoretical and methodological ones. They raise larger questions about the prevalence of stereotypes around relationships, and gender identities and associations in India. These questions engender discussions and deliberations multifaceted in nature. First, how reflective are these stereotypical biases found in advertisements of real-life conceived notions of gender roles in Indian communities? For example, despite the increased participation of women in the workforce, women are still continued to be looked at as homely and lack any acknowledgement or requirement of their ability to work in these advertisements. On the contrary, men are often associated with their financial stability and occupational class. This ensues dialogue and contemplation of reasons for prolonged gender biases, and how they translate into preferences for matrimony.

Second, the extant nature and relevance of these advertisements across time also probe the emphasis of continued ideas of heterosexual alliances and expectations in Indian society. It is important to note here that this observation does not simply imply the absence of advertisements seeking homosexual alliances —which is still not legally recognized in India. Rather, this means that there seems to exist an apparent societal consensus about an ideal man or woman, with some gender polarization. Not only does this impose constraints on people seeking matrimony, but it also excludes people who do not fit into these traditional norms of heterosexual relationships from the matchmaking process (Ramasubramanian and Jain, 2009).

Third, what is the role of these matrimonial advertisements in shaping or advancing perceptions of gender identities in India? According to a UNICEF report on gender bias and inclusion in Indian advertising, the portrayal of women in Indian advertisements was

considered problematic because it furthered gender stereotypes. More recent research has shown that advertisements can also lead to gender socialization wherein individuals internalize societal norms (Gutmann et al., 2022). This is more so in the case of matrimonial advertisements which endorse individual preferences and desires. Therefore, cultural patterns observed in these advertisements not only lead to questions of existing societal prejudices but also about their broader implications on instigating and carrying forward these biases.

Lastly, how does preferential seeking through matrimonial advertisements compare against that on online dating platforms? While physical appearance and occupational preferences are common themes across both modalities, the prevalence of religion and caste-driven desires seems more intrinsic to the Indian context. However, most studies on online dating platforms are not based on Indian populations, and thus it might be hard to make a definitive claim without further studying the cultural preferences of individuals on online dating platforms in India. This gives rise to two questions instrumental to understanding shifts in individual preferences. First, is there an inherent difference between people’s preferences when they seek partners with the intention to marry versus date them? If not, are the foundations for what individuals seek in partners generally, set in roots of bias and stereotypes? Secondly, what implications would these findings have on notions of gender, race, occupation, body objectification and so on, as the world becomes interconnected and partner-seeking becomes a more inter-cultural phenomenon? These are important questions to answer to understand the importance of this study against the broader sociological literature.

### *Limitations*

This research does have certain limitations. Firstly, the sample of advertisements for the study is collected from a newspaper distributed in six states of northern India, thereby

making it less representative of the Indian population, and restricting the generalizability of these results. Moreover, these advertisements were scraped from digital archives of the website of The Tribune from 2001 to 2014. The website of the newspaper has changed over time, and this could have resulted in inaccuracies in the data collected. This is closely associated with the embedding models where the vocabulary for different years could be misrepresented in one model trained on all years. Additionally, the word lists used to construct the dimensions were primarily used from existing literature where they are not necessarily used in the Indian context, and this could affect the meaningfulness of the results. Furthermore, there could also be limitations to using Word2Vec as opposed to other algorithms for training embedding models, and this can induce its own systemic bias. Lastly, while the study aims to capture the relevance of newspaper and online advertisements, in restricting itself to users from MTurk, this research is exposed to a possible selection bias. Workers on MTurk have access to internet and might have biased opinions towards using online platforms when compared to the population of India that does not have access to internet.

### *Future research*

Future research must consider data from a more representative sample to extrapolate the external validity of these findings. It must also try and reach out to a larger diaspora of the Indian population to capture a representative opinion about the relevance and success of different media of matrimonial advertisements. Further studies must also explore the literature and data sources thoroughly to develop exogenous word lists for constructing dimensions which are more nuanced and catered to the Indian context. Moreover, other algorithms such as FastText or GloVe must be used to compare their ability to capture some of the larger semantic structures along preferences of physical attributes, religion, and occupation. Future work can also potentially compare how these methods and observations fare for data beyond 2014, and online advertisements, to truly establish the relevance of

these findings and if and how some of these key observations regarding cultural bias in the Indian matchmaking ecosystem have changed over time. More importantly, future research must extend this study to contexts outside of Indian matchmaking. It can compare how the narrative observed in Indian matrimonial advertisements fits against the paradigm of research in other cultures in similar domains. It can also draw from theories of feminism, race, objectification, perceived bias in gender roles in relationships, and several other rich theoretical frameworks in sociology, anthropology, and psychology that can qualitatively and theoretically corroborate the origins and reasoning of the findings of this paper.

## Conclusion

The Indian matchmaking tradition has come a long way from discussions amongst kin of brides and grooms, to newspapers, and finally to online platforms. However, these advertisements continue to be explicit in their preferences for persons with specific physical attributes, occupations, and religious affiliations. Since the tradition has adjusted its medium to changes in demographics, generations, and time, one might expect preferences of individuals to change with these changes in circumstances. However, surprisingly, as of 2014, preferences of men and women remain consistent across time. This reflects the potentially conservative nature of the Indian matchmaking landscape — at least in Northern India —, and its conformity towards orthodox beliefs.

This then raises questions about how gender roles of men and women are perceived in society and the accordance between this societal perception or preference, and ground reality of gender roles. In order to understand why these preferences and perceptions of individuals continue to remain consistent across time, it is important to understand broader sociological aspects of the Indian matchmaking process. This research hoped to augment existing sociological frameworks on such sociological aspects of mate selection

and understand how analog and digital spaces facilitate matchmaking and norms around matchmaking. The majority of research in these areas is currently ethnocentric in the West, and this study hence hoped to extend such research to Indian contexts.

Results from this project indicate that gender and geography might influence one's ability to find mates through a certain modality. In order to understand such complex processes, it is essential to understand extensive social systems such as the roles of families in matchmaking in future research. If families are instrumental, are preferences of individuals reflective of their families' preferences, or is the lack of change of preferences across years a function of continued cultural biases of older generations? Deliberating upon such questions will further substantiate conclusions from this paper.

Lastly, it is imperative to understand how the arranged marriage ecosystem itself has changed over time. For example, how has the process changed beyond seeking matrimony through advertisements? There is a possibility that with changing times, people use matrimonial advertisements to seek partners, but are open to other channels of seeking individuals such as dating applications, or other avenues. While scholarship has shown similar patterns in online dating avenues for other cultures, it is important to understand its significance in the Indian context to derive further conclusions about how partner-seeking in India compares to that in other cultures.

Hence, while this research makes key contributions towards understanding the content and relevance of newspaper matrimonial advertisements over time, it raises points of deliberation around several broader questions of the sociology of preferential seeking of partners and implications of Indian matchmaking on the Indian societal fabric. Future work that delves into these questions could give meaningful insights into a holistic understanding of the Indian matchmaking environment. Therefore, this paper aims to draw the attention of readers to the potential implications of the continued prevalence

of cultural biases in matrimonial advertisements and calls for further investigation into larger sociological factors that can extend our understanding of the Indian matchmaking landscape.

### **Data and Code Availability Statement**

All the data and code used to generate the results in this paper can be found at this GitHub repository (Iyer, n.d.)

## REFERENCES

- Anand, K. (1965). An analysis of matrimonial advertisements. *Sociological bulletin*, 14(1), 59–71.
- Auletta, K. (2013). 13. why india’s newspaper industry is thriving. *The best business writing 2013* (pp. 281–304). Columbia University Press.
- Bolton, C. D. (1961). Mate selection as the development of a relationship. *Marriage and Family Living*, 23(3), 234–240.
- Buss, D. M. (1985). Human mate selection: Opposites are sometimes said to attract, but in fact we are likely to marry someone who is similar to us in almost every variable. *American scientist*, 73(1), 47–51.
- Chattopadhyay, D., & Chattopadhyay, S. (2019). Colorism and love for fair skin: Exploring digitization’s effect on india’s arranged marriage matrimonial advertisements. *Asian Communication Research. Academia. edu*.
- Dash, M., & Belgaonkar, P. (2012). Comparative effectiveness of radio, print and web advertising. *Print and Web Advertising (August 15, 2012)*.
- Feingold, A. (1992). Gender differences in mate selection preferences: A test of the parental investment model. *Psychological bulletin*, 112(1), 125.
- Garg, N., Schiebinger, L., Jurafsky, D., & Zou, J. (2017). Word embeddings quantify 100 years of gender and ethnic stereotypes. *CoRR*, abs/1711.08412. <http://arxiv.org/abs/1711.08412>
- Gupta, G. R. (1976). Love, arranged marriage, and the indian social structure. *Journal of Comparative Family Studies*, 7(1), 75–85.
- Gutmann, B., Jha, S., O’Doherty, E., & Banerji, R. (2022). Selling stereotypes: Reviewing the impact of business advertisements on gender norms and socialization. *Business and Human Rights Journal*, 7(1), 157–162.
- Hill, R. (1958). Sociology of marriage and family behaviour, 1945-56. *Current Sociology*, 7(1), 1–33.



- Hitsch, G. J., Hortacsu, A., & Ariely, D. (2010). What makes you click?—mate preferences in online dating. *Quantitative marketing and Economics*, 8, 393–427.
- Hollingshead, A. B. (1950). Cultural factors in the selection of marriage mates. *American Sociological Review*, 15(5), 619–627.
- Hunter, E. (2007). Change and continuity in american marriage. *Handbook of the new sexuality studies* (pp. 217–222). Routledge.
- Iyer, P. (n.d.). *Deciphering Indian matrimonial advertisements*.
- Jaiswal, T. (2014). *Indian arranged marriages: A social psychological perspective*. Routledge.
- Jurafsky, D., & Martin, J. H. (2009). *Speech and language processing (2nd edition)*. Prentice-Hall, Inc.
- Kozlowski, A. C., Taddy, M., & Evans, J. A. (2019). The geometry of culture: Analyzing the meanings of class through word embeddings. *American Sociological Review*, 84(5), 905–949.
- Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The annals of mathematical statistics*, 22(1), 79–86.
- Lin, K.-H., & Lundquist, J. (2013). Mate selection in cyberspace: The intersection of race, gender, and education. *American Journal of Sociology*, 119(1), 183–215.
- Lindstädt, N., & Budzinski, O. (2011). Newspaper vs. online advertising—is there a niche for newspapers in modern advertising markets? *Online Advertising—Is There a Niche for Newspapers in Modern Advertising Markets*.
- McClendon, D. (2016). Religion, marriage markets, and assortative mating in the united states. *Journal of Marriage and Family*, 78(5), 1399–1421.
- Potârca, G., & Mills, M. (2015). Racial preferences in online dating across european countries. *European Sociological Review*, 31(3), 326–341.

- Rajadesingan, A., Mahalingam, R., & Jurgens, D. (2019). Smart, responsible, and upper caste only: Measuring caste attitudes through large-scale analysis of matrimonial profiles. *CoRR*, *abs/1904.04176*. <http://arxiv.org/abs/1904.04176>
- Ramakrishnan, S. (2012). ‘wheatish’grooms and ‘innocent’divorcées: Commodifying attributes in the discourse of indian matrimonials. *Discourse & Society*, *23*(4), 432–449.
- Ramasubramanian, S., & Jain, P. (2009). Gender stereotypes and normative heterosexuality in matrimonial ads from globalizing india. *Asian Journal of Communication*, *19*(3), 253–269.
- Ramchandani, I. (2000). *Students’ britannica india* (Vol. 1). Popular Prakashan.
- S., R. (2018). Urban indians still get married the way their grandparents did. <https://www.livemint.com/Politics/mnVzCfIEbqvzEu01LTxqLM/Urban-Indians-still-get-married-the-way-their-grandparents-d.html>
- Schubert, L. (2020). Computational Linguistics. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring 2020). Metaphysics Research Lab, Stanford University.
- Seth, N. (2011). Online matrimonial sites and the transformation of arranged marriage in india. *Virtual communities: Concepts, methodologies, tools and applications* (pp. 951–974). IGI Global.
- Stuers, C. V.-D. (1969). The relevance of matrimonial advertisements for the study of mate selection in india. *Bijdragen tot de Taal-, Land-en Volkenkunde*, (1ste Afl), 103–117.
- Sweeney, M. M. (2002). Two decades of family change: The shifting economic foundations of marriage. *American Sociological Review*, 132–147.
- Titzmann, F.-M. (2011). Medialisation and social change: The indian online matrimonial market as a new field of research. *social dynamics*, *2*, 49–66.

- Upreti, H. (1967). Matrimonial advertisements: A brief sociological analysis. *Journal of Family Welfare*.
- Voutilainen, A. (2003). Part-of-speech tagging. *The Oxford handbook of computational linguistics*, 219–232.

# Appendix

## *Mturk survey*

01/05/2023, 13:35

Have you used matrimonial advertisements?

If you have questions or concerns about the study, you can contact the researchers at Pranathi Iyer, [pranathiiyer@uchicago.edu](mailto:pranathiiyer@uchicago.edu)

If you have any questions about your rights as a participant in this research, feel you have been harmed, or wish to discuss other study-related concerns with someone who is not part of the research team, you can contact the University of Chicago Social & Behavioral Sciences Institutional Review Board (IRB) Office by phone at (773) 702-2915, or by email at [sbs-irb@uchicago.edu](mailto:sbs-irb@uchicago.edu).

### **Consent:**

Participation is voluntary. Refusal to participate or withdrawing from the research will involve no penalty or loss of benefits to which you might otherwise be entitled.

By clicking "Agree" below, you confirm that you have read the consent form, are at least 18 years old, and agree to participate in the research. Please print or save a copy of this page for your records.

*Mark only one oval.*

- ☐ I agree to participate in the research
- ☐ I do not wish to participate

2. Have you been married/are seeking partners through matrimonial ads/websites? \*  
(not eligible if answer is no)

*Mark only one oval.*

- ☐ Yes
- ☐ No

**You are eligible to participate!**

<https://docs.google.com/forms/d/1bIUvakEwsCJ--ApjHZjJDEYaEaB6INtR-SCnAM9fWc/edit>

3/7

# Appendix

## *Mturk survey*

01/05/2023, 13:35

Have you used matrimonial advertisements?

3. What year did you use matrimonial advertisements in? \*

*Mark only one oval.*

- ☐ 2000-2005  
☐ 2005-2010  
☐ 2010-2015  
☐ 2015-2020  
☐ After 2020

4. What gender do you identify with? \*

*Mark only one oval.*

- ☐ Male  
☐ Female  
☐ Prefer not to say

5. Are you an Indian \*

*Mark only one oval.*

- ☐ Yes, I stay in India      *Skip to question 11*  
☐ Yes, I am an NRI  
☐ No

6. Did you use newspaper matrimonial advertisements? \*

*Mark only one oval.*

- ☐ Yes  
☐ No

<https://docs.google.com/forms/d/1bIUvakEwsCJ--ApjHZjJDEYaEaB6INtr-SCnAM9fWc/edit>

4/7

# Appendix

## *Mturk survey*

01/05/2023, 13:35

Have you used matrimonial advertisements?

7. Did you use online matrimonial advertisements? \*

Mark only one oval.

☐ Yes

☐ No

8. Were you successful in finding your partner through the service? \*

Mark only one oval.

☐ Yes, through newspaper ads

☐ Yes, through online ads

☐ Yes, through both

☐ No

9. If yes, why was the platform successful in helping you? \*

---

---

---

---

---

10. Enter your worker ID \*

---

Where in India?

11. Which state in India are you from? \*

---

This content is neither created nor endorsed by Google.

<https://docs.google.com/forms/d/1bIUvakEwsCJ--ApjHZjJDEYaEaB6INtR-SCnAM9fWc/edit>

5/7