Holding Space for Voices that Do Not Speak:
Design Reform of Rating Systems for Platforms in GREAT Economies

Anuttama Dasgupta
Indian Institute for Human Settlements
adasgupta@iihs.ac.in

Prasanna P. Karhade
University of Hawai'i at Mānoa
karhade@hawaii.edu

Abhishek Kathuria
Indian School of Business
abhishek_kathuria@isb.edu

Benn R. Konsynski
Emory University
benn.konsynski@emory.edu

Abstract

Researchers can examine ethical implications of online rating systems to understand how they function as 'knowledge instruments' and affect social relations and networks connected with them. Research should address the fact that the underlying economic structures that design and deploy knowledge producing 'technical objects' on online platforms are not egalitarian and may create new circles of exclusion. Exploring implications of this for a starkly unequal country like India, we illustrate our ideas by integrating induction and abduction to study rating systems on a pan-India food discovery and delivery platform. Rating systems are borrowed from WEIRD contexts and our findings imply that the instrument studied here is designed to hear only some of many voices. Consequently, they might be 'institutionalizing' knowledge that is problematic for GREAT domains in which they are imposed. We highlight the need for decolonization of research approaches for GREAT domains and critical research of technical knowledge objects.

1. Introduction

In a Digital and Social Media (DSM) intensive world, all users are not equal and in India, where stark inequalities of income and opportunity already exist between urban and rural populations and across genders and social groups [1], this inequality is exacerbated by their ability to afford, access and participate on DSM platforms. In terms of getting their voices heard, even decades after the term was first used in postcolonial literature, the subaltern is continuing to be un-made and re-made by the deployment of technical objects like rating systems on digital platforms that are not 'fluid' [2]. In trying to understand why a relatively small portion of population is venturing its feedback on an online food delivery platform, the idea of a 'capability' is drawn upon [3]. Participation is not always a simple choice

of giving a rating or not. The difference may be likened to that between someone who is fasting and someone who is starving. The one who is fasting has a choice to not fast, but the one who is starving may not have the freedom to choose. In a country that is beset with fundamental developmental inequalities, this is a framing that researchers of DSM can explore [4].

Economic advisors to the Prime Minister (PM) of India over the past decade have all believed that innovation and entrepreneurship, especially technical entrepreneurship is the 'silver bullet' that will solve India's development and inequality problems [5]. But "in casting street hawkers and technological elites alike as entrepreneurs in potentia, proponents collapse the vast gaps in money, formal knowledge and authority that separate the two" [5]. Narendra Modi, won the 2015 national elections in India in a sweeping majority on the key promise of "Acche din" (literally translated as "good times" but figuratively meaning prosperity and wellbeing.) Through his first and second term in office, the push for "Digital India" has had an impact on the lives of many Indians but at the same time, for the first time in the history of the nation, the Indian Economy is facing a recession and Indian society is beset by social and communal polarization. At this juncture, it is pertinent to ask what is "accha" or "good" for everyone, and if structural inequalities are being institutionalized in digital infrastructure like DSM platforms, it is necessary to conduct 'critical research' [6] of the same.

If decades of discourse on post-colonialism has argued that the subaltern has difficulty speaking and also being heard, has the digital world allowed those without a voice so far to speak freely and communicate with all others as equals? What role has access to technology and the ability to afford certain types of technology played in facilitating this conversation? What kind of knowledge is getting institutionalized when such knowledge creating technology is in

HICSS

common use? Are rating systems on platforms creating new digital elites and subalterns out of potential users? Is it widening existing circles of exclusion and / or creating new ones? Using English as the medium of communication is just one circle of exclusion as only 12% of India's urban population speaks English.

Are a lot of voices getting lost in the differential; represented in DSM as "missing" data/ voices that have not spoken and hence not been accounted for? Defining context as Western, Educated, Industrialized, Rich, and Democratic - WEIRD [7, 8] or Growing, Rural, Eastern, Aspiring, and Transitional - GREAT [9],─we suggest that when technical objects are deployed from one context to the other, researchers should study how they are effecting the social relations and networks in the context where they are deployed and what broader implications to the receiving society this is having. Not everyone in the WEIRD context is rich or democratic. The largest democracy in the world is a GREAT economy, and although aspiring of "Acche din" and economic prosperity, is seeing a contraction of economic growth. DSM's power in interrogating these issues cannot be over emphasized but its true potential can be realized only through proper participation. Discourses on the 'decolonization of research' also talk about the role space plays in participatory research, not only social but contextual and physical [10]. Part of this context, as in the submerged part of the iceberg whose tip in being seen lends itself to a popular proverb, is in the differential.

We also look at theory from various intersectional disciplines: post-colonial studies from where we revisit Spivak's essay "Can the Subaltern speak" [11], and education and philosophy from where we look at Paulo Freire's "Pedagogy of the Oppressed" [12]. We make note of Southern Urban Theory that urges practitioners and academics to "dislocate the center of research" to the Global South [13]. We advocate the need for developing a new decolonizing pedagogy based on Freire's idea of "praxis" – iterative process that is based on critical reflection as well as a conscious awareness of the "oppression" of imposed constructs [12]. Approaches to social media research [14, 15] and theory of design of technological objects [2, 16] provide a framework for our analysis.

We maintain that while employing a quantitative research instrument may be sufficient for gathering data in a WEIRD context [7, 8], but when applied to a GREAT context [9], as in the case of the Zimbabwe Bush pump [2] it may be creating changes in social relations networked with this technical design object.

While it may not be of direct concern to the platform designers to consider these implications, as researchers of DSM these unintended consequences of the DSM need to be studied.

We demonstrate this issue by using the GREAT context of India to understand that data (including "missing" data) yields insights that may be useful for users of digital platforms to make strategic decisions [9]. To better motivate the cultural context in which the design of the platform and the data collection/analysis are situated, we focus on two aspects of India that is relevant to this study. First, as described in the earlier section, there are *staggering inequalities* of income and opportunity in India which affects the population's access to basic goods like food, electricity, education and health care to the more aspirational goods such as the internet and smartphones. 1% of the population owns 73% of its wealth [1]. The richest man in Asia is an Indian and 6.7% of the population (about 88 million people) live in extreme poverty, on less than $2 a day.

Second, India is *culturally and socially very diverse*. Indian cuisines reflect this diversity with there being no pan-Indian cuisine. Indian cuisine can be broadly split into five categories – northern, southern, eastern, western, and northeastern, with the cuisine of each region reflecting its local produce, cultural diversity, and varied demographics. The restaurants in our dataset also reflect this diversity; they collectively serve 133 unique cuisines, out of which 35 are Indian and the remaining 98 are international cuisines. That India is a vegetarian country is a myth as only 20% of Indians are vegetarians. Owing to long coast lines and lots of perennial rivers, fish is eaten extensively in India and the consumption of meat too is quite prevalent. However, owing to the preferences of cultural elites, and dominant groups, people often tend to under-report meat eating, especially beef and over report vegetarian food.

The consumption of alcohol also is subject to similar cultural biases. Alcohol is a state subject as per the seventh schedule of the Indian Constitution, therefore laws governing the retail and sale of alcohol vary from state to state. There are four 'dry' states in the country – Bihar, Gujarat, Mizoram, and Nagaland, along with the Union Territory of Lakshwadeep where the retail and sale of alcohol is banned. Kerala and Tamil Nadu do not allow private retailing of alcohol, and alcohol is sold through government owned shops only. Each state (among 29 states and 7 Union Territories) has its own list of days when liquor sale is prohibited, though all states and union territories

announce prohibition days during elections. There is also a highway liquor ban in India that prohibits the sale of liquor within 500 meters of national and state highways, except within municipal boundaries. Despite regulations, consumption of alcohol in India increased by 55% between 1992 and 2012 [17].

This paper takes the case of user ratings on a food delivery platform for a data set spanning 37 Indian cities. We illustrate by analyzing quantitative data qualitatively [18] and demonstrate the complexity of the social cultural and political context, leading us to critically examine the fact that so many restaurants across different cities have not received any ratings and suggest that this might not be a simple case of consumer preference but also a limitation of the rating system to capture all possible voices. Consequently, given the central role of ratings in prior research, our finding implies that there is a need to hold space for the voices that do not speak especially in a growing, aspirational and transitional context like India.

In conclusion, we argue that for businesses that depend on user ratings to compete, understanding the context and related networks may help in providing better solutions rather than being blind to it. Research instruments need to be adaptable and fluid [2] to generate and institutionalize knowledge that ethically speaks across a population that is unequal and diverse.

2. Background literature and theory development

We commence our discussion with a fundamental question from post-colonial discourse - "Can the Subaltern speak?" In the lecture of the same name, the subaltern was defined as the differential between the total Indian population and those described as elite. The elite included both dominant foreign as well as dominant indigenous groups, and the task of (post-colonial) research was to "investigate, identify and measure the specific nature and degree of deviation" of (the voice) of the regional indigenous people from the ideal represented by the elite. What was true of GREAT economies before the IT revolution and the proliferation of the internet, still remains a relevant need as it learns through praxis, learning by doing.

This question is further investigated through a vast body of literature on decolonizing research which argues that "decolonization is not a metaphor" for all social injustices imposed upon indigenous people across the globe, but a very particular form of social injustice where the identity of the colonized people is not subsumed, but effaced by the constructs imposed

upon them, as in the case of the Native American students in the public school education system in the United States [10], through which Settler colonialists attempt to either destroy or assimilate the Natives in order to take over their land, which is the primary objective of colonization. Digital colonization, especially through exported WEIRD constructs [7, 8] follows a similar pattern of impositions.

Subaltern groups are represented as the "insider-within" the participatory research space, which is as much rooted in place as the social context and in being so, offers the subaltern a space for "healing, recovery and development" [19]. Social spaces cannot be defined through a Cartesian coordinate system but are produced by a triad comprising of spatial space, representational space and representations of space [19]. We present a case wherein analysis of the 'place' where the rating system is administered generates insights that are used to expand the research findings in a non-traditional way, which in turn re-informs the constructs that were used in designing it. This illustrates the triangular relationship between pedagogy, research and praxis-learning through reflective practice [12], and departs from the earlier settler colonialist methods of research and learning.

As research projects are "subverted" by funders and their interests [5], we contend that learnings from user generated data is potentially more representative if it is freed of its "settler colonialist" design through decolonization of its research methodology and critical research of the design as well. Such data and its analysis may be of practical value to its users, as well as to the platforms that support it.

Information Systems (IS) literature encourages researchers to "venture into industries not commonly studied" and posits that studying new (to IS) industries can reveal novel phenomena and lead the development of new theory [18, 20-24]. The industry that we scrutinize, along with an examination of the industry structure and profitability [25], is a digital platform for food delivery operational in urban India, and the complex social, cultural, political and economic context of food and alcohol retail in India that influences it.

As digital platforms such as food delivery rating systems cater to the urban population (the data set studied in this paper comprises of 37 urban agglomerations), the service they provide may be included within the gamut of Urban Practice, particularly Southern Urban Practice [13], which is also grappling with the problem that Urban Theory

borrowed from the global north (WEIRD contexts [7, 8]) are inadequate to explain or understand phenomenon observed in cities of the global south (GREAT contexts [9]). As the majority of the world's urban population lives in the cities of the global South, there's a call for "dislocating the center" of urban theory making to the Global South and a "recalibration of the geographies of authoritative knowledge" [9, 13]. Urban theory remains "unrooted" in context and "thus seems impossible to translate, apply or use to influence practice in particular places" [14]. A phenomenon like why so many food outlets registered on a platform in urban India have not been rated is one such case.

Wellbeing, one of the goals of development, may be a permanent state of economic prosperity or the immediate gratification of having a good meal home conveniently delivered at the shortest possible time that meets a consumer's budget and tastes. Implicit to this, is the ability to make free choices which is what digital platforms claim to maximize. In the context of the internet, freedom is not what we get as a result of making a decision but what makes our decisions possible [14]. Rating systems on digital food delivery platforms enable decisions and their design and functioning as a "networked object" has an impact on social relations and other networks as they "simultaneously embody and measure a set of relations between heterogenous elements" [16].

Technical objects have political strength. They may change social relations but after their causal links are stabilized, it appears that is how things always were. "Once technical objects are stabilized, they become instruments of knowledge" [16] and they must be analyzed critically for their impact on society at large beyond their immediate intended use.

Digital platforms bring together the services of many providers, entrepreneurs and innovators in a sector along with consumers. They create knowledge regarding their transactions through rating systems which collates information on the subjective preferences of consumers.

As in the case of the Zimbabwe Bush pump [2], when technical objects designed for a certain type of user are deployed in areas where the user is different, a "fluid object that tries to serve its users" – fluid – adaptive, flexible, responsive – might be more successful than an object that is "firm" or rigid. The bush pump was adaptive in a way that the rating system we are studying here is not. What are the implications of this firmness or rigidity? Is the absence of ratings an indication of its "firmness" or rigidity?

During the Covid-19 crisis, cities went under lockdown, photographs of empty city streets garnered interest on the net. There was a 'fearful' thrill in considering the surrealness of deserted streets in cities that we know are densely populated. When we extend this metaphor to the domain of food delivery in urban areas in India, (particularly the data set we studied), we are struck by a similar disquiet: so many restaurants across different cities have not received any ratings. What does this say about the design of the rating system as well as the users and user networks that they affect? This question becomes even more relevant considering that the urban dining industry has been severely impacted by the Covid-19 crisis and given the social distancing norms being put into place, digital knowledge systems such as rating platforms may play a crucial role in its recovery and re-formation. According to industry experts, a large proportion of restaurants, even close to 40% of restaurants, are likely to shut down permanently.

A platform by definition implies a raised podium designed to facilitate certain activity [26]. From a political and architectural sense, a platform implies a raised structure from where a politician may address an audience or a passenger may stand to board a train. A popular social media platform like YouTube by describing itself as a platform implies that it is egalitarian and will support all users equally but this is in direct contradiction of the fact that it is supported almost entirely by advertising [15].

"Critical design is a research through design methodology that foregrounds ethics of design practice, reveals potentially hidden agendas and values, and explores alternative design values" [6]. DSM Platforms 'institute a way of being' [15] by producing knowledge and critical research must be undertaken to understand what this means for society.

3. Methods

In alignment with recent recommendations [27, 28], we leverage a recent multi-method approach that integrates induction and abduction for theory development [28-31]. Induction is a methodology for discovering patterns from big data [9, 30]. After discovering patterns in data [32], we make sense of the induced patterns by abductively developing the most generalizable explanations. In summary, abduction is an approach to theory building that completes the knowledge production process by making sense of the data-driven patterns discovered by induction [9, 31]. This method of inductive data-driven analytics

followed by abductive discovery has been demonstrated to yield novel theoretical insights from large datasets [28].

3.1. Research context

India is an apt setting for examining our research question. A largely agrarian society, India has been one of the fastest growing major economies in the world [33, 34]. This has led to the rise of a large group of consumers who aspire to western products and services [35-37], resulting in a transitional economy [9]. Hence, it has been the subject of increasing research inquiry (e.g., [9, 28, 33, 35, 38-41]). We chose to collect data from the largest food delivery platform in India. We do not explicitly name the platform to protect the confidentiality of the platform.

The motivation for this choice is given below. In recent years, India has been attracting a lot of investment in platforms across various industries [28]. A lot of platform companies are trying to enter and to dominate the Indian economy which has a massive population exceeding a billion people. While many food delivery platforms in India have received funding, given the diversities and pluralities of India, many platforms have not been able to succeed in India. In fact, after consolidation of the industry, by the end of 2017, five platforms left in operation. We chose the largest platform for our analysis. Three reasons motivated this choice. First, this platform is a comprehensive review and rating site that provides food discovery and delivery services. All restaurants are listed on the platform, irrespective of whether they participate in the delivery service. Second, this platform consistently does not levy fees from customers and thus does not cross-subsidize restaurant participation. In doing so, the restaurants' platform participation choices are not influenced by the dynamics of underlying fee/payment structure. Third, this platform has a pan-India presence and has been in operation for more than 2 years in all large cities in India. We chose to collect data from this platform given that it contains the most comprehensive collection of Indian restaurants.

We started with a population sample of 95,735 restaurants, serving a total of 135 different cuisines, located in 37 cities of India as our dataset. Restaurants across India are part of the sample if they are listed on the digital platform. Any consumer can list a restaurant on the website; listed restaurants can garner reviews and ratings from other consumers.

Ratings for about a third of the restaurants are missing; investigating the voice of the missing, restaurants with missing ratings is the focus of our analysis. Since induction yields easy-to-interpret decision rules [32, 42, 43] organized in a tree, this user-friendly methodology is often preferred by top management executives. Induction opens up the black box of decision making and represents emergent interrelationships between decision attributes and outcomes (information attributes are inputs to induction; outputs of induction, namely the attributes included in the tree, are referred to as decision attributes) [28].

Data partitioning creates non-overlapping training and validation partitions necessary for ascertaining the generalizability of knowledge. Knowledge is discovered from the training partition and validated using unseen data from the validation partition. In this study, we use 10-fold validation via data partitioning for avoiding the overfitting trap. In summary, we assess generalizability of the knowledge discovered on training data by testing its prediction accuracy on unseen data from the validation data partition. Following up induction with abduction is vital for theory development as it enables us to develop generalizable explanations for making sense of the data-driven patterns induced from big data.

3.2. Learning from data and missing data

For some restaurants, ratings on the digital platform are missing. Next, we describe information attributes included in our theory. The first key attribute we included was the price range. The cost of a meal for two persons reflects the strategic positioning of the restaurant (cost leadership [28]). Specifically, a restaurant that offers a meal for two persons for 1000 INR and above was assigned a value of high price range (approximately 14 US Dollars). The cost for a restaurant that offers a meal less than or equal to 300 INR (4 US Dollars) assigned a value of low. Restaurants where a cost for a meal was between 300 and 1000 INR were in the medium price range.

Cuisine Variety was assigned a value of low if the restaurant offered a single cuisine, medium if two or three cuisines were offered. A value of high was assigned to this variable if the restaurant offered more than three cuisines. If the restaurant is a vegetarian only restaurant or not is captured by using a dummy called Vegetarian. A value of Yes was assigned to this attribute if the restaurant was a vegetarian only restaurant, otherwise a value of No assigned to this attribute. Similarly, if the restaurant provides only

Indian food (versus world cuisines) is captured using a dummy called Only Indian. A value of Yes was assigned to this attribute if the restaurant serves only Indian food, otherwise a value of No assigned to this attribute. If the restaurant serves alcohol is captured using a dummy called Alcohol. A value of Yes was assigned to this attribute if the restaurant serves alcohol, otherwise a value of No was assigned.

A key institutional attribute that we captured corresponds to whether a restaurant is part of a group of restaurants; a restaurant chain reflected by the same or similar names. These restaurants may be part of a chain or might share a common name that reflects a well-established "institutional" identity (e.g., [44]). Thus, we capture this attribute by assigning Institutional Chain a value of high if nine or more other restaurants had the same name as the focal restaurant. A value of medium is assigned at least one other restaurant, and less than nine other restaurants, shared their names with the focal restaurant. If the restaurant's name was unique, low value is assigned.

From the point of view of the customer, restaurant variety (calculated for each focal restaurant) is captured by density of restaurants relative to a focal restaurant. Restaurant variety which represents spatial concentration of competition was calculated for each focal restaurant as the number of restaurants that lie within 1-kilometre distance of that focal restaurant. Restaurant variety variable was assigned three values. We assigned a value of low if the number of restaurants was less than 9, medium if the number of restaurants was between 10 and 99 (both inclusive) and high if the number of restaurants that lie within 1-kilometre distance of that focal restaurant was greater than or equal to 100.

Customers can either choose (home delivery) convenience by having the platform deliver food to their homes or choose the (restaurant dining) experience. These two scenarios of interest; (home delivery) convenience vs. (restaurant dining) experience were captured based on the restaurant's participation on the food delivery component of the online platform. Restaurants that did not participate on the delivery platforms were grouped together to study the (restaurant dining) experience scenario. Restaurants that participated on the delivery platforms were grouped together to study the (home delivery) convenience scenario. Finally, the focal variable of our analysis, if a restaurants online rating is available or missing was captured. 35,815 restaurants did not have ratings and we examine which factors guide or potentially explain when the restaurant's ratings are missing. The missing data across our entire pan-India dataset is presented in Table 1.

India is home to seven big metros including the old four (Mumbai, Kolkata, Chennai and Delhi) and the newer three metros (Bangalore, Pune and Hyderabad). Outside these top seven metro, in the thirteen tier two cities identified above, we find that ratings are available for fewer restaurants. This leads us to believe that more can be learned from the missing data. The thirteen cities where more data on ratings is missing (than available) could point to one of the following three explanations. The digital platforms have not penetrated the tier two cities. Or offline word of mouth mechanisms are much stronger in such cities as compared with online ratings. Alternatively, English could be a barrier whereby expressing ratings in English is not routinized consumer behavior.

Table 1. Pan-India missing ratings by city

City	Total Restaurants	Ratings Available	Ratings Missing	More Missing
Agra	507	214	293	Yes
Ahmedabad	3206	2084	1122	
Allahabad	345	94	251	Yes
Amritsar	387	152	235	Yes
Aurangabad	374	156	218	Yes
Bangalore	10580	7210	3370	
Bhopal	613	304	309	Yes
Bhubaneswar	609	440	169	
Chandigarh	1969	1188	781	
Chennai	5859	3904	1955	
Coimbatore	1561	466	1095	Yes
Dehradun	585	278	307	Yes
Delhi NCR	19068	11657	7411	
Goa	2391	1234	1157	
Guwahati	781	680	101	
Hyderabad	5839	3616	2223	
Indore	1071	701	370	
Jaipur	2048	1408	1000	
Kanpur	529	214	315	Yes
Kochi	1519	783	736	
Kolkata	4918	3793	1125	
Lucknow	1386	765	621	
Ludhiana	580	314	266	
Mangalore	374	243	131	
Mumbai	14487	10323	4164	
Mysore	445	270	175	
Nagpur	1218	468	750	Yes
Nasik	330	205	125	
Patna	304	159	145	
Puducherry	497	252	255	Yes
Pune	7664	4776	2888	
Ranchi	262	155	107	
Surat	611	325	286	
Udaipur	538	263	275	Yes
Vadodara	810	376	434	Yes
Varanasi	292	147	145	
Vishakhapatnam	818	314	504	Yes

3.3. Alcohol consumption in restaurants and missing ratings

Serving alcohol has strong implications for governance as the sale of alcohol generates significant

cash revenues. In India, sale and consumption of alcohol supports a parallel cash economy. Restaurants that serve alcohol do not participate on the platform. Systematic analysis of the participation question reveals that perhaps this strategic omission is to avoid scrutiny of the taxman [17, 28]. Restaurants that serve alcohol also suffer from low ratings [9]. Thus, restaurants that serve alcohol deserve more attention in our examination of the missing. Restaurants serving alcohol are fundamentally different from restaurants not serving alcohol. Thus, we decided to further scrutinize the patterns of missing ratings across restaurants that do serve alcohol as compared with restaurants that do not serve alcohol. Model free findings are reported in Table 2.

Table 2. Alcohol consumption and the missing

City	Total	Ratings Available		Ratings Missing	
		No Alcohol	Serves Alcohol	No Alcohol	Serves Alcohol
Allahabad	345	89	5	241	10
Bangalore	10580	6322	888	3278	92
Chennai	5859	3682	222	1925	30
Delhi NCR	19068	10382	1275	7288	123
Goa	2391	353	881	399	758
Hyderabad	5839	3166	450	2132	91
Kolkata	4918	3371	422	1101	24
Mumbai	14487	8283	2040	3835	329
Pune	7664	3996	780	2709	179
Ranchi	262	137	18	98	9
Vishakhapatnam	818	286	28	481	23
Ahmedabad	*3206*	*2084*		*1122*	
Vadodara	*810*	*375*	*1*	*430*	*0*
Surat	*611*	*325*	*0*	*285*	*1*

Some key findings emerge from Table 2. Ratings are more readily available in the larger metros on India. Data from Mumbai reveals that ratings are available for about 70% of the restaurants. In most tier two cities like Allahabad, ratings are available for only about 27% of restaurants. The patterns in data for most tier two cities are like that of the Allahabad data represented here. When restaurants serve alcohol, in most cases, ratings are available for fewer such restaurants. A notable exception to these findings is the state of Goa. Known for its beautiful beaches, nightlife and hospitality, Goa is a popular spot with tourists. A long Portuguese history makes Goa a culturally rich and laid-back state known for merry making. People go to Goa to enjoy a drink or two; this also seems to be reflected in the ratings. More restaurants that serve alcohol are rated in Goa that in any other city. In sharp contrast, the state of Gujarat (where the cities of Ahmedabad, Vadodara and Surat are located) is a "dry" state. The sale of alcohol in Gujarat is officially not allowed.

3.4. Induction: Discovering common traits of restaurants with missing ratings

Two steps define knowledge discovery via induction. First, the C4.5 algorithm is used to induce a decision tree on training data [45]. Second, the tree grown in step 1 is pruned by validating it with unseen data from the validation partition. By employing high levels of pruning, we discover the tacit structure of data and demonstrate robustness of discovered knowledge. The Weka software application, an open-source platform is used for data partitioning, and for growing and pruning trees [42]. The C4.5 algorithm relies on the concept of purity and utilizes informative attributes to recursively partition the training data to reduce impurity in terminal nodes. Entropy is chosen as the impurity measure, as entropy is easy to interpret for a two-class decision problem [43, 45, 46].

Tree induction iteratively groups together observations (i.e., restaurants) such that they are similar not only in certain information attributes but also similar in terms of their (missing or available) ratings. There are two inputs to tree induction: (1) restaurants described by all information attributes (as described in Table 1), and (2) restaurant ratings. We investigate commonalities in terms of restaurant attributes collectively associated with missing ratings.

The objective of tree induction is to discover tacit combinations of information attributes associated with similar final outcomes (i.e., similar ratings) [42, 43]. Trees only retain the most pertinent attributes for explaining decisions and organize decision attributes in a context-dependent manner; certain questions are only raised depending on answers obtained to other questions [42, 43]. Trees discovered by induction are not reflective of the exact rules or "scripts" used by the decision makers, but rather represent credible approximations of customer journeys. To ensure that decision rationale is comprehensively discovered, a process of drawing mutually exclusive, training and testing subsamples is repeated multiple times. In each iteration, we draw random, mutually exclusive subsamples of restaurants from the original data; one set, known as the training set, from which the tacit decision rationale is discovered by the C4.5 induction algorithm [43], and another disjoint set of initiatives, known as the testing set, which is used to test the predictive accuracy of this discovered rationale. We used 10-fold validation where the full sample is divided into 10 partitions of which 9 partitions are used for building the tree and the last partition is used for validation. Accuracy of the tree discovered from

training set is assessed by predicting decisions for restaurants from unseen data from the validation set.

Multiple approximations of the rationale are derived by conducting computational experiments whereby the 10-fold validation process is repeated at varying levels of pruning. Using prediction accuracy of the decision tree as the sole criterion when choosing the best representative tree can be misleading and would be akin to falling into the overfitting trap. We rely on three heuristics [9, 28], namely (i) high prediction accuracy, (ii) high parsimony and (iii) high reliability to select the best representative trees across the two scenarios of (home delivery) convenience vs (restaurant dining) experience. Thus, we high confidence that trees presented here are the "best", most credible approximations of the customer journeys for rating restaurants on the platform.

All seven information attributes characterizing restaurants across the two scenarios of (home delivery) convenience versus (restaurant dining) experience in conjunction with the final restaurant rating are inputs to induction. All information attributes deemed informative for explaining ratings are included in the trees as decision attributes and the algorithm excludes all the non-informative attributes from the tree. The most informative decision attribute is the top-most attribute in the tree. Importance of attributes decreases as we move away from the top of the tree to its leaves.

3.5. Induction Results: Common traits of restaurants with missing ratings

The results of decision tree induction, reported in Figure 1, validate our model free analysis. Ratings are often missing for low-cost restaurants serving meat. This finding shines the light on aspects of economic inequality and the imperative of subsistence over the need to voice opinions on DSM.

4. Discussion and concluding comments

4.1. Implications of findings from induction

Research approaches on DSM and online platforms that originated in the WEIRD contexts need to be reexamined before they are imposed on to the GREAT contexts. We examined the ratings system for a food delivery platform in India, yielding three main findings. First, for thirteen cities, we found that ratings were missing for more restaurants (when compared to available restaurant ratings). Arguably the voice is mostly missing in these thirteen cities of India as the

rating system imposed onto this GREAT domain is not designed to "hear" or capture the variety and complexity of food delivery market in those cities. Second, in a large majority of cases, many restaurants that serve alcohol do not receive online ratings. The sale of alcohol in India is arguably good for business; however, it is not good for online ratings. Third, online ratings are also mostly missing for cheap meat serving eateries across India. A large proportion of Indians eat at these low-cost meat serving restaurants. The offline word of mouth is arguably so strong that relying on the digital platform to voice out opinions and ratings about the food is not routine behavior.

```
Decision Tree

Price Range = Low
|  Cuisine Variety = No
|  |  Vegetarian Only = No: Ratings Missing (10948 restaurants)
|  |  Vegetarian Only = Yes
|  |  |  Institutional Chain = No: Ratings Missing (4970 restaurants)
|  |  |  Institutional Chain = Small Chain
|  |  |  |  Restaurant Variety = Low: Ratings Missing
|  |  |  |  Restaurant Variety = Medium: Ratings Available
|  |  |  |  Restaurant Variety = High: Ratings Available
|  |  |  Institutional Chain = Large Chain: Ratings Available
|  Cuisine Variety = Medium
|  |  Institutional Chain = No
|  |  |  Vegetarian Only = No: Ratings Missing (6277 restaurants)
|  |  |  Vegetarian Only = Yes
|  |  |  |  Restaurant Variety = Low: Ratings Missing
|  |  |  |  Restaurant Variety = Medium: Ratings Available
|  |  |  |  Restaurant Variety = High: Ratings Available
|  |  Institutional Chain = Small Chain: Ratings Available
|  |  Institutional Chain = Large Chain: Ratings Available
|  Cuisine Variety = High: Ratings Available
Price Range = Medium
|  Only Indian = No: Ratings Available (40199 restaurants)
|  Only Indian = Yes
|  |  Platform Delivery = No
|  |  |  Vegetarian Only = No: Ratings Missing
|  |  |  Vegetarian Only = Yes: Ratings Available
|  |  Platform Delivery = Yes: Ratings Available
Price Range = High: Ratings Available (9005 restaurants)
```

Figure 1: Missing restaurant ratings

4.2. Abduction and conclusion

Abducting away from these findings induced from data, we present two key implications. First, research is mostly written from a WEIRD perspective; asks WEIRD questions and necessarily provided WEIRD answers [7, 8]. The rest of the world is not WEIRD. In adopting a mostly WEIRD stance, research has fallen into the "deep" trap of blindness and gullibility. A Type I error in statistics refers to the error of blindness where a true relationship between X and Y is not "seen" / detected by "blind" researchers. A type II error is where "gullible" researchers incorrectly "see" a relationship between X and Y when in fact such a relationship does not exist. Errors in statistical analyses are relatively easy to fix. The deep

trap is that researchers are making errors of blindness and gullibility as they choose their research questions.

We maintain that WEIRD research and WEIRD questions [7, 8] cannot be imposed upon GREAT contexts; a deeper thinking and discourse is necessary. The decolonization of research methodologies and tools (rating systems in this case) is often essential and researchers need to carefully think about their research choices and not impose WEIRD research tools onto GREAT domains. Our research raises awareness on key ethical issues of doing DSM research and encourages researchers not to be blind to the obvious domain-specific nuances and questions staring them in the eye. We encourage research to not be gullible and ask the same WEIRD questions all over again in GREAT contexts. Hopefully, mindful researchers will not be blind or gullible when choosing GREAT research questions.

Second, elaborating on the decolonization of WEIRD research, liberation is not a "gift" to be given by the "oppressor" to the "oppressed" but a "mutual process" where practitioners and researchers iteratively adapt the research method to a GREAT domain. Resonating with Friere's [12] idea of praxis - action which relies on critical self-reflection for informing itself and creating its own learning pathways, we encourage researchers to challenge power structures within existing systems of knowledge creation and reinvent the wheel where necessary to ask GREAT questions and reinvent the tools to suitably yield GREAT answers. An important part of this process is also to simultaneously raise awareness about the ethical issues of conducting research on digital and social media [4]. In particular, given the prevalence of rating systems, we encourage research on new forms of rating systems for GREAT contexts such that restaurant owners, and other participants of other digital platforms and DSM [4], can receive tangible benefits from user generated data.

In conclusion, we identify opportunities for *design reform* whereby online rating systems can be redesigned such that they are more contextually sensitive and "fluid" to hear voices that were being excluded before. Such design reform is vital for exploiting the full potential of online platforms in GREAT domains. In summary, such a process of decolonization has implied learnings for Big Data Research also, which deals with even more unstructured, disjointed and random datasets where the problem of the missing may be further exacerbated by the presence of noise [15].

5. References

[1] H. A. H. Dang and P. Lanjouw, "Inequality in India on the Rise," in "WIDER Policy Brief 2018/6," UNU WIDER Policy Brief, Helsinki, 2018.

[2] M. de Laet and A. Mol, "The Zimbabwe Bush Pump: Mechanics of a Fluid Technology," *Social Studies of Science,* vol. 30, no. 2, pp. 225-263, 2000.

[3] M. C. Nussbaum, "Creating Capabilities: The Human Development Approach and Its Implementation," *Hypatia,* vol. 24, no. 3, pp. 211-215, 2009.

[4] T. Sutherland and J. Pierre, "Introduction to the Minitrack on Critical and Ethical Studies of Digital and Social Media," in *Hawaii International Conference on System Sciences*, Hawaii, 2019.

[5] L. Irani, *Chasing Innovation: Making Entrepreneurial Citizens in Modern India*. Princeton and Oxford: Princeton University Press, 2019.

[6] J. Bardzell and S. Bardzell, "What is "critical" about Critical Design?," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Paris, France, 2013, pp. 3297-3306.

[7] J. Henrich, S. J. Heine, and A. Norenzayan, "The Weirdest People in the World?," *Behavioral and Brain Sciences,* vol. 33, no. 2-3, pp. 61-83, 2010.

[8] J. Henrich, S. J. Heine, and A. Norenzayan, "Most People are not WEIRD," *Nature,* vol. 466, no. 7302, p. 29, 2010.

[9] P. P. Karhade and A. Kathuria, "Missing Impact of Ratings on Platform Participation in India: A Call for Research in G. R. E. A. T. Domains," *Communications of the Association for Information Systems,* vol. 47, 2020.

[10] M. Zavala, "What do we mean by decolonizing research strategies? Lessons from decolonizing, Indigenous research projects in New Zealand and Latin America," *Decolonization: Indigeneity, Education & Society* vol. 2, no. 1, pp. 55-71, 2013.

[11] G. C. Spivak, "Can the Subaltern Speak?," in *Can the Subaltern Speak?: Reflections on the History of an Idea*, R. Morris Ed. New York: Columbia University Press, 1983, pp. 237-292.

[12] P. Freire, *Pedagogy of the Oppressed*. Bloomsbury Publishing USA, 1968.

[13] G. Bhan, "Notes on a Southern Urban Practice," *Environment and Urbanization,* vol. 31, no. 2, pp. 639-654, 2019.

[14] W. H. K. Chun, *Control and Freedom: Power and Paranoia in the Age of Fiber Optics*. Boston: MIT Press, 2008.

[15] T. Gillespie, "The Politics of 'Platforms'," *New Media & Society,* vol. 12, no. 3, pp. 347-364, 2010.

[16] M. Akrich, "The De-scription of Technical Objects," in *Shaping Technology / Building Society: Studies in Socio-Technical Change*, W. E. Bijker and J. La Eds. Boston: MIT Press, 1992, pp. 205-224.

[17] OECD. Alcohol consumption [Online] Available: https://www.oecd-ilibrary.org/content/data/e6895909-en

[18] M. G. Pratt, "From the editors: For the lack of a boilerplate: Tips on writing up (and reviewing)

qualitative research," ed: American Society of Nephrology Briarcliff Manor, NY, 2009.

[19] H. Lefebvre and D. Nicholson-Smith, *The Production of Space*. Oxford Blackwell, 1991.

[20] P. Karhade and J. Dong, "Innovation Outcomes of Digitally Enabled Collaborative Problemistic Search Capability," *MIS Quarterly,* vol. forthcoming, 2020.

[21] P. Karhade and J. Dong, "Information Technology Investment and Commercialized Innovation Performance: Dynamic Adjustment Costs and Curvilinear Impacts," *MIS Quarterly,* vol. forthcoming, 2020.

[22] J. Q. Dong, P. Karhade, A. Rai, and X. S. Xu, "How Firms Make Information Technology Investment Decisions: Toward a Behavioral Agency Theory," *Journal of Management Information Systems,* Forthcoming.

[23] J. Q. Dong, J. He, and P. Karhade, "The Penrose effect in resource investment for innovation: Evidence from information technology and human capital," *European Conference on Information Systems*, 2013.

[24] J. Q. Dong, P. Karhade, A. Rai, and S. X. Xu, "Information Technology in Innovation Activity of the Firm: Theory and Synthesis," *European Conference on Information Systems*, 2013.

[25] M. W. Chiasson and E. Davidson, "Taking industry seriously in information systems research," *MIS Quarterly,* pp. 591-605, 2005.

[26] S. Vitzthum, A. Kathuria, and B. Konsynski, "Toys become Tools: From Virtual Worlds to Real Commerce," *Communications of the Association for Information Systems,* vol. 29, no. 1, p. 21, 2011.

[27] V. Grover, A. Lindberg, I. Benbasat, and K. Lyytinen, "The Perils and Promises of Big Data Research in Information Systems," *Journal of the Association for Information Systems,* vol. 21, no. 2, p. 9, 2020.

[28] A. Kathuria, P. P. Karhade, and B. R. Konsynski, "In the Realm of Hungry Ghosts: Multi-Level Theory for Supplier Participation on Digital Platforms," *Journal of Management Information Systems,* vol. 37, no. 2, pp. 396-430, 2020.

[29] C. S. Peirce, *Collected papers of Charles Sanders peirce*. Harvard University Press, 1974.

[30] P. Karhade, M. J. Shaw, and R. Subramanyam, "Patterns in Information Systems Portfolio Prioritization: Evidence from Decision Tree Induction," *MIS Quarterly,* vol. 39, no. 2, 2015.

[31] P. P. Karhade, M. J. Shaw, and R. Subramanyam, "Evolution of Decision Rules used for IT Portfolio Management: An Inductive Approach," in *Americas Conference on Information Systems*, 2009, pp. 307-320.

[32] P. Karhade and M. Shaw, "Rejection and Selection Decisions in the IT Portfolio Composition Process: An Enterprise Risk Management Based Perspective," *Americas Conference on Information Systems,* p. 221, 2007.

[33] R. Kathuria, N. N. Kathuria, and A. Kathuria, "Mutually supportive or trade-offs: An analysis of competitive priorities in the emerging economy of India," *The Journal of High Technology Management Research,* vol. 29, no. 2, pp. 227-236, 2018.

[34] R. Kathuria, S. J. Porth, N. Kathuria, and T. Kohli, "Competitive priorities and strategic consensus in emerg-ing economies: evidence from India," *International Journal of Operations and Production Management,* vol. 30, no. 8, pp. 879-896, 2010.

[35] J. Khuntia, A. Kathuria, T. Saldanha, and B. Konsynski, "Benefits of IT-enabled Flexibilities for Foreign versus Local Firms in Emerging Economies," *Journal of Management Information Systems,* vol. 36, no. 3, pp. 1-38, 2019.

[36] J. Khuntia, A. Kathuria, M. Andrade Rojas, T. Saldanha, and N. Celly, "How Foreign and Domestic Firms Differ in Leveraging IT-enabled Supply Chain Information Integration in BOP Markets: The Role of Supplier and Client Business Collaboration," *Journal of the Association for Information Systems,* forthcoming.

[37] J. Khuntia, T. Saldanha, and A. Kathuria, "Dancing in the Tigers' Den: MNCs versus Local Firms Leveraging IT-enabled Strategic Flexibility," *International Conference on Information Systems*, 2014.

[38] N. Celly, A. Kathuria, and V. Subramanian, "Overview of Indian multinationals," *Emerging Indian Multinationals: Strategic Players in a Multipolar World,* 2016.

[39] A. Kathuria, A. Mann, J. Khuntia, T. Saldanha, and R. J. Kauffman, "A Strategic Value Appropriation Path for Cloud Computing," *Journal of Management Information Systems,* vol. 35, no. 3, pp. 740-775, 2018.

[40] V. Venkatesh, A. Rai, T. A. Sykes, and R. Aljafari, "Combating Infant Mortality in Rural India: Evidence from a Field Study of eHealth Kiosk Implementations," *MIS Quarterly,* vol. 40, no. 2, pp. 353-380, 2016.

[41] V. Venkatesh, J. D. Shaw, T. A. Sykes, S. F. Wamba, and M. Macharia, "Networks, Technology, and Entrepreneurship: A Field Quasi-experiment among Women in Rural India," *Academy of Management Journal,* vol. 60, no. 5, pp. 1709-1740, 2017, doi: 10.5465/amj.2015.0849.

[42] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," *ACM SIGKDD Explorations Newsletter,* vol. 11, no. 1, pp. 10-18, 2009.

[43] J. R. Quinlan, "Induction of decision trees," *Machine learning,* vol. 1, no. 1, pp. 81-106, 1986.

[44] M. T. Hannan and J. Freeman, "The population ecology of organizations," *American Journal of Sociology,* vol. 82, no. 5, pp. 929-964, 1977.

[45] R. J. Quinlan, "C4. 5: Programs for Machine Learning," 1993.

[46] J. R. Quinlan, "Decision trees and decision-making," *IEEE Transactions on Systems, Man, and Cybernetics,* vol. 20, no. 2, pp. 339-346, 1990.