The Measurement of Social Difference

Rohini Somanathan

October, 2019

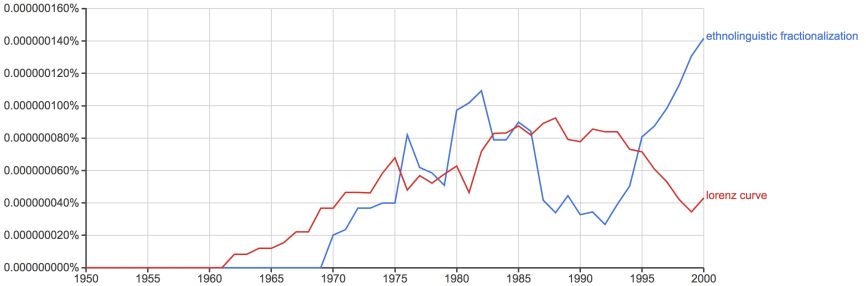
Our changing vocabulary

Ethnicity
Diversity
Multiculturalism
Minorities
Representation
Culture

Dissimilarity
Conflict
Segregation
Polarization
Discrimination
Fractionalization

.

Increased focus on groups and identity



Source: Ngram frequency for google book corpus

Why groups?

To ensure equal treatment

criteria: groups historically disadvantaged through slavery, ostracism, geographic or linguistic isolation

To understand social cooperation and conflict

criteria: strong identity within groups, social distance, alienation or rivalry between them

Common premises

Identity is

inherited, not chosen and primarily ethnic (broadly defined)

heterogeneity

is a disadvantage to be managed, groups do not share common goals

social data

is comparable over time and space

relative group shares

measure the potential for cooperation and conflict

Three claims

social data

is rarely comparable, it is worth studying the political and ideological influences that shape it

Identity is

fluid, responds to incentives in group-based policies

the study of group inequality and collective action

can benefit from combining class with identity

Outline

Categories of difference

A brief history of social data collection in several countries, a case-study of India

Measures of difference

Fractionalization, Polarization, Dis-similarity

Evidence on shifting identity

Identity is fluid and subjective, it responds to the social environment, questionnaire format, incentives created by group-based policies

Combining class and groups

Results from past work: Inequality can affect segregation, group-based policies can be meritocratic

Social data: cross national variation

The United Nations Statistical Division has archives census forms and data

- 138 countries conducted a census around the year 2000
- 87 collect some type of ethnic information
  - 20 per cent (mainly Europe) collect nationality
  - 15 per cent (mainly Latin America) collect indigenous origin
  - Former slave-holding societies use race
- Many others (outside the 87) use religion, language and legal citizenship as markers group identity
- India: caste and tribe for disadvantaged groups, language and religion for all

As we will see, differences reflect a combination of demographics, politics and ideology

Social data..The United States

Started census operations in 1790 to determine representation and taxation

1790: Counted whites, other free persons, and slaves- not native americans

1850: Color appeared - blacks whites, mulattoes

1870: Race appeared as an explicit question with many terms for mixtures:

Instructions to enumerators: *Be particularly careful in reporting the class Mulatto. The word here is generic, and includes quadroons, octoroons, and all persons having any perceptible trace of African blood. Important scientific results depend upon the correct determination of this class.*

Early 20th C: Chinese, Japanese, Korean, Mexican, Hindu, Cuban, Vietnamese and Asian Indian were added under Race

1970: Self-reporting of race and ethncity (separate questions)

1997: Federal directive to increase number of race categories and allowed multiple responses to more accurately report minorities.

Tiger Woods on the Oprah Show in 1997 talked of being bothered when called African-American:

Growing up, I came up with this name. I'm a Cablinasian. (Caucasian, black, Indian and Asian)

Father half-black, one-quarter American Indian, and one-quarter Chinese and mother half-Thai, one-quarter Chinese and one-quarter white.

Social data..Canada

Also started census operations in the 18thC, similar history of immigration, but very different approach to social data

pre-1951: the census asked about race

1951 ethnic origin replaced the race question and included options for race, religion and country of origin.

canadian was a possible response but discouraged and not tabulated

1986: counts for canadian became available, 0.5% reported it

1996: one-third reported canadian as main ethnicity

experimented with ordering ethnicity options, with french appearing before english in some years and not others

2011: the long form of the census with ethnicity questions was made voluntary- selection biases are now inevitable in ethnic datasets.

Social data..Europe

Post WW2: Antisemitism and conflict led most countries to prohibit the collection of ethnic data

The European parliament has suggested that the lack of ethnic data may hinder anti-discrimination policies by making ethnic gaps less visible.

Simon (2012) studies social data collection in 41 countries and finds:

- 22 collect data on ethnicity or nationality
- 23 collect data on religion
- 26 record mother-tongue

Social data.other examples

Latin America:

under colonial rule, ethnic counts used for conscription to forced labor and taxation
when census operations began, four categories: of white, black, indian and mixed (sometimes indigenous listed first)
After 1950, many replaced race with language

Israel: records religion and birthplace but not ethnicity such as Arab

Rawanda: outlawed the use of ethnic labels such as Hutu and Tutsi after the 1994 genocide

Social data..British India

Violently caste contests today- debates originated with the colonial census

pre-1872: scattered counts, detailed reports for some provinces, limited aggregation

1881: castes above 100,000 enumerated, others arbitrarily included- alphabetically listed

1891: occupational classification adopted

1901: emergence of the ethnographic state and the use of caste to test race science

1931: more limited counts, return to emphasis on occupation, lists of exterior castes and primitive tribes

1941, 1951: very limited operations, the two above groups eligible for the affirmative action after the 1950 constitution.

Affirmative action categories after independence

Data collection driven by affirmative action legislation: 3 categories, SCs STs, OBCs

- 1950-51: two separate lists for scheduled castes and scheduled tribes
- 1955: first backward classes commission appointed to create a new list of other backward classes (OBCs) (2,399 groups, 50%)
- 1976: most territorial restrictions for scheduled castes and tribes removed within states
- 1978: second backward classes (mandal) commission, updated previous lists
- 1990: 27% quotas for OBCs in federal jobs
- 2008: 27% quotas for OBCs in university admissions

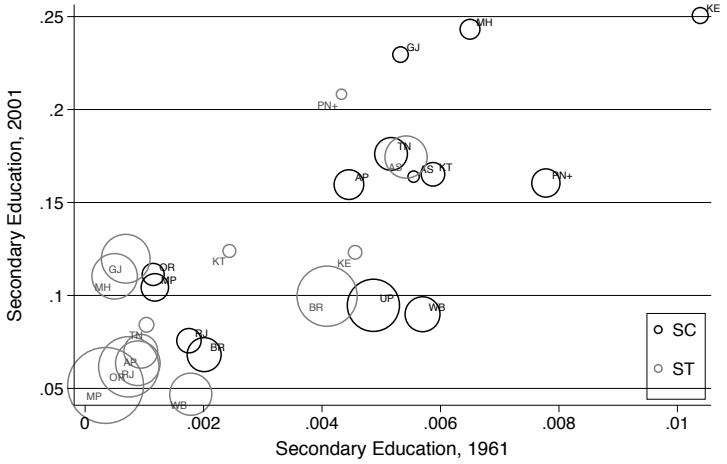
The demand for disadvantage

British India: The exterior castes protested inclusion.

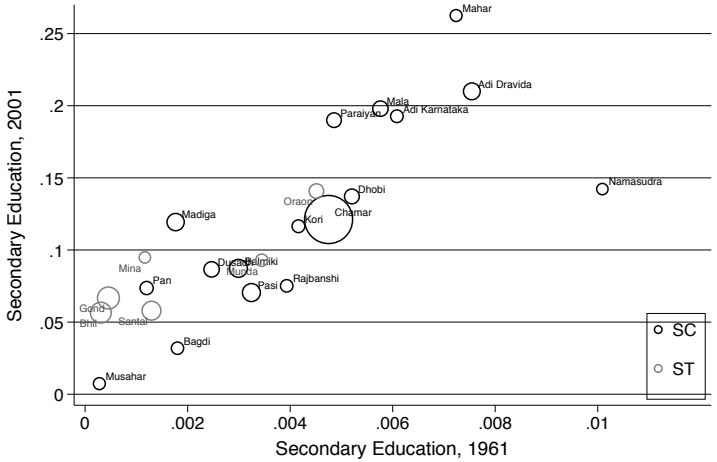
> *[In Bengal] ..the Suklis have definitely protested against their inclusion and the attitude of the Rajbansis has been equivocal (J. Hutton: Census of India, 1931.)*

Independent India:The excluded castes lobby for inclusion:

Divergence in secondary schooling, 1961-2001..by state



Divergence in secondary schooling, 1961-2001..by jati



Was the original classification justified?

What did the census commissioners think?

1885: Eustace Kitts, while creating a compendium of Indian castes and tribes in 1885 using census reports, described himself in "a mighty maze without a plan."

1891: Caste and Religion: "Thus we find that the various tribes of Jat and Rajput, in the north of India, contain nearly an equal number of Brahmanic and Musalman members, not to mention the Sikhs that prevail in certain localities."

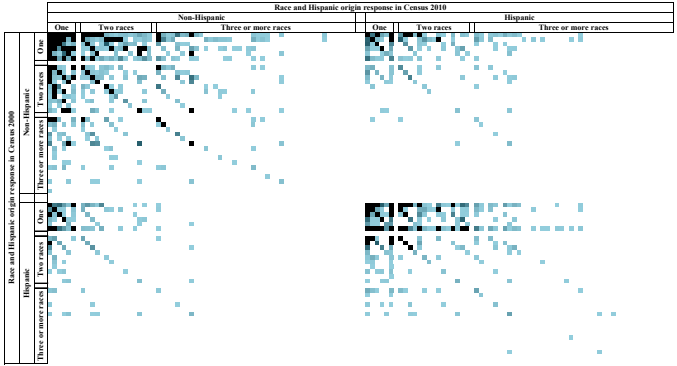
1901: Bengal report in response to Risley's demand to classify caste by status based on public opinion : It is very difficult to say precisely what constitutes Hindu public opinion. The Hindus as a body are strangely indifferent to the circumstances of castes that do not clash with their own. Those of a good position know very well from whom they can take water and those whose touch defiles, but they neither know nor care much regarding their relative position. The lower castes are even more ignorant of the caste of the higher ones.

1931: "All subsequent census officers in India must have cursed the day when it occurred to Sir Herbert Risely, no doubt in order to test his admirable theory of the relative nasal index, to attempt to draw up a list of castes according to their rank in society. He failed, but the results of his attempt are almost as troublesome as if he had succeeded, for every census gives rise to a pestiferous deluge of representations, accompanied by highly problematic histories, asking for recognition of some alleged fact or hypothesis of which the census as a department is not legally competent to judge and of which its recognition, if accorded, would be socially valueless."

On fluid identities..United States

6 million people change their race or hispanic origin responses between 2000 and 2010 (Liebler et al, 2014)

re-ordering of race and ethnicity improves ethnicity response (Anderson and Fienberg,2000)



On fluid identities..Brazil

Race questions asked in different formats (Bailey, 2008)

mixed races, when constrained opt for white

mentioning quotas for blacks doubles the fraction in that category

TABLE 2
PESB: DISTRIBUTION OF SELF-CLASSIFICATION ACROSS FORMATS

			Dichotomous	
Color/Race	Census (1)	Open-Ended (2)	Mulattos (3)	All (4)
White	49.1	42.7	44.1	66.7
Moreno		24.0		
Mulatto (*pardo*)	38.6	15.4		
Negro		6.8		
Black (*preto*)	12.4	2.9	55.9	33.3
Moreno claro		2.8		
Amarelo		1.7		
Mixed		.8		
Claro		.8		
Mulato		.6		
Indian		.4		
Mestiço		.2		
Other		.8		
N	2,225	2,306	830	2,203

NOTE.—Data are percentages except for *N*. All columns sum to approximately 100%.

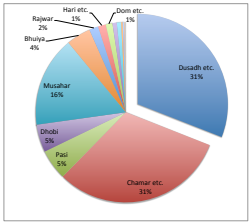
Identity manipulation..India

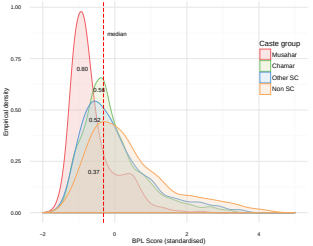
Political pressure to widen eligibility for affirmative action

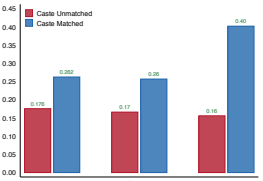
The mahadalits or ultra-disadvantaged category was gradually taken over by all

poverty rates did not warrant the expansion

sub-identity matters







Measures

Fractionalization

$$ELF = 1 - \sum_{i=1}^{n} s_i^2 \qquad \text{or equivalently as} \qquad \sum_{i=1}^{n} s_i(1 - s_i)$$

Polarization

$$P = \sum_{i=1}^{n} s_i^2(1 - s_i)$$

Dis-similarity

$$D = \frac{1}{2} \sum_{i=1}^{n} \left| \frac{B_i}{B} - \frac{W_i}{W} \right|.$$

The first two could potentially incorporate distance but it is hard to do and hasn't entered the literature

Incorporating inequality into group research..segregation

Inequality and Segregation, JPE 2004.

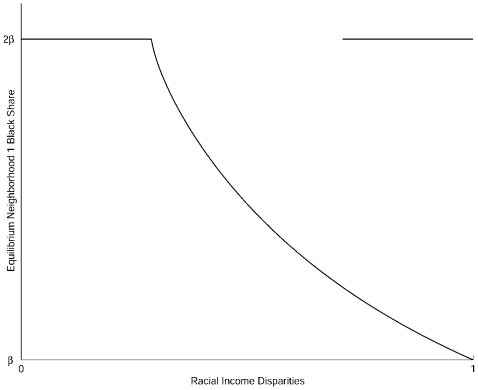
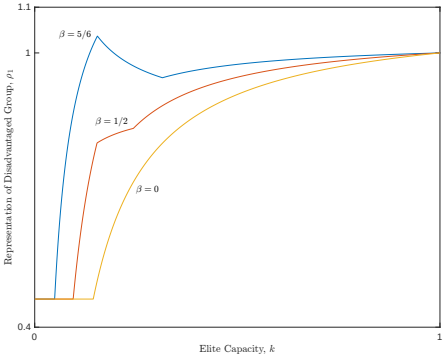


FIG. 4.—Racial income disparities and equilibrium segregation

Incorporating inequality into group research...representation

Meritocracy in the face of group inequality, 2019.

$$\text{performance} = \beta \text{ ability} + (1 - \beta) \text{ training}$$