Exploring cultural diversity of the people of India

N. V. Joshi*, Madhav Gadgil*† and Suresh Patil**

*Centre for Ecological Sciences, Indian Institute of Science, Bangalore 560 012, India
**Anthropological Survey of India, Southern Regional Office, 2963, Gokulam Road, Mysore 570 002, India
†Jawaharlal Nehru Centre for Advanced Scientific Research, Indian Institute of Science, Bangalore 560 012, India

The People of India project of the Anthropological Survey of India has assigned the entire Indian population to 2753 communities distributed over 32 states/union territories. This article explores: (a) number of communities in relation to population size, (b) number of and similarity in the traits exhibited by different sets of communities, (c) clustering of the traits in relation to geography, food habits, subsistence strategy and (d) impact of modernization on the division of labour within the Indian society.

THE Indian subcontinent, with its tremendous range of environmental regimes, supports a large human population that is a fascinating mosaic of varied cultural traits. This cultural diversity is organized in the form of tens of thousands of endogamous caste groups that function as largely closed breeding units. Furthermore, in rural India, social interactions and hence transmission of cultural traits overwhelmingly involve members of an endogamous caste group, with inter-group interactions primarily focusing on economic transactions. Members of an endogamous caste group therefore tend to share cultural traits, so that between-group variation in cultural traits is significantly larger than the within-group variation. India's cultural diversity is therefore best investigated in terms of such endogamous caste groups. Presently, however, this is an impossible task given the very large numbers of groups involved. These endogamous caste groups, however, form a smaller number of culturally homogeneous clusters. These correspond to the 'communities' investigated by the People of India project[1,2]. We then have at our disposal extensive information on 2753 such communities making up 4635 elements when a community population in each state/union territory is counted as a separate element. A yes/no response to 776 individual items of information ranging over identity, ecology, food habits, occupation, kinship patterns, marriage rules, art and music, as well as educational status and impact of development programmes is now available for these 4635 community elements. Undoubtedly this represents information of varying degrees of reliability and value from an ecological, anthropological and developmental perspective. Detailed investigation of this material must therefore be based on careful sifting. But before that begins it is worthwhile enquiring into the broad patterns emerging from this rich set of data; this is what we propose to undertake here.

How many communities?

A good way to begin is by asking how many different communities occur in a given geographical area. The POI data furnishes this number for the 25 states and 7 union territories with population ranging from 40,000 in Lakshadweep to over 110 million in Uttar Pradesh[3]. In addition, the total number of communities for India as a whole is 2753 (Table 1). The relationship between the number of communities and human population for the states and union territories is described well by the equation (see Figure 1):

Table 1. States with their populations and number of traits

State/UT	Population (millions)	No. of communities	Mean no. of traits	Mean dissimilarity	Total no. of traits	Abbrev.**
Andhra Pradesh	53.550	386	215.8	0.394	723	AP
Assam	19.897	115	220.9	0.393	684	ASM
Bihar	69.915	261	223.2	0.335	677	BHR
Gujarat	34.086	289	175.8	0.437	682	GUJ
Haryana	12.923	82	231.7	0.336	636	HAR
Himachal Pradesh	4.281	116	229.2	0.348	698	HP
Jammu and Kashmir	5.987	111	201.1	0.454	691	J&K
Karnataka	37.136	300	178.4	0.399	684	KAR
Kerala	25.454	225	195.2	0.363	676	KER
Madhya Pradesh	52.179	342	182.4	0.430	695	MP
Maharashtra	62.784	305	204.7	0.433	735	MHR
Manipur	1.421	29	225.0	0.382	564	MNP
Meghalaya	1.336	27	235.4	0.434	636	MGH
Nagaland	0.775	25	285.2	0.320	629	NGL
Orissa	26.370	279	224.3	0.360	708	ORS
Punjab	16.789	95	192.7	0.400	644	PUN
Rajasthan	34.262	228	171.6	0.417	649	RAJ
Sikkim	0.316	25	208.7	0.379	539	SIK
Tamil Nadu	48.408	364	213.7	0.374	733	TN
Tripura	2.053	78	218.3	0.312	566	TRP
Uttar Pradesh	110.862	307	206.4	0.413	721	UP
West Bengal	54.581	203	210.4	0.376	701	WB
Andaman and Nicobar	0.189	20	179.6	0.482	571	A&N
Arunachal Pradesh	0.632	66	218.3	0.358	572	ARU
Chandigarh	0.452	42	192.0	0.388	600	CHN
Dadara	0.104	14	180.9	0.308	409	DDR
Delhi	6.220	147	214.4	0.330	681	DEL
Goa	1.008	35	177.6	0.401	525	GOA
Lakshadweep	0.040	7	215.3	0.123	287	LAX
Mizoram	0.494	17	256.2	0.197	460	MIZ
Pondicherry	0.604	67	213.2	0.346	629	PON
Daman	0.079	28	188.2	0.314	493	DMN
Total (All India)	685.185	2753*	204.1	0.449	776	ALL

*If a community occurring in more than one states is counted separately for each state, there are 4635 community elements.

**These abbreviations are employed in Figures 1, 3, 4, 5, 6 and 7.

$$C = 0.074\ P^{0.46}, \qquad (1)$$

where C is the number of communities in a state/UT, P the population of that state/UT. (For the log–log least square fit, $r = 0.94$, the 95% confidence limits for r are 0.88 and 0.97).

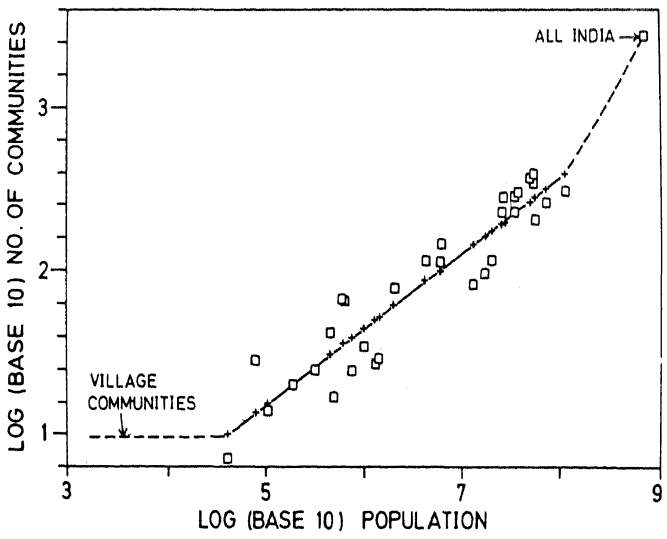

Figure 1. The number of communities occurring within a state/union territory as a function of its population in 1981. The dotted lines represent the number of communities co-occurring at the village level and at the level of the country as a whole.

Union territories are reasonably homogeneous; states are somewhat more heterogeneous, in particular, large ones like Uttar Pradesh that at once embrace the thickly populated Gangetic plains and the more thinly populated hill areas. For units of this level of homogeneity with populations ranging between 1 million and 100 million, the total population size effectively predicts the number of communities. However, the total number of communities predicted for the country as a whole on the basis of 1981 population (685 million) using equation (1) is 900, which is far smaller than the actual number of 2753. This implies that new communities are added *across* states/UTs at a much higher rate than they are *within* the culturally more homogeneous units like the states/UTs (see Figure 1).

Extrapolating equation (1) at the lower end of population size suggests that a village settlement, typically with 500–1000 inhabitants, should support either 1 or 2 communities. Extensive data are available on the composition of village communities throughout India, the best source being the 188 monographs ac-

companying the 1961 census[4]. Only a very small number of such villages located in tribal tracts support just 1 or 2 communities. More commonly villages with populations ranging from 500 to 1000 support 8–16 communities. This much larger number of communities at the lower end of population size reflects the traditional division of labour in rural India, so that while 1 or 2 cultivator (or more rarely, fishing or pastoral) communities make up the bulk of the population, several traditional service/artisan communities occur in smaller numbers (see Figure 1).

How many traits per community: α-diversity

One may explore the diversity of individual communities and groups of communities in terms of measures of diversity used in community ecology[5–7]. If an individual human community is taken to be analogous to a biological community, the number of traits reported by it corresponds to the number of species of a biological community. This is species richness, one of the measures of α-diversity. When one compares several biological communities along an environmental gradient, the number of new species added and old ones deleted is termed species turnover, a measure of β-diversity. The number of traits which are not shared in common by two human communities, or trait dissimilarity would then be an analogous measure of diversity for the POI data. Finally the total number of species reported by a whole set of biological communities of any region is termed γ-diversity. In the case of POI data this would correspond to the total number of traits reported by a group of human communities. This may be termed as trait accumulation.

The currently available computerized POI data base is a list of those attributes (out of a possible total of 776) for which any given community has a positive response. Figure 2 depicts the frequency distribution of

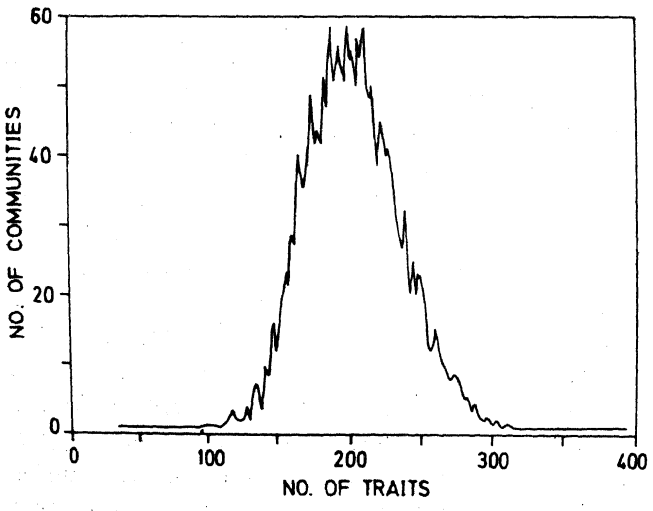

Figure 2. Frequency distribution of the number of traits (out of a possible maximum of 776) reported by the 4635 community elements.

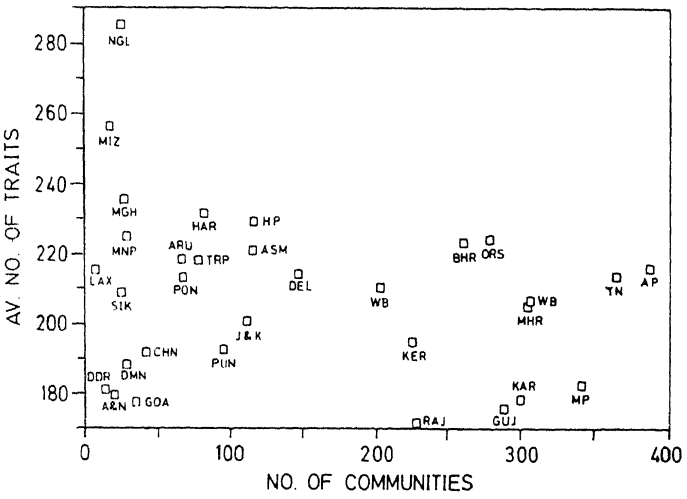

Figure 3. Mean number of traits reported by a community as a function of the number of communities in a state/union territory.

the number of traits which a community reports: the minimum being 32, the maximum 406 with a median of 202 and a mean of 204 with an s.d. of 35.4. The 32 different states and UTs show noteworthy differences in the number of traits reported on an average by communities belonging to that territory (Figure 3). North-eastern states of Nagaland (285) and Mizoram (256) report much higher and the north-western states of Gujarat (176) and Rajasthan (172) report a much lower number of traits than the bulk of other territories (Table 1). This corresponds to the fact that the north-eastern hill states are not only more heterogeneous topographically, but harbour tribal populations which have few food taboos, rather fluid cultural practices and minimal occupational specialization. Indeed most communities reporting large number of traits from other states are also tribals or groups dependent much more directly on natural resources such as Khas of Himachal Pradesh or Gonds of Maharashtra. The rather arid territories of Rajasthan and Gujarat are topographically more homogeneous, and harbour caste populations with extensive food taboos, more rigid cultural practices and considerable caste-based occupational specialization. Groups reporting very small number of traits from other states are either little-studied tribals of Andaman and Nicobar islands such as Sentinelese, Jarwas and Shompens or communities constituted out of normal endogamous groups through some special process like Faqirs (ascetics) of Maharashtra or Hijras (sexually abnormally developed individuals such as gynandromorphs) of Uttar Pradesh.

Differences between communities: β-diversity

The second component of diversity would then be trait dissimilarity, or the proportion of unshared traits amongst two communities (see Appendix). Of the 2753 POI communities, 791 occur in two or more states:

These communities may be expected to exhibit lower levels of trait dissimilarity in comparison with two randomly picked communities. Although the data do exhibit such a tendency with Mann–Whitney test resulting in significant differences at $p<0.01$, the tendency is weak in that the means of the two distributions are not significantly different. Evidently geographical proximity is an important component governing the distribution of the POI traits.

In addition to territory, the 2753 POI communities can be classified on a variety of other bases, such as religion, scheduled caste or tribe, current occupation, rural or urban, nomadic or settled. Amongst such categories, the clearest trend with respect to dissimilarity obtains in case of religious categories. Followers of the ancient Jain religion overwhelmingly follow business/trade as an occupation and are concentrated in a contiguous geographical tract encompassing Rajasthan, Gujarat, Maharashtra and Madhya Pradesh. They exhibit the lowest levels of trait dissimilarity, 0.31 within a state and 0.36 across states. Followers of Christianity have converted much more recently, over a wide geographical tract and from very different cultural backgrounds. They follow an exceedingly wide range of traditional occupations. Trait dissimilarity is highest amongst them for any religious category, 0.38 within a state and 0.46 across states. Sikhs, Hindus, Muslims and Buddhists exhibit, in that order, increasing levels of trait dissimilarity. The trends amongst occupational categories are less clear; nomadic pastorals exhibit the highest levels of trait dissimilarity and agriculturists the lowest with business/trade communities occupying an intermediate position.

Figure 4 depicts the levels of trait dissimilarity as a function of number of communities for the 32 states/UTs. Lakshadweep harbours only 7 communities, all followers of Islam, all Malayalam speakers and all subsisting on fishing and coconut cultivation. They exhibit a very low level of trait dissimilarity of 0.12 (Table 1). So does the predominantly tribal north-eastern state of Mizoram, whose 17 communities exhibit a dissimilarity of 0.2. On the other hand, the Andaman and Nicobar islands have complemented the 6 original tribal communities with 14 others derived from widely different cultural background, so that these have a high dissimilarity level of 0.48. Amongst the more populous mainland states with over 250 communities, Bihar (261 communities) has the lowest dissimilarity level of 0.34 and Maharashtra (305 communities) the highest dissimilarity level of 0.44. Bihar was the seat of the 2500-year-old agricultural civilization of Magadha, and appears to have acquired over centuries a high level of cultural homogeneity. Maharashtra achieved widespread settled cultivation much later and is known as a state transitional between the northern and southern cultural zones. This seems to be reflected in the trait dissimilarity level. We expect low trait richness to be correlated with high levels of trait dissimilarity. As Figure 5 shows, this is borne out by the statistically significant negative correlation ($r=-0.438$, the 95% confidence limits are -0.68 and -0.11) between these two variables, represented by the equation, $\beta=0.627-0.0012\alpha$, where β is the average proportion of unshared traits in a pair of communities, and α is the average number of traits per community in the state/UT.

How many traits in a state: γ-diversity

Trait accumulation, analogous to γ-diversity, is a third measure of diversity that may be examined on the basis of the POI data. Trait accumulation is the total set of traits represented in a group of communities, such as

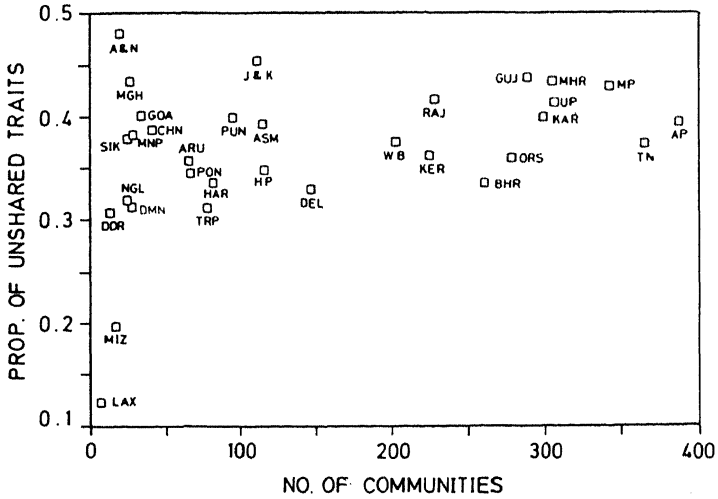

Figure 4. Mean value of dissimilarity (proportion of traits not shared) between communities within a state/union territory as a function of the number of communities in the state/union territory.

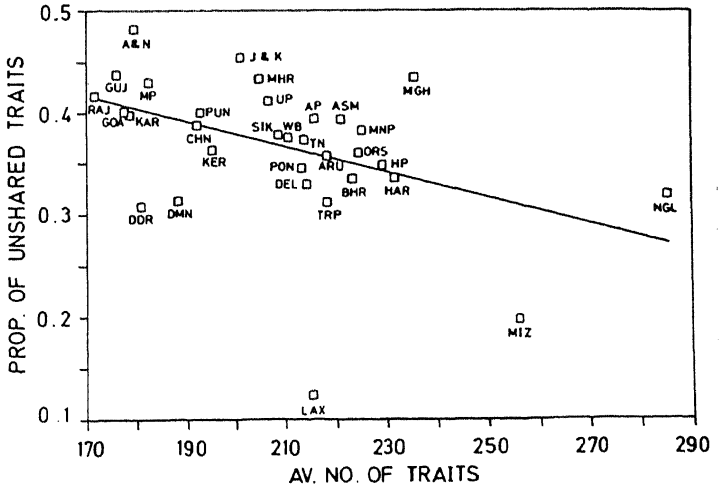

Figure 5. Mean value of dissimilarity between communities as a function of the mean number of traits reported by communities in the 32 states/union territories.

those co-occurring in a state/UT. Obviously trait accumulation is a function of trait richness and dissimilarity. In addition, it depends on how the traits are distributed across communities. If there is a wide variation in the distribution of traits (i.e. some traits being very common, some very rare, etc.) an equation of the form

$$T = An/(B+n) \quad (2)$$

should provide a reasonable fit[8], where A and B are functions of trait richness and dissimilarity, and n is the number of communities. Empirically, the POI territory level data on trait accumulation (Figure 6) are well-fitted by the equation:

$$T = 730\,n/(10+n) \quad (3)$$

(least square fit for $1/T$ vs $1/n$, $r = 0.95$, the 95% confidence limits for r are 0.90 and 0.97).

As the number of communities becomes large (say > 100), T asymptotically approaches a value of 730. In the POI data Maharashtra with 305 communities has the highest T value of 735, with Tamil Nadu with 364 communities close behind at 733 (Table 1). The very high accumulated trait set of Maharashtra seems to reflect its intermediate position in the north–south cultural gradient, with communities representing features typical of north as well as south. Lakshadweep with only 7 rather specialized and homogeneous set of communities has the lowest accumulated trait set of 287, well below the maximum of 406 for a single community, that of Khas in the Himachal Pradesh.

Correlation analysis: searching for clusters of traits

Beyond measuring diversity, one may look for patterns in the distribution of diversity in terms of geography as well as social and economic stratification. The simplest geographical hypothesis may be that the set of traits held by groups of communities diverges with distance between them. Thus while neighbouring Bihar and Orissa share 63% of traits with each other both share only 54% of traits with the distant Gujarat (see Appendix). But further exploration reveals this hypothesis to be overly simplistic. Thus while Himachal Pradesh shares 55% of traits with neighbouring Jammu and Kashmir, it shares 56% with distant Assam, and Haryana shares 58% of traits not only with neighbouring Punjab, but distant Tripura as well. However, while distance is not a simple predictor of divergence between two groups of communities as in states/UTs, a multivariate analysis reveals these territorial elements to cluster in an interpretable fashion. Figure 7 is a representation of the similarities between different territorial elements through multi-dimensional scaling. Geographical proximity evidently does explain cultural similarity to a fair extent.

Indian society continues to be predominantly an agricultural society. It is therefore of considerable interest to enquire whether there are clusters of cultural traits associated with the use of specific food plants (see Appendix). An exploration of the People of India data reveals several such clusters; six of these, associated with wheat/moong/masur/mustard, jowar/ragi/tur groundnut, barley/maize/mustard, coconut, peas/beans and ragi merit further discussion. The first of these, associated with wheat/moong/masur/mustard spans over the Indus–Ganga–Brahmaputra valleys and includes a set of linked cultural traits involving *gotra*-based exogamy, *kanyadan* during marriage, use of vermilion

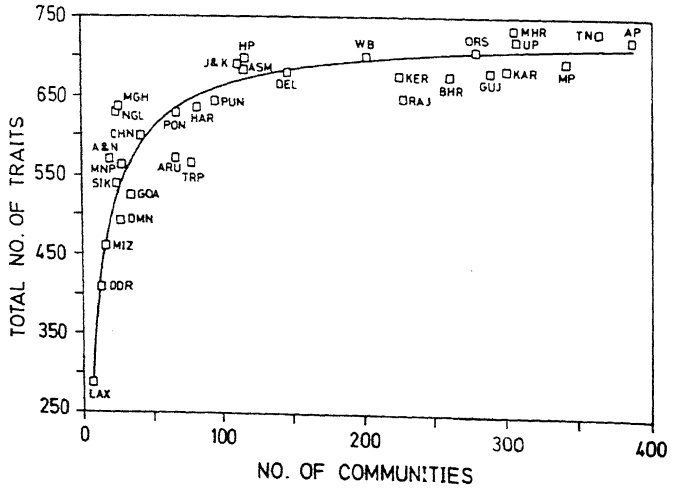

Figure 6. Total number of traits reported as a function of the total number of communities for the 32 states/union territories.

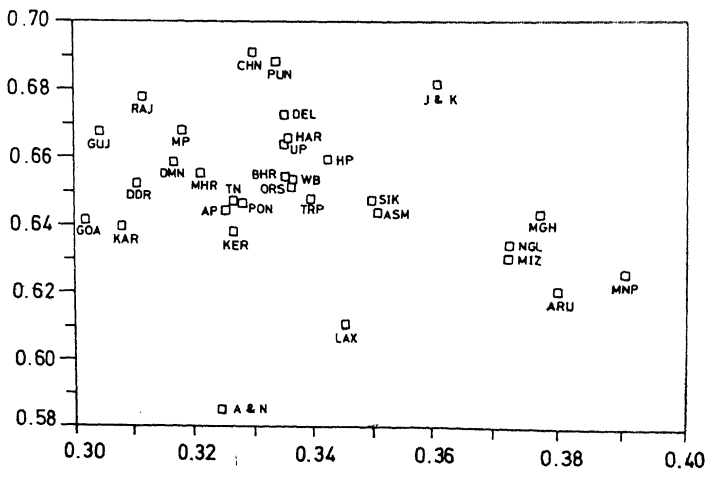

Figure 7. Extent of similarity between states and union territories for the traits reported by the constituent communities. This depiction is based on a multi-dimensional scaling analysis.

and bangles as symbols of a woman's married status, cremation of dead, joking relationship between sister-in-law and younger brother-in-law, avoidance between daughter-in-law and father-in-law as well as between sister-in-law and elder brother-in-law, and acceptance of water from members of other communities. The complex of jowar/ragi/tur/groundnut characterizes the south Indian plateau and Tamil Nadu plains and is associated with a set of linked cultural traits including surname-based exogamy, practice of uncle–niece and cross-cousin marriages and the use of *tali* and toerings as symbols of married status of women. Coconut, universally used on the west coast is linked to use of betel leaf and snuff, cross-cousin marriages, performance of marriage ceremony in bride's residence, use of *tali* as a symbol of married status and observance of puberty and pre-delivery rituals. The complex of barley/maize/mustard characterizes the hilly areas of Western Himalayas and is linked to terrace cultivation, smoking of loose tobacco, acceptance of levirate form of marriage, participation of caste council in resolution of disputes over land, practices of exorcism and divination and attribution of putative kinship to members of other communities. We are, of course, not stating anything novel in identifying the geographical spread of these traits, since diffusion of cultural traits is well known to have taken place within these four broad geographical tracts. However, the identification of such expected patterns from our data strengthens our faith in other, less obvious patterns.

Hilly, still partially forested regions of central India and northeast harbour many tribal communities which share several traits, not so much because they have been in communication, but because these traits adapt them to their more primitive mode of resource use. There is indeed a large cluster of traits associated with shifting cultivation, primarily an activity of tribals. This cluster includes habitation in forest habitat and extensive use of forest produce, use of streams as a water source, hunting, fishing and gathering activities, terrace cultivation, maintenance of chicken and pigs, basket weaving, women's active participation in fishing, agriculture and fuel collection, role of traditional village council in punishing offences like adultery and theft, animistic beliefs, practice of exorcism and participation of both sexes in singing folk songs and in dancing.

Less expected, and therefore more noteworthy is the linkage of a number of traits related to occupation with specific food habits. The consumption of peas and beans is strongly linked to a community's participation in modern service, trade and industrial sectors, high levels of education for both boys and girls, use of piped drinking water, watching television and saving through banks and other institutions. On the other hand, use of ragi is strongly linked to unskilled agricultural labour, employment of child labour, use of fuelwood, women's involvement in collecting fuelwood, fetching water and participation in agricultural and animal husbandry activities, as well as wage earning, community's involvement in barter economy and weekly markets and worship of village deities.

A significant component of Indian population is nomadic, and POI records 276 non-pastoral and 37 pastoral nomads amongst its 4635 community elements. One expects pastoral nomads, i.e. shepherds, cowherds, etc. who move over large distances with their livestock primarily to occur in habitats where grazing resources are highly seasonal. This would be the case in regions of scanty rainfall or seasonal snowfall. Indeed while only 26 out of 3954 (0.65%) communities in regions of high or medium rainfall practice nomadic pastoralism, as many as 9 out of 646 (1.5%) of communities in tracts of low rainfall, 4 out of 103 (4%) in tracts of scanty rainfall and 15/196 (7.5%) in tracts of snowfall do so. Part of cultural diversity of the Indian subcontinent can thus be neatly explained by the ecological setting.

Occupational diversity: traditional and contemporary patterns

Finally, cultural traits of Indian communities are in no way stagnant; they are undergoing rapid change. The POI data set has substantial information on this aspect as well, information pertaining to changing patterns of religious practices, marriage relationships, family planning practices, health care, food habits, education, occupation and so on. We illustrate the emerging patterns here by looking at occupation, since transmission of occupation from one generation to the next was a notable feature of the Indian caste society. The POI data classify occupations in 42 categories and record traditional as well as current practices. As expected, traditional occupations show a strong mode at 1 (Figure 8), the average (computed by excluding

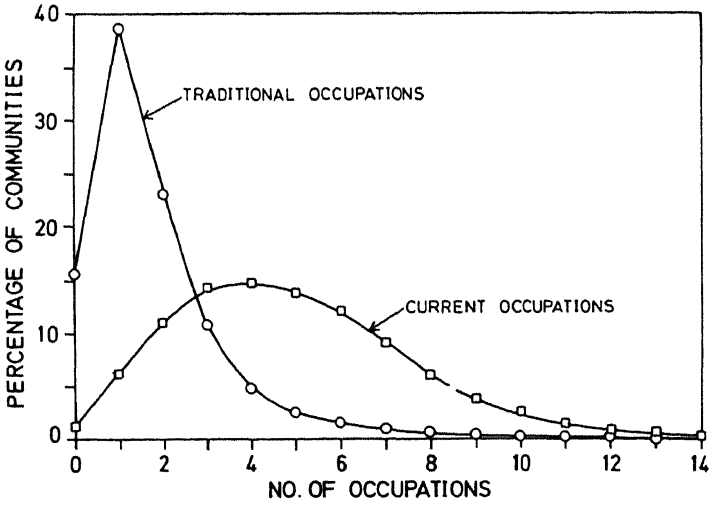

Figure 8. Frequency distribution of the number of traditional and current occupations (out of a possible maximum of 42) reported by the 4635 community elements.

communities for which no information exists) being 2.8 with a s.d. of 1.8. The maximum is 17. It is notable that all communities reporting large numbers of traditional occupations are tribal communities, not within the ambit of the traditional caste-based division of labour. In contrast, the mode of current occupation has shifted to 4, with a mean of 5.1, an s.d. of 2.2 and a maximum of 25. Indian communities are clearly widening their base of occupation. It is also noteworthy that while the five commonest traditional occupations, in decreasing order, were settled cultivation, unskilled labour, animal husbandry, trade and business and hunting and gathering; the five commonest current occupations (in terms of involving at least some members of a given community), again in the same decreasing order are Government service, settled cultivation, unskilled labour, self-employment and private service. People of India are evidently part of a dynamic society. However it is significant that fully 528 out of the 4635 community elements have apparently failed to enter even marginally the most valued occupational categories of Government and private service.

Appendix

Similarity between communities

For a pair of communities, this is defined to be the probability that a randomly picked trait reported by one of the communities is also reported by the other.

If T_1, T_2 are the number of traits reported by the two communities, and T_{12} is the number of traits seen in both the communities, then the similarity will be: $S_{12}=2\cdot T_{12}/(T_1+T_2)$, and the dissimilarity is given by $D_{12}=1-S_{12}$.

The similarity and dissimilarity are thus the proportions of shared and unshared traits, respectively, for a pair of communities.

Similarity within a group of communities

Similarity within a group is the expected similarity between two communities, randomly chosen from the same group. If a group (e.g. a state) contains N communities there can be $N(N-1)/2$ comparisons of pairs of communities. The *mean similarity*, computed over the $N(N-1)/2$ similarities, has been termed as the similarity between communities within a group. Thus the statement that 'the dissimilarity within Mizoram is 0.2' implies that the similarity between two randomly picked communities from Mizoram is expected to be 0.8.

Similarity between groups of communities

Similarity between two groups is the expected similarity between a pair of randomly picked communities, one from each of the two groups. If the two groups contain N_1 and N_2 communities, there can be $N_1 \times N_2$ such pairs, and the mean similarity computed over all these pairs is termed as the similarity between the two groups. Thus, the statement that 'Bihar and Orissa share 63% of traits' implies that the similarity between two randomly picked communities, one each from Bihar and Orissa, is expected to be 0.63.

Correlation between traits

For two traits X and Y, let

a = Number of communities showing both X and Y
b = Number of communities showing X but not Y
c = Number of communities showing Y but not X, and
d = Number of communities showing neither X nor Y.

For the present data set, for any pair of traits,

$$a+b+c+d=4635.$$

The Pearson product moment correlation coefficient between the two traits is given by

$$r=\frac{ad-bc}{[(a+b)(a+c)(b+d)(c+d)]^{1/2}}.$$

Testing the significance of correlations

When the number of observations (N) is large, the standard error of the correlation coefficient (r) is $1/(N-3)^{1/2}$. For the correlation to be significantly different from zero at 1% level, the absolute value of r should be higher than 2.58 times the standard error. We have computed correlations between all pairs of 711 traits (such that each trait was reported by at least 30 communities). However, since r between each trait with all the 710 others is tested for significance, about 7 traits are expected to turn out to be significant at 1% level, by chance alone. For such a situation of multiple comparisons, the Bonferroni confidence intervals have been used by carrying out the testing of significance at 1/710 per cent, i.e. at 0.0014% level.

Clusters of traits

A cluster of traits is a group of traits such that all the traits in the group are significantly and positively correlated with each other. To obtain such a cluster centered around trait X, say, we initially identify all the traits significantly positively correlated with X. Suppose there are M such traits. We next examine the correlation of each of the M traits with the remaining $M-1$ traits. In case not all pairs are significantly correlated with each other, that trait which is correlated to the least number of other traits is dropped from the set, to leave a set of $M-1$ traits. This process is iterated till a set of M' traits is left, which are all significantly positively correlated to each other. This is the desired cluster. The analysis presented in the text is based on an examination of a large number of such clusters, obtained using many different traits as starting points.

1. Singh, K. S., *People of India: An Introduction*, Seagull Books, Calcutta, 1992.
2. Singh, K. S., *Curr. Sci.*, 1993, **64**, 5–10, this issue.
3. Padmanabha, P. (ed.), *Census of India 1981*, Controller of Publications, Delhi, 1981.
4. *Census of India*, 1961, Village Survey Monographs.
5. MacArthur, Robert, H., *Geographical Ecology*, Princeton University Press, Princeton, 1972.
6. Pielou, E. C., *Ecological Diversity*, John Wiley & Sons, New York, 1975.
7. Cody, M. L. and Diamond, J. (eds.), *Ecology and Evolution of Communities*, The Belknap Press of Harvard University Press, Cambridge, 1975.
8. Ranjit Daniels, R. J., Joshi, N. V. and Gadgil, M., *Biol. Conserv.*, 1990, **52**, 37–48.

ACKNOWLEDGEMENTS. We are grateful to Dr K. S. Singh and his colleagues in the Anthropological Survey of India for making this fascinating data set available, and for many stimulating discussions as well as financial and moral support. K. C. Malhotra kindly made available his summary of village survey monograph data to us. Anil Gore and A. V. Kharshikar read an earlier draft of this MS and made a number of useful suggestions. This work has also been supported by a grant from the Ministry of Environment and Forests, Government of India.

Received 27 October 1992; accepted 2 November 1992