Title
Segregation by Racial and Demographic Group: Evidence from the San Francisco Bay Area

Authors
Miller, Vincent P.
Quigley, John M.
Wiseman, Michael L.

Publication Date
1988-02-01

Peer reviewed

iberi Institute of Business and Economic Research
University of California, Berkeley

CENTER FOR REAL ESTATE AND URBAN ECONOMICS
WORKING PAPER SERIES

WORKING PAPER NO. 88-140

SEGREGATION BY RACIAL AND DEMOGRAPHIC GROUP: EVIDENCE FROM THE SAN FRANCISCO BAY AREA

BY

VINCENT P. MILLER

JOHN M. QUIGLEY

MICHAEL L. WISEMAN

GRADUATE SCHOOL OF BUSINESS ADMINISTRATION

CENTER FOR REAL ESTATE AND URBAN ECONOMICS
UNIVERSITY OF CALIFORNIA AT BERKELEY

The Center was established in 1950 to examine in depth a series of major changes and issues involving urban land and real estate markets. The Center is supported by both private contributions from industry sources and by appropriations allocated from the Real Estate Education and Research Fund of the State of California.

INSTITUTE OF BUSINESS AND ECONOMIC RESEARCH
J. W. Garbarino, Director

The Institute of Business and Economic Research is a department of the University of California with offices on the Berkeley campus. It exists for the purpose of stimulating and facilitating research into problems of economics and of business with emphasis on problems of particular importance to California and the Pacific Coast, but not to the exclusion of problems of wider import.

SEGREGATION BY RACIAL AND DEMOGRAPHIC GROUP: EVIDENCE FROM THE SAN FRANCISCO BAY AREA

by

Vincent P. Miller
John M. Quigley
Michael L. Wiseman

University of California
Berkeley

February 1988

Abstract

Segregation By Racial and Demographic Group: Evidence from the San Francisco Bay Area

This paper considers residential segregation by race and by type of household in 1970 and in 1980. The paper presents entropy indices of segregation for the San Francisco Bay Area and for its five metropolitan areas. The methodology permits an investigation of the effects of group definition upon segregation measures, and an analysis of the degree of independence in the spatial clustering of households by race and demographic group.

The results indicate that the level of segregation by race, as well as the level of segregation by household type, has declined modestly during the 1970's, at least in this region. More importantly, however, the results indicate a remarkable independence in the spatial distribution of households by race and demographic group. Only a very small fraction of the observed levels of segregation by race could be "explained" by the prior partitioning of households by demographic group. The principal results of the analysis are invariant to changes in the definition of racial or household groups.

I. INTRODUCTION

Even the most casual observer notices that residential patterns in American urban areas are highly segregated by race. It is only slightly less obvious that urban areas are segregated by income, by household size and composition, and by other demographic characteristics. Presumably, residential clustering by sociodemographic group reflects similarity of tastes for local public goods and locational amenities and similarity in disposable income. Residential clustering by race may reflect the same phenomenon. It may also reflect the outcomes of a discriminatory market in which minority households are denied access to the entire housing stock or in which minority households feel less threatened by choosing to reside in close proximity.

Disentagling "natural" segregation by sociodemographic group from that which arises from racial discrimination is no easy task. Yet the distinction is important to interpreting trends in segregation and to promoting equal opportunity. If, for example, levels of housing market discrimination have *declined* while demographic differences among races have increased, the observable result may be *increased* tendencies towards spatial segregation by race. This situation would imply that the determinants of housing segregation are increasingly rooted in economic differences among households and not in the resistance of actors in the real estate market to the granting of equal access to rental and sales markets.[1]

This issue in interpretation arises in other contexts. For example, in the decade between the 1970 and 1980 census, the poverty rate for U.S. black

---

[1] See Yinger [1986] for an extensive discussion.

households declined very modestly from 29.5 percent to 28.9 percent. However, this aggregate outcome masks quite significant changes within individual categories of households. The persistence of the high average poverty rate reflects the offsetting effects of simultaneous increases in the incomes of intact families and increases in the proportion of black households in groups with "high risks" for poverty, in particular in families headed by single women. In fact, if the composition of black households in 1980 had been the same as in 1970, then the decline in poverty rates that occurred within each group would have led to a 9.6 percentage point decline in the black poverty rate (to 19.9 percent), instead of the 0.6 percentage point actually observed.[2]

These developments in household composition have spatial consequences. It is important to examine the extent to which racial segregation reflects reductions in the residential segregation of household types counteracted by shifts in household composition towards single-earner (or no-earner) households who are more likely to be in poverty and thus to have fewer options for housing and location.

Unfortunately most empirical studies of residential segregation have ignored these distinctions among discrete demographic groups in focusing attention on the occupancy patterns of two racial groups. Even at this aggregate level, however, similar problems in interpretation arise. For example, reductions in the level of racial segregation between blacks and nonblacks reported in the 1970's may reflect, to an unknown extent, the increasing spatial integration

---

[2] See Green and Welniak [1982].

of blacks with other minorities (especially hispanics) combined with an increasing segregation of white and minority households.[3]

These problems in interpretation arise because analyses of segregation are based empirically upon a binary representation of residential location -- black-nonblack or white-nonwhite -- by census tract or urban neighborhood.

In this paper we consider the decomposition of residential segregation by several distinct household types and races, as well as location. The paper begins with a careful definition of residential segregation, a cursory comparison of some common measures of the phenomenon, and presents an analysis of segregation patterns in the San Francisco Bay Area as indicated by the 1970 and 1980 census reports.

We focus on the the San Francisco area for several reasons. First, it is a large and economically diverse metropolitan area. Segregation patterns observed are thus more likely to be meaningfully compared with other large and diverse metropolitan areas. Second, the region is among the fastest growing in the United States, and demographic trends are starkly visible. Third, the region has several large and well-defined ethnic populations, including black, hispanic and Asian Americans. Finally, the spatial character of the region is well defined. Like most western metropolitan areas, it is relatively recently developed. But unlike most others, the density and housing price gradients reveal a monotonic decline from a central business district. Section II below introduces the methodological issues in the comparisons. Sections III and IV introduce the data and the comparative measures employed. Section V presents

---

[3] See Massey [1979] for a discussion.

an extensive analysis of segregation by race, demographic group and location in the region.

II. SEGREGATION

Residential segregation can refer to both a process and an outcome. This paper is concerned with measures of the outcome.

A. Definitions

Consider households' choices to reside at various points in an urban area, where conflicting choices are resolved by some impersonal mechanism. Now if the conditional distribution of households by race in space differs from the unconditional distribution, the population may be said to be segregated by race. Of course, if conflicting choices are resolved by a price mechanism, then differences in income or wealth among races may lead to this segregation. Also, in these circumstances systematic differences by race in other factors which affect preferences for location, for example family size and household composition, can cause some degree of residential segregation by race.

This definition of segregation is scale invariant (Allison, 1978); it is independent of the number of racial groups involved in the analysis, the size of the total population, and the overall distribution of area households among races. Scale invariance permits direct comparisons of segregation to be made for differing geographical areas or time periods. Of course, scale invariance may not be an appropriate property for all purposes. For example, suppose segregation were of concern principally because it inhibits social interaction among people of different races, and suppose the likelihood of interaction were a function of the probability that one sees members of another racial group. Then a doubling of population density would lead to a reduction in the impact of segregation (since it would raise the expected number of individuals of a different race within a fixed distance of any household).[4]

The spatial implication of these definitions is that, in the absence of segregation, all subdivisions of a larger area will have the same distribution of households by race. This suggests that segregation be measured empirically by partitioning the area to be analyzed into subareas and examining the racial composition of each; under this approach, measures of the variation in racial composition across subareas may be taken as measures of the degree of segregation. Two related problems arise with this approach. One is that, as the size of the subareas increases, the same physical area appears less segregated. In the limit, when the subarea subsumes everything, the metropolitan area must be "integrated." A second is that the way in which the area is partitioned can affect the degree of segregation implied. For example, a checker-board pattern of residential occupancy by race can give rise to extreme differences in residential occupancy by subarea or to identical measures of the racial composition of subareas, depending only upon how the checker-board is partitioned.

Despite the potential importance of this boundary problem, any empirical analysis of patterns of U.S. residential segregation must ultimately begin with counts of individuals or households by predetermined geographical areas: census tracts or perhaps block faces.[5] Census tracts were established to have stable

---

[4] See Lichter [1985] for a discussion of racial concentration, density, and racial segregation.

[5] Any analysis of data by block faces is severely compromised by Census Bureau confidentiality rules which lead to the suppression of population counts by various categories, including simple counts of households by race or housing type.

boundaries, and were "designed to be relatively homogeneous areas with respect to population characteristics, economic status and living conditions" (U.S. Bureau of the Census, 1982, p.8). Any measure of segregation is conditional upon the prior partitioning of the urban area into these geographical subareas.

B. Measurement

The empirical analysis in this paper relies principally upon the entropy measure of residential segregation which seems ideally suited to the problem of analyzing segregation by race and demographic group. The discussion begins with a cursory review of other more common measures of segregation.[6]

As noted above, any quantitative measure of segregation must begin with counts of households by racial or other group residing in subareas (census tracts).

Let n_{it} be the number of individuals of group i residing in census tract t. Thus $n_{*t} = \Sigma_i \, n_{it}$ is the total number of individuals residing in t, and $n_{i*} = \Sigma_t \, n_{it}$ is the total number of individuals of type i in the entire area. Finally, $n_{**} = \Sigma_i \, n_{i*} = \Sigma_t \, n_{*t}$ is the area population.

The most common quantitative measure of the level of residential segregation computed from these counts of individuals by census tract is the so-called "dissimilarity index" popularized by Taeuber and Taeuber (1965).

---

[6] See White [1986] for a more extensive comparison of many common measures of segregation.

The dissimilarity index measures the level of segregation between two groups. Suppose each neighborhood t is composed of n_{1t} and n_{2t} individuals of group 1 and 2 respectively. The dissimilarity index D,

$$(1) \quad D = \frac{\Sigma_t \; n_{*t} |(n_{1t}/n_{*t}) - (n_{1*}/n_{**})|}{2 \; n_{**}(n_{1*}/n_{**})(n_{2*}/n_{**})} ,$$

is the sum of the absolute deviations of the racial proportions of census tracts from the overall racial proportion, normalized. The normalization factor, the denominator of (1), represents the maximum of this sum under complete segregation. The index thus ranges from 0 to 1.

The index does have an appealing intuitive interpretation; its value represents the minimum proportion of the population that would have to relocate to eliminate segregation. The traditional measure is seriously deficient, however, on other grounds. First, its properties violate the common sense principle of transfers (Allison, 1978), namely that an index of segregation should decrease in value when members of a group move from an area of higher group concentration to one of lower concentration. The value of the dissimilarity index, however, depends only upon the distribution of group members between areas with above and below average concentrations.[7] Second, the index is

---

[7] After some manipulation, equation (1) can be restated:

$$(1') \quad D = \Sigma_{t \varepsilon A} \; (n_{1t}/n_{1*}) - \Sigma_{t \varepsilon A} \; (n_{2t}/n_{2*})$$

where A represents those census tracts for which $(n_{1t}/n_{1*}) > (n_{1*}/n_{**})$, *ie*, those for which the proportion of group 1 exceeds the proportion in the entire area. Thus the value of D is unaffected by the assignment of individuals within the set A or its complement. Schnare (1980) has presented some examples of the curious properties of this index in the context of residential segregation by race.

not well defined when there are more than two groups. Increasingly U.S. metropolitan areas are characterized by several identifiable minority groups, and the dissimilarity index is deficient in representing that heterogeneity.[8]

An alternative index measures the "exposure" of one group of residents to others. The exposure index is the weighted average proportion of agents in each area who are not members of the same group. The exposure of any population group i, to all other groups, $\bar{i}$, is defined as

$$(2) \quad E_{i\bar{i}} = (1/n_{i*})\ \Sigma_t\ n_{*t}\ (n_{it}/n_{*t})(1-n_{it}/n_{*t})$$

In contrast to the dissimilarity index, the value of the exposure index does depend upon the distribution of population groups within each subarea. Moreover, the exposure index can be decomposed into a weighted average of the exposure of members of group j to each of other subgroups.

$$(3) \quad E_{i\bar{i}} = \Sigma_{k\neq i}\ (n_{k*}/n_{**})\ E_{ik}$$

---

8 A number of papers have used the dissimilarity index to analyze the segregation of hispanic households. Massey [1979] and Massey and Bitterman [1985] considered the segregation of hispanic from "other" households in two metropolitan areas, while Hwang, *et al* [1985] considered the segregation of hispanics in Texas. The latter study computed dissimilarity indices for three pairwise combinations of households: white-black, white-hispanic, and hispanic-black. In addition, at least two papers have attempted to generalize the dissimilarity index to three or more groups (Sakoda [1981], Morgan and Norbury [1981]). There is, however, no convenient way to extend the dissimilarity index to several groups (See Theil [1972] for an extensive discussion).

The exposure index cannot, however, be decomposed spatially or geographically.

In contrast, the entropy index (H) is the only measure of segregation which satisfies the properties of symmetry, continuity, and full additivity. The entropy index is defined as

(4) $H = \Sigma_t \Sigma_i (n_{it}/n_{*t}) \log (n_{*t}/n_{it})$

The principal advantage of the entropy index in representing the segregation of households by household type, race, and location is illustrated in the analysis that follows.[9]

III. THE DATA

As noted above, this analysis of spatial segregation is based upon data from the San Francisco Bay Area (The "San Francisco-San Jose-Oakland Consolidated Metropolitan Statistical Area") which includes nine counties and five Standard Metropolitan Statistical Areas (SMSA's). The analysis is based upon census tract data for 1970 and 1980, consisting of 1079 census tracts (according to 1970 boundaries). Figure 1 presents, in schematic terms, the five SMSA's which

---

[9] One deficiency with the index remains. The entropy index does not overcome the boundary problem arising from the arbitrary partitioning of space into subareas. This can be addressed crudely by accounting for the distance between each pair of subareas and assuming that population is concentrated at the centroid of each subarea. See White [1983, 1984] for a discussion. We do not pursue this extension here, because, in contrast to the entropy index, these distance-related measures do not preserve the property of additivity. They also require extensive geocoding of subareas.

make up the Bay Area, the central cities of each SMSA,[10] and the census tracts which form the ultimate building blocks for the analysis.

The empirical analysis considers the segregation of households by race and household type as well as location. Unfortunately, the census definitions of race and household type were changed for the 1980 census in a way that complicates intertemporal comparisons by race and ethnicity. One difference arises in the classification of persons of hispanic origin. In the 1970 census, hispanic was treated as a separate race category, but in the 1980 census, race and hispanic origin were assessed in separate questions. The result was, as might be expected, a considerable range of race-hispanic mix. For comparability over time we have created a separate hispanic racial category for 1980 by re-allocating all persons of any race citing hispanic origin. The result is a mutually exhaustive six-category race-ethnicity classification, comparable, but more detailed, than that available in the 1970 census.

Table 1 reports the 1970 and 1980 racial composition of the Bay Area in these mutually exclusive categories. The table also reports the raw data for 1980 from which these totals were calculated, as well as the less detailed racial composition reported in the 1970 census. For the nine county region as a whole, about 70 percent of the population is classified as white, and 12 percent is of hispanic origin. About 9 percent of the population is classified as black or Asian. During the decade of the 1970's, the total population of the region increased by almost 13 percent. The hispanic population increased by 45,000, and the black population increased by about 100,000.

---

[10] Note that one SMSA, the "Vallejo-Napa-Fairfield" SMSA (hereafter "Napa") contains three central cities.

In contrast, the non-black non-hispanic population declined by 36,000 people. One suspects that the Asian population increased substantially during this period, but census data provide no evidence on this.

The classification of population into household types is less problematic, but is still not without some complications. According to U.S. census conventions, the population is counted by family and by household. Families are defined on the basis of relationships; households are defined on the basis of living quarters.[11] Households are of two basic types. Family households include two or more related persons living together. Non-family households are persons living alone or sharing living quarters with persons to whom they are not related.[12]

Average household size in the Bay Area is 2.6 persons, and 97.7 percent of the population resides in households. Table 2 presents data on the distribution of Bay Area households by the six races defined above and by seven major types of household. These types include traditional husband-wife families with and without children, single adults living alone, by sex, single parent households, by sex, and non family households containing two or more adults.[13] Note that Asian, hispanic, and "other" households are far more likely to involve married couples with children than is true for white, black, or native American households. Also, black households are three times more likely to be made up of an

---

[11] Persons not living in households live in group quarters without separate cooking facilities, such as college dormitories.

[12] This latter category includes the famous "Persons of Opposite Sex Sharing Living Quarters" category. Of particular importance in San Francisco, this category also includes homosexual couples.

[13] Race is defined by the race of the "householder," generally the adult cited first by the census respondent.

unmarried female head with children than is the case for other groups. Among households with children, 45 percent of black households are headed by single women, compared to 16 percent for all other groups. 27 percent of all the households in the Bay Area are white, non-family households. Only 22 percent of all households are white married couples with children. Married couples of all races with children account for only 27 percent of Bay Area households.

Less detail about household types is available from the 1970 census. In particular, the 1970 census did not distinguish between families with children and those without children. Households were recorded in only four categories: families with married couples, headed by unmarried males or females, and non-family groups. Table 3 presents the comparable race and household type information available for the Bay Area for 1970 and 1980. For each of these four household types, counts are available separately for black, hispanic, and all other (i.e., non-black, non-hispanic) households. The table presents a comparison of these twelve categories as reported in the 1970 and 1980 census.

Between 1970 and 1980 the number of black households increased by 39 percent, and the number of hispanic households increased by 16 percent. Despite these increases, however, the number of black households consisting of married couples increased by only 4 percent, and the number of hispanic households consisting of married couples actually declined by 7 percent. The largest comparable race household category in 1970 and in 1980 is other (non-black, non-hispanic) married couples, but the fraction of Bay Area households comprising this category fell from 53 percent to 45 percent during the decade.

IV. RESIDENTIAL SEGREGATION BY DEMOGRAPHIC AND RACIAL GROUP[14]

Households in the Bay Area are partitioned into race categories and household types and into a variety of spatial groupings, for example central city or suburban location in each of the five SMSA's.

Again, let the subscript i denote the category of household, race or household type, let t denote an index of the geographical areas (census tracts) included in the area, and let M_s be the set of census tracts in some spatial aggregation s, such as a central city or a metropolitan area.

Thus

(5) $\rho_{it} = n_{it}/n_{*t}$

is the number of households of category i as a fraction of all households in census tract t, and

(6) $\omega_t = n_{*t}/n_{**}$

is the number of households in census tract t as a fraction of all households. Similarly,

(7) $W_s = \Sigma_{t \varepsilon M_s} \ \omega_t$

[14] A more extensive discussion and proof of the results summarized in this section is contained in Theil [1972]

is the number of households in aggregation s (a central city or a metropolitan area, for example) as a fraction of area total. Following (4), the entropy at the level of census tract is

(8) $$H_t = \Sigma_i \; p_{it} \log(1/p_{it})$$

and the entropies for some aggregate level s and for the entire region level are, respectively,

(9) $$H_s = \Sigma_{t \varepsilon M_s} H_t$$

(10) $$H = \Sigma_t H_t = \Sigma_s H_s$$

Similarly the average entropy of the census tracts in region s is

(11) $$\overline{H}_s = \Sigma_{t \varepsilon M_s} (\omega_t / W_s) H_t,$$

and, the average entropy of all tracts in the area is

(12) $$\overline{H} = \Sigma_t \omega_t H_t = \Sigma_s W_s \overline{H}_s \quad .$$

Note that each of the entropy measures in (9) through (12) is a simple linear combination of the entropy at the level of the census tract. Clearly, for any number of household categories, I, entropy is maximized when each of the underlying probabilities in (8) is equal to (1/I). It follows that the maximum entropy of any region depends upon the aggregate distribution of population among each of the categories.[15]

The maximum possible entropy of any region, $\overline{K_s}$, is

$$(13) \quad \overline{K_s} = \Sigma_I \, P_{is} \log(1/P_{is}) \quad ,$$

where P_{is} is the proportion of population of type i in region s

$$(14) \quad P_{is} = \Sigma_{t \varepsilon M_s} (\omega_t / W_s) \quad \rho_{it} \quad .$$

As before

$$(15) \quad \overline{K} = \Sigma_s \, W_s \, \overline{K_s} \quad .$$

For the region as a whole, K measures the maximum possible entropy level, given the overall distribution of households among categories, H measures the actual level of entropy, given the observed pattern of segregation by census tract.

Thus

$$(16) \quad Z = (\overline{K}-\overline{H})/\overline{K} \quad ,$$

$$Z_s = (\overline{K_s}-\overline{H_s})/\overline{K_s}$$

measures the relative reduction in entropy arising from the spatial segregation of household types in the entire region or in any aggregation s.

---

[15] This merely restates the commonplace observation that all schools cannot be integrated "fifty-fifty" when the aggregate student body is composed of eighty percent members of one race.

The previous discussion deals with classifications in one dimension, say racial categories. The extension to the bivariate case -- the joint distribution of race, r, and household type, h -- is straightforward.

As before

(17) $\rho_{rht} = n_{rht}/n_{**t}$

is the number of households of race r and housing type h as a fraction of all housholds in census tract t. The probabilities of the two marginal distributions are

(18) $\rho_{r*t} = \Sigma_h \; \rho_{rht}$

$\rho_{*ht} = \Sigma_r \; \rho_{rht}$

and the entropies of these distributions are

(19) $H(r)_t \quad = \Sigma_r \; \rho_{r*t} \log (1/\rho_{r*t})$

$H(h)_t \quad = \Sigma_h \; \rho_{*ht} \log (1/\rho_{*ht})$

$H(r,h)_t = \Sigma_r \; \Sigma_h \rho_{rht} \log (1/\rho_{rht})$

These entropies can clearly be aggregated to some spatial level by analogy to equations (9), (10), (11) and (12).

Further it can be shown that

(20) $H(r,h) = H(r) + H_r(h) = H(h) + H_h(r)$

where $H_h(r)$ and $H_r(h)$ are the average entropies of r conditional upon h and vice versa

$$(21)\quad H_h(r) = \Sigma_r \Sigma_h \rho_{rh} \log (\rho_{r*}/\rho_{rh})$$

$$H_r(h) = \Sigma_r \Sigma_h \rho_{rh} \log (\rho_{*h}/\rho_{rh}) \quad .$$

These conditional entropies have a convenient interpretation in terms of segregation. $H_h(r)$, the average conditional racial entropy, measures the extent of racial integration of a geographic area conditional upon the extent of segregation by household type. Similarly $H_r(h)$, the average conditional household type entropy, measures the extent of integration of household types conditional upon the extent of segregation by race. These conditional entropies must always be smaller than the unconditional entropies unless the distributions of race and household type are completely independent.[16]

V. RESULTS

Table 4 compares, for 1980, the household type and racial entropy of the geographic components of the San Francisco Bay Area with the maximum entropy possible. The comparison is based upon both the four and six racial classifications and the four and seven household classifications noted in Tables 1-3. Considering all six races, the maximum racial entropy in the region is 0.978,

---

[16] Specifically

$$H_h(r) \leq H(r)$$

$$H_r(h) \leq H(h)$$

with the equality holding if and only if

$$\rho_{rht} = \rho_{r*t}\, \rho_{*ht} \qquad \text{for all r,t, and t}$$

which would be obtained if each and every census tract had the racial composition of the region as a whole -- that is, if each tract had the racial proportions reported on the last line of Table 2. The actual racial entropy of the region is lower, 0.759, due to the segregation of races (see Appendix Table A1). The reduction in entropy due to racial segregation is 0.219 or 22.43% of the maximum.

Taking the 5 SMSA's individually, the maximum racial entropy is largest in San Francisco and Oakland, the two SMSA's with the smallest fractions of white households. The measures of segregation are also largest in these two SMSA's, 25.16% and 23.22% respectively. The least segregated SMSA is clearly Santa Rosa, but it is also the one with the smallest non white population.

The table also presents similar information for the central city and suburban rings of each SMSA. These entries must be interpreted somewhat judiciously since the maximum possible entropy is conditional upon the racial composition of only a part of each SMSA. The table indicates that the level of segregation within the suburbs of each SMSA is substantially lower than the level of segregation in the central cities. This indicates that minorities fortunate enough to reside in the suburbs are less segregated in those suburbs than minorities are segregated in central cities. It should be noted however, that the maximum possible entropy is as much as fifty percent higher in the central cities, reflecting the intense segregation of minorities into the central cities of these SMSA's.

The levels of entropy by racial and demographic grouping, and the interpretation of the segregation indices themselves, are dependent upon the prior classification of the underlying population into meaningful groups. If the groups are too finely divided, their spatial integration will be less remarkable (as when Danish-American and Norwegian-American households are observed to live in adjacent houses). If the groups are too aggregated, their spatial integration may be misleading (as when the increasing integration of blacks and Puerto Ricans in Spanish Harlem is reported as a decrease in the level of black-nonblack segregation). Accordingly, column 2 of Table 4 presents segregation indices computed at a higher level of aggregation, using the three racial groupings noted in Table 3.

A comparison of Columns 1 and 2 reveals that the index of segregation is increased significantly from 22.4% to 26.6% when the population is divided into three racial groups (blacks, hispanics, and others) rather than six. This difference reflects the relatively more integrated Asian and white communities. Spatial integration of Asians with whites (and also with the small population of native Americans and others) "counts" in the disaggregated analysis in the sense that it leads to a reduction in the measure of racial segregation. In the analysis reported in column 2, all non black, non hispanic households are considered together. The difference in the index of segregation reported for the Oakland SMSA is particularly striking.

Columns 3 and 4 present analogous information for 1980 on the segregation of households by demographic type within the region. In column 3 the comparison is based on the seven classifications of household type noted in Table 2. For the region as a whole, the maximum entropy is 1.485, which would be obtained if each census tract had a distribution of household types identical to that reported in the last column of Table 2. The maximum entropy by household type is a good bit larger than the racial entropy, reflecting in part the more equal classification of households into groups. For the region as a whole, segregation by household type reduces actual entropy to 1.363 (see Appendix Table A1), or by 8.19 percent. Thus, for the region as a whole racial segregation is about two and a half times more intense than segregation by demographic group. When the entropy measures are disaggregated by SMSA, the results are similar. The index of segregation varies from 2.9 percent in the Santa Rosa SMSA to 8.5 percent in the Oakland and San Francisco metropolitan areas. In contrast, the index of racial segregation varies from 8.7 percent in Santa Rosa to 23.2 percent in Oakland and 25.2 percent in San Francisco.

The table also indicates the level of segregation by household type within the central city and suburbs of each SMSA. In contrast to the results by race, there is no systematic difference in the maximum possible entropy between central cities and their surrounding rings. There seems, however, to be a slightly greater level of segregation within central cities than within suburbs (in at least four of the five SMSA's), but the differences are rather small. Spatial segregation by household type is far less intense than segregation by race, and differences between central cities and suburbs are far less pronounced.

The fourth column in the table presents the indices of residential segregation by household type in 1980 when households are classified into only four groups: families headed by married couples, single females, single males, and non-family households. Computed this way, the indices of spatial segregation are slightly larger. The overall segregation measure is 1.1 percentage points higher; for three of the five SMSA's the index is also higher, by 0.6 to 1.4 percentage points. The differences are rather small, however, and the base is also rather small.

For either of these groupings, the level of racial segregation is estimated to be about two and a half times as intense as the level of segregation by household type.

Table 5 uses the actual entropy levels reported in Appendix Table A1 to compare the conditional and unconditional entropies by race and household type for the various geographical components of the San Francisco Bay Area. Knowledge of the three unconditional entropies H(r), H(h), H(r,h), permits the average conditional entropies to be calculated from equation (20), as well as the expected mutual information, J(r,h):

$$(22)\quad J(r,h) = H(r) - H_h(r)$$
$$= H(h) - H_r(h)$$
$$= H(r) + H(h) - H(r,h)$$

It is clear from (22) that J(r,h) = 0 when the distributions of r and h are independent; otherwise J(r,h) > 0. As indicated in Appendix Table A1, the values of J(r,h), and hence the differences between the conditional and unconditional entropies, are quite small indeed for the region as a whole and for each of its subareas. It can be shown that the upper bound of J is the smaller of the two marginal entropies, but in fact, the values of J for this region are only about one tenth as large as the smaller marginal entropy. This indicates a substantial degree of independence in the spatial distribution of households by race and household type -- for the region as a whole and for its various components. Stated another way, incorporating prior knowledge of the spatial distribution of household types does not affect the expected level of racial segregation very much.

Table 5 indicates the proportionate change in the conditional and unconditional entropies for 1980. The entries in the table have a convenient interpretation. Suppose the spatial distribution of household types in the metropolitan region is governed by "economic forces." Under these circumstances, recognizing the known and prior distribution of household types explains only a small fraction of the observed segregation of households by race. Using the most disaggregated definitions of race and household type and for the region as a whole, only 8.3 percent of the racial segregation observed could be attributed to segregation by household type. For the central cities of San Francisco and Oakland, only about 10.5 percent of the racial segregation observed could be attributed to the segregation by household type arising from economic forces.

Alternatively, only about 4.6 percent of the spatial segregation of household types could be explained by the prior segregation of households by race. For the largest central cities of San Francisco and Oakland, the upper limit is less than 8 percent.

Using more aggregated groupings, the mutual information is even smaller and the proportionate changes in conditional and unconditional entropies are even less. These results are also reported in Table 5. Despite the differences arising from group definition, the more aggregated analysis confirms the principal results: *Only a small fraction of segregation by household type can be explained by a prior segregation of households by race. An even smaller fraction of segregation by race can be explained by economic forces leading to a clustering by demographic group.*

The analysis of 1980 census data in more aggregated race-household categories provides a confirmation that the findings do not depend upon the definitions of race or type of housing. Consideration of these groupings of households does, however, permit a direct comparison of levels of segregation during the decade of the 1970's. Exactly the same information is available for the same census tracts from the 1970 census: Counts of households by each of the three racial and four household types.[17]

Table 6 summarizes the identical analysis of segregation by race and household type conducted using 1970 census information for the same 1079 census tracts and for the same racial and household definitions. The index of racial segregation for the region as a whole is 28.30% as compared to 26.58% in 1980. During the decade of the 1970's, the level of racial segregation declined in four of the five metropolitan areas -- by 2 to 3 percentage points in the Oakland, San Francisco, and Napa SMSA's. Increases in the level of racial integration were most pronounced within the central city of Oakland, and within the suburbs of the San Francisco SMSA. Residential segregation by race increased in the San Jose metropolitan area, both in the central city and its suburbs.

A comparison of Tables 4 and 6 also indicates that the residential segregation of households of different types declined during the 1970's. For the area as a whole, the index of segregation declined from 11.61% in 1970 to 9.32%

---

[17] The only difference is in the suppression of data for small samples. In the 1970 census, household counts by race were suppressed if the number of households of that race was less than 5. The analysis underlying Tables 6 and A2 was undertaken by distributing the number of suppressed households in any census tract into household types according to the distribution of housing types by that race in the central city or suburban ring containing that census tract.

in 1980. This reduction was observed in each of the 5 SMSA's, and in nine of the ten central city-suburban subareas. The only exception is in the central city of Oakland, where the segregation index increased very slightly, from 8.17% to 8.37%.

Finally, a comparison of the proportionate differences in conditional and unconditional entropies in 1970 with 1980 is obtained by comparing the results in Tables 5 and 6. For the entire region, the difference proportionate in racial entropy went from 4.02% to 4.62% and the difference by household type went from 2.08% to 1.82%.

The evidence is equally compelling that the spatial distributions of households by race and household type were independent in 1970 as well as in 1980. The level of mutual information is quite small relative to the joint entropy, and the difference between the conditional and unconditional entropies by race and by household type are very small indeed. For all the entries for 1970 and 1980 reported in Tables 5 and 6 the proportionate difference is on the order of 4% by race and 2% by household type.[18]

The socioeconomic forces which lead to spatial clustering of different types of households explain practically none of the spatial segregation of the races in 1980. They explained practically none in 1970, either.

---

[18] As one further check on the importance of group definition in reaching these conclusions, we conducted the entire analysis using two groups (black-nonblack) and two household types (female headed-all others) for 1970 and 1980. The qualitative results were the same as those reported in the text for the more disaggregated groupings.

VI. REFERENCES

Allison, Paul D., "Measures of Inequality," *American Sociological Review*, Vol 43, December 1978, pp 865-880.

Cortes, Charles F., *et al*, "Further Considerations of the Methodological Analysis of Segregation Indices," *American Sociological Review*, Vol 41, August 1976, pp 630-637.

Duncan, Otis Dudley and Beverly Duncan, "A Methodological Analysis of Segregation Indexes," *Amperian Sociological Review*, Vol 20, April 1955, pp 210-217.

Green, Gordon and Edward Welniak, *Changing Family Composition and Income Differential*, U.S. Department of Commerce, Bureau of the Census, Special Demographic Analysis CDS-80-7, August 1982.

Hwang, Sean-Shong, *et al*, "The Effects of Race and Socioeconomic Status on Residential Segregation in Texas, 1970-1980," *Social Forces*, Vol 63, #3, March 1985, pp 732-747.

Lichter, Daniel T., "Racial Concentration and Segregation across U.S. Counties, 1950-1980," *Demography*, Vol 22, #4, 1985, pp 603-609.

Lieberson, Stanley and Donna K. Carter, "A Model for inferring the Voluntary and Involuntary Causes of Residential Segregation," *Demography*, Vol 19, #4, November 1982, pp 511-526.

Massey, Douglas S., "Residential Segregation of Spanish Americans in United States Urban Areas," *Demography*, Vol 16, #4, November 1979, pp 553-564.

Massey, Douglas S. and Brooks Bitterman, "Explaining the Paradox of Puerto Rican Segregation," *Social Forces*, Vol 64, #2, December 1985, pp 306-331.

Morgan, Barrie S. and John Norbury, "Some Further Observations on the Index of Residential Differentiation," *Demography*, Vol 18, #2, May 1981, pp 251-256.

Sakoda, James, "A Generalized Index of Dissimilarity," *Demography*, Vol 18, #2, May 1981, pp 245-250.

Schnare, Ann B., "Trends in Residential Segregation by Race: 1960-1970," *Journal of Urban Economics*, Vol 7, #3, May 1980; pp 293-301.

Taeuber, Karl E. and Alma F. Taeuber, *Negroes in Cities*, Chicago: Aldine, 1965.

Taeuber, Karl E. and Alma F. Taeuber, "A Practitioner's Perspective on the Index of Dissimilarity," *American Sociological Review*, Vol 41, pp 884-889.

Theil, Henri, *Statistical Decomposition Analysis*, Amsterdam: North Holland Publishing Co., 1972.

White, Michael J.,"The Measurement of Spatial Segregation," *American Journal of Sociology*, Vol 88, #5, March 1983, pp 1008-1018.

White, Michael J., "Reply to Mitra," *American Journal of Sociology*, Vol 90, #1, July 1984, pp 189-191.

White, Michael J., "Segregation and Diversity Measures in Population Distribution," *Population Index*, Vol 52, #2, Summer 1986, pp 198-221.

Yinger, John, "Measuring Racial Discrimination with Fair Housing Audits: Caught in the Act," *American Economic Review*, Vol 76, #5, December 1986, pp 881-893.

Zoloth, Barbara S., "Alternative Measures of School Segregation, *Land Economics*, Vol 52, #3, August 1976, pp 278-298.

FIGURE 1

Census Tracts and Metropolitan Areas
San Francisco Bay Area

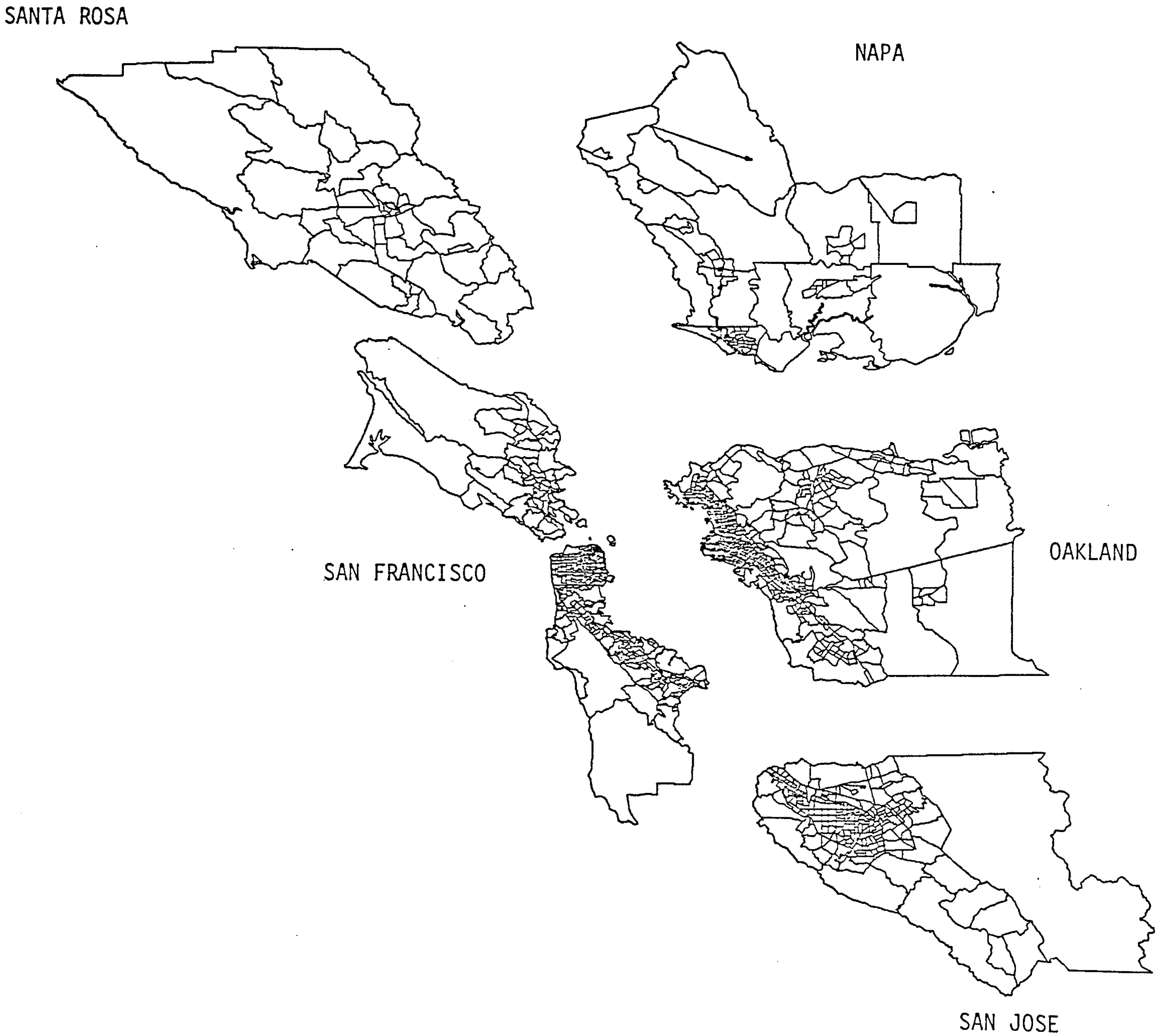

TABLE 1

Race and Hispanic Origin, San Francisco Bay Area, 1970-80
(Individuals)

Race	1980					1970	
	Reported Origin			Revised			
	Hispanic	Non-Hispanic	Total	Total	Percent	Total	Percent
Hispanic	--	--	--	632,650	12.2%	587,503	12.8%
Black	7,474	458,800	466,274	458,800	8.9%	365,893	7.9%
Asian	0	462,890	462,890	462,890	8.9%		
White	334,255	3,605,829	3,940,084	3,605,829	69.6%	3,641,962*	79.3%*
Native Amer.	32,025	5,162	37,187	5,162	0.1%		
Other	258,896	14,453	277,349	14,453	0.3%		
Total	632,650	4,547,134	5,179,784	5,179,784	100.0%	4,595,358	100.0%

Note:* Non-black, non-hispanic individuals.

Source: Calculations by authors from 1980 Census STF3, and from 1970 Census of Housing, Fourth Count Summary.

TABLE 2

Household Type by Race, Bay Area, 1980
Forty-two Race-Household Categories
(Households)

Household Type	White Number	White Prop.	Black Number	Black Prop.	Native Number	Native Prop.	Asian Number	Asian Prop.	Hispanic Number	Hispanic Prop.	Other Number	Other Prop.	Total Number	Total Prop.
Family														
Married Couple														
With Children	331,493	0.16	28,834	0.01	2,314	0.00	44,208	0.02	70,110	0.04	31,657	0.02	508,616	0.26
No Children	427,324	0.22	25,863	0.01	1,582	0.00	29,979	0.02	36,530	0.02	20,243	0.01	541,521	0.27
Male Householder (Unmarried)														
With Children	16,900	0.01	5,612	0.00	2,623	0.00	3,592	0.00	3,574	0.00	3,757	0.00	36,058	0.01
No Children	19,821	0.01	3,410	0.00	161	0.00	3,505	0.00	4,209	0.00	1,715	0.00	32,821	0.02
Female Householder (Unmarried)														
With Children	66,317	0.03	27,706	0.01	1,116	0.00	4,628	0.00	17,346	0.01	6,861	0.00	123,974	0.06
No Children	45,004	0.02	9,679	0.00	397	0.00	4,914	0.00	7,937	0.00	2,973	0.00	70,904	0.04
Non-Family	524,036	0.27	53,845	0.03	2,786	0.00	29,160	0.01	41,741	0.02	21,184	0.01	672,752	0.34
Total	1,430,895		154,949		10,979		119,986		181,447		88,390		1,986,646	
Proportion		0.72		0.08		0.01		0.06		0.09		0.05		1.00

Source: Calculations by authors from 1980 census, STF3.

Note: Male and female "householder" classes may include other adults.

TABLE 3

Household Type by Race, Bay Area, 1970 and 1980
Twelve Race-Household Categories
(Households)

Household Type	All Other Number	All Other Prop.	Black Number	Black Prop.	Hispanic Number	Hispanic Prop.	Total Number	Total Prop.
1980								
Family								
Married Couple	888,173	0.45	55,238	0.03	106,640	0.05	1,050,051	0.53
Male Householder (Unmarried)	42,262	0.02	6,250	0.00	7,783	0.00	56,295	0.03
Female Householder (Unmarried)	133,712	0.07	35,585	0.02	25,283	0.01	194,580	0.10
Non-Family	577,148	0.29	53,790	0.03	41,741	0.02	672,679	0.34
Total	1,641,294	0.83	150,864	0.08	181,447	0.09	1,973,605	1.00
1970								
Family								
Married Couple	813,205	0.53	53,351	0.03	115,058	0.07	981.614	0.64
Male Householder (Unmarried)	25,704	0.02	2,993	0.00	3,882	0.00	32,579	0.02
Female Householder (Unmarried)	92,329	0.06	22,750	0.01	15,319	0.01	130,398	0.08
Non-Family	343,799	0.22	29,790	0.02	22,777	0.01	396,366	0.26
Total	1,275,037	0.83	108,884	0.07	157,036	0.10	1,540,957	1.00

Source: Calculated by the authors from 1980 census, STF3, and from 1970 Census of Housing, Fourth Count Summary.

TABLE 4

Indices of Residential Segregation by Race and Household Type
San Francisco Bay Area, 1980

	Race		Household Type	
	Six Groups	Three Groups	Seven Groups	Four Groups
Entire Region	22.43%	26.58%	8.19%	9.32%
SMSA's				
Oakland	25.16	31.43	8.49	9.87
San Francisco	23.22	24.52	8.53	9.90
San Jose	12.06	14.38	6.36	6.97
Santa Rosa	8.73	4.21	2.94	2.80
Napa	13.25	13.67	5.16	4.86
Central City/Suburbs				
Oakland				
Central City	21.28	23.69	7.54	8.37
Suburbs	18.79	23.26	7.40	8.45
San Francisco				
Central City	20.32	25.34	8.33	9.22
Suburbs	22.66	20.13	5.50	6.27
San Jose				
Central City	11.80	13.87	5.93	6.39
Suburbs	9.42	11.06	6.12	7.00
Santa Rosa				
Central City	12.68	4.92	4.68	4.69
Suburbs	4.64	3.39	1.53	1.29
Napa				
Central City	12.45	13.13	5.56	5.51
Suburbs	8.29	6.67	3.48	2.20

Note: Table entries are (K-H)/K where K is the maximum entropy possible in each geographical region and H is the actual entropy computed from the census tracts in that region. Values of H appear in Appendix Table A1.

TABLE 5

Proportionate Differences in Conditional and Unconditional Entropies
San Francisco Bay Area, 1980

Race Groups / Household Groups	Race: Six / Seven	Race: Three / Four	Household Type: Six / Seven	Household Type: Three / Four
Entire Region	8.30%	4.06%	4.62%	1.82%
SMSA's				
Oakland	8.34	4.18	4.65	2.03
San Francisco	9.82	4.98	5.72	2.13
San Jose	6.76	3.71	3.95	1.71
Santa Rosa	7.56	3.46	2.46	0.83
Napa	7.22	3.59	3.98	1.72
Central City/Suburbs				
Oakland				
Central City	10.52	5.26	7.20	3.44
Suburbs	7.55	3.56	3.94	1.54
San Francisco				
Central City	10.57	5.33	7.74	2.65
Suburbs	8.77	4.15	3.99	1.49
San Jose				
Central City	6.23	3.37	4.03	1.81
Suburbs	7.46	3.77	3.80	1.41
Santa Rosa				
Central City	8.02	3.65	2.73	0.84
Suburbs	7.24	3.33	2.28	0.82
Napa				
Central City	7.72	4.27	4.39	2.18
Suburbs	5.41	2.06	2.72	0.92

Note: For columns 1 and 2, table entries are $[H(r)-H_h(r)]/H(r)=J(r,h)/H(r)$.

For columns 3 and 4, table entries are $[H(h)-H_r(h)]/H(h)=J(r,h)/H(h)$.

TABLE 6

Indices of Residential Segregation by Race and Household Type and Proportionate Differences in Conditional and Unconditional Entropies
San Francisco Bay Area, 1970

	Segregation Index		Proportionate Difference in Conditional and Unconditional Entropies	
	Race Three Groups	Household Type Four Groups	Race	Household Type
Entire Region	28.30%	11.61%	4.02%	2.08%
SMSA's				
Oakland	33.59	12.09	3.59	1.98
San Francisco	27.06	12.09	4.30	2.07
San Jose	12.56	8.80	3.77	2.12
Santa Rosa	4.71	3.44	3.77	2.12
Napa	16.42	5.71	3.14	1.92
Central City/Suburbs				
Oakland				
Central City	31.34	8.17	4.44	2.68
Suburbs	28.39	10.37	3.27	1.72
San Francisco				
Central City	26.19	10.01	4.62	3.12
Suburbs	23.85	7.33	3.36	1.39
San Jose				
Central City	12.94	10.07	4.31	2.85
Suburbs	9.40	7.40	3.87	1.82
Santa Rosa				
Central City	4.14	4.93	3.40	0.92
Suburbs	4.77	2.10	2.59	0.83
Napa				
Central City	16.68	6.48	3.70	2.05
Suburbs	7.54	2.50	2.32	1.15

Note: See Tables 4 and 5 for definitions.

APPENDIX TABLE A1

Actual Entropy Levels by Race and Household Type
San Francisco Bay Area, 1980

	Race, H(r)		Household, H(h)		Joint, H(r,h)	
Race Groups Household Groups	Six	Three	Seven	Four	Six Seven	Three Four
Entire Region	0.759	0.419	1.363	0.936	2.059	1.338
SMSA's						
Oakland	0.767	0.455	1.376	0.937	2.079	1.373
San Francisco	0.764	0.402	1.312	0.940	2.001	1.322
San Jose	0.814	0.431	1.394	0.932	2.153	1.347
Santa Rosa	0.463	0.231	1.425	0.964	1.853	1.187
Napa	0.762	0.446	1.381	0.903	2.088	1.333
Central City/ Suburbs						
Oakland						
Central City	0.960	0.684	1.403	1.047	2.262	1.695
Suburbs	0.715	0.393	1.369	0.907	2.030	1.286
San Francisco						
Central City	0.908	0.469	1.240	0.942	2.052	1.386
Suburbs	0.627	0.337	1.380	0.937	1.952	1.386
San Jose						
Central City	0.915	0.505	1.416	0.939	2.274	1.427
Suburbs	0.697	0.345	1.369	0.924	2.014	1.256
Santa Rose						
Central City	0.474	0.219	1.391	0.955	1.827	1.166
Suburbs	0.456	0.240	1.449	0.970	1.872	1.202
Napa						
Central City	0.790	0.468	1.388	0.916	2.117	1.364
Suburbs	0.684	0.388	1.362	0.866	2.009	1.246

Appendix Table A2

Actual Entropy Levels by Race and Household Type
San Francisco Bay Area, 1970

	Race, H(r)	Household, H(h)	Joint, H(r,h)
Race Groups Household Groups	Three	Four	Three Four
Entire Region	0.423	0.819	1.225
SMSA's			
Oakland	0.446	0.810	1.240
San Francisco	0.419	0.868	1.269
San Jose	0.424	0.753	1.161
Santa Rosa	0.255	0.855	1.102
Napa	0.423	0.783	1.191
Central/City/Suburbs			
Oakland			
Central City	0.585	0.970	1.529
Suburbs	0.398	0.754	1.139
San Francisco			
Central City	0.498	0.932	1.407
Suburbs	0.327	0.793	1.109
San Jose			
Central City	0.487	0.737	1.203
Suburbs	0.362	0.770	1.118
Santa Rose			
Central City	0.235	0.868	1.095
Suburbs	0.270	0.845	1.108
Napa			
Central City	0.433	0.782	1.199
Suburbs	0.388	0.785	1.164