

# Explaining changes in US residential segregation through patterns of population change

Received: 21 August 2023

Accepted: 11 January 2024

Published online: 09 February 2024

 Check for updates

Benjamin Elbers  

While overall residential segregation in US cities has declined in the past 30 years—especially between the Black and white populations—relatively little is known about the patterns of population change that caused these changes. Here I investigate changes in racial residential segregation in US cities between 1990 and 2020, and focus on answering two key questions: first, which racial groups are driving changes in segregation, and second, where in the metropolitan area these changes are produced. By connecting segregation to population changes in specific locations, this study highlights how segregation is connected to core urban problems and processes such as suburbanization, gentrification and other forms of spatial inequality in cities. To answer these questions, the paper develops a flexible decomposition method that allows us to draw a direct link from changes in the distribution of racial groups—brought about by residential mobility, births and deaths—to changes in racial segregation. This demographic approach to explaining segregation change quantifies how much Black suburbanization, ‘white flight’ and other group-specific population changes contribute to changes in segregation. The results show that almost all decreases in segregation were produced by the suburbanization of Black, Hispanic and Asian people, as well as the population growth of these groups in the formerly majority white areas of central cities. Changes in the distribution of the white population, instead, are mostly associated with increasing segregation. Hence, segregation has decreased despite the majority group’s efforts to resegregate themselves. The upshot is that most metropolitan areas are shaped by simultaneous and ongoing desegregation and resegregation (‘racialized reshuffling’), and that this process is not restricted to the difference between central cities and suburbs. For the white population, both growth and decline in suburban places contributed toward increasing segregation, which indicates that there is substantial resorting happening also within suburbs. The results of this study suggest that focusing only on the integration of minority populations, without also limiting the resegregation of the white population, will be ineffective in bringing about substantial reductions in segregation.

Changes in US residential segregation in the past decades are well documented<sup>1–5</sup>. Hispanic–white and Asian–white segregation are usually reported as relatively stable, while the segregation of the Black population has declined substantially (Fig. 1a). Concurrently with the decreases in segregation between 1990 and 2020, the US population has become increasingly diverse, with the relative shares of the Hispanic and Asian populations doubling (Fig. 1b).

There are two broad theoretical perspectives on changing segregation<sup>2</sup>. Spatial assimilation theory argues that that minority groups become residentially integrated as they are integrated along other dimensions, such as income or education. Members of minority groups therefore ‘move up’ spatially when they move up in the distribution of other status characteristics. Place stratification theory, on the other hand, focuses on the considerable barriers that hinder the integration of minorities, such as housing market discrimination and diverging preferences of racial groups. Empirical tests of these theories have not been conclusive: the declines in segregation at the aggregate level suggest the applicability of spatial assimilation theory<sup>6</sup>, although empirical studies point out that the Black population still faces considerable barriers to convert upward mobility into improvements in residence<sup>7</sup>. In support of place stratification, many studies find continued housing market discrimination<sup>8</sup>, ‘white flight’<sup>9</sup> and a clear preference of white people to live in majority-white neighborhoods<sup>10</sup>.

These findings present a puzzle: the continuing disadvantage of the Black population suggests stable segregation, while the comparatively fewer barriers that are faced by the Hispanic and Asian populations imply declining segregation. However, as Fig. 1 shows, although Black–white segregation remains high, it saw a considerable decrease since 1990, unlike Hispanic–white and Asian–white segregation. To resolve this puzzle, it is necessary to understand how micro-level processes that determine neighborhood location aggregate up to changes in segregation. This mirrors the call by ref. 11 to focus segregation research on the ‘racialized reshuffling’ that is brought about by residential mobility/immobility, and that ultimately produces persisting, not declining, segregation.

In this Article, I contribute empirically to this research agenda by directly linking changes in the distributions of racial groups to changes in residential segregation. The study takes a demographic approach to explaining segregation change, which ultimately must be related to different rates of natural population change in different neighborhoods (brought about by fertility and mortality), and residential mobility patterns. While we know how segregation changed in the aggregate, we know surprisingly little about which racial groups are driving the changes in segregation. For instance, white people have been known to practice ‘white flight’ (a process that increases segregation), but also gentrification (a process that reduces segregation). Similarly, minority groups may want to self-segregate (increasing segregation), or move into more privileged white areas in search of new opportunities and better schools. This study will disentangle the aggregate process of segregation decline into contributions for each racial group, and thereby answer the question of how much different processes of population change contribute to changes in segregation.

The demographic perspective that is adopted in this paper also highlights the fact that a segregation index can obscure countervailing trends. For instance, zero change in the segregation index can be accounted for either by no population movement or by offsetting movements. For instance, white flight and Black suburbanization could happen simultaneously in a metropolitan area, canceling each other out. A similar offsetting argument can be made for groups that are composed both of newly arrived immigrants and native-born residents. While recent immigrants cluster in racially homogeneous neighborhoods after arriving to the United States, long-term residents may be able to ‘move up’ in the spatial hierarchy in accordance with the spatial assimilation model<sup>12</sup>. This means that, within the same racial group, both resegregation and desegregation may be happening

simultaneously. While the net change in segregation that these dynamics produce may be zero, this does not indicate the absence of change, supporting the thesis that US metropolitan areas are shaped by ‘racialized reshuffling’.

The paper therefore advances the literature both empirically and conceptually: the empirical study applies a decomposition approach to understand which population changes contribute to changing segregation. At the conceptual level, the study shifts the emphasis from the aggregate segregation measure to the mechanical demographic processes that shape changes in the population distribution and, consequently, segregation. The population changes of different racial groups may be considered as ‘proximate causes’ for segregation change, compared to the remote causes that affect population movements (or their absence). Empirically, the results show that almost all decreases in segregation were produced by the suburbanization of Black, Hispanic and Asian people, as well as the population growth of these groups in the formerly majority white areas of central cities. The fact that the white population mostly contributed to increasing segregation lends strong support to the continuing importance of place stratification. If large-scale reductions in segregation are to be achieved, policy cannot focus solely on the integration of minorities, but needs to focus as well on avoiding the resegregation of the white population.

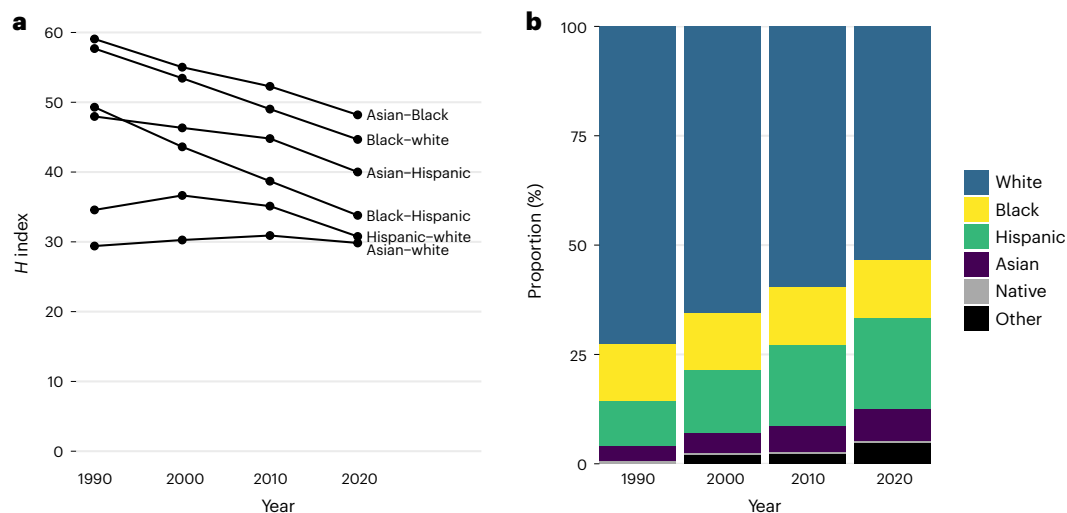
## Prior research

The number of studies that have directly connected residential mobility to segregation change is small. This is probably due to both data constraints and the absence of a methodology to decompose changes in segregation. Two exceptions to this are the studies by ref. 13 and ref. 14. The latter distinguish between three kinds of segregation trajectories at the neighborhood level: durable segregation, racial change and durable integration. They argue that segregation should not be studied in the cross-section, because this gives only a snapshot of what is ultimately a segregation process. The snapshot might capture an intermediate, more integrated picture, although the trajectory points toward resegregation. For instance, a neighborhood where the Black population has started to move in and the white population has started to move out might look integrated when observed in the middle of this process. However, ref. 14 does not connect their findings to aggregate segregation measures.

The approach that ref. 13 pursues is similar to the research design of the current paper. The authors show that, between 1990 and 2010, between-county migration served to decrease segregation, while natural population change would have increased segregation. With this approach, the authors go far beyond usual studies on segregation, but the paper is not without limitations. First, importantly, due to data constraints, the smallest unit of analysis is only the county, which excludes a large—and arguably the most relevant—amount of residential segregation, which occurs at the Census tract and block levels<sup>15</sup>. Second, while the authors compute counterfactual segregation indices, their results are not full decompositions in the sense that the sum of all factors reflects the total change in segregation.

## Contribution of this study

To answer the question of how much different racial groups contribute to changes in segregation, the paper develops a novel decomposition method. The change in segregation between 1990 and 2020 is decomposed into several factors that describe population changes for each of the racial groups. The decomposition is based on counterfactual scenarios: a factor such as ‘Black population growth in suburban places’ is calculated by keeping the block-level population counts constant, and only varying the Black population according to the observed growth in suburban places. When this is done for many factors, the sum of all factors may deviate strongly from the actual observed change in total segregation. Hence, the importance of each factor is estimated using a Shapley decomposition<sup>16</sup>, which ensures that the sum of all factors



**Fig. 1 | Pairwise segregation indices and population composition by race/ethnicity. a**, Population-weighted averages of segregation in a consistent set of 224 US metropolitan areas. Segregation is calculated using Theil's Information Index  $H$  on block-level Decennial Census data. **b**, The population composition by race and ethnicity in the same set of areas. 'White', 'Black' and 'Asian' refer to the

single-race, non-Hispanic white, Black or African American, and Asian or Pacific Islander population, respectively. 'Hispanic' refers to the Hispanic or Latino population of any racial group. 'Native' refers to the single-race, non-Hispanic American Indian and Alaska Native population. 'Other' comprises individuals that identify as any other race or as two or more races.

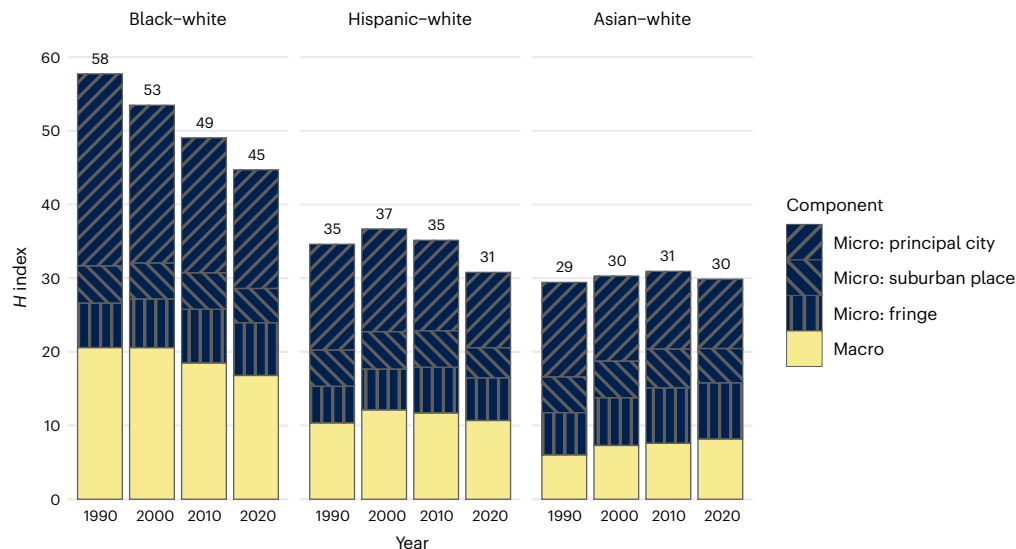
equals the observed change in segregation. Because some decompositions have too many factors to make a closed-form computation of the Shapley decomposition possible, a simulation approach is used to approximate the solution. While the Shapley decomposition has been used in segregation research before<sup>17,18</sup>, this is the first study to use the Shapley decomposition to decompose changes in segregation into contributions for each racial group, and to make use of a simulation approach that allows for the estimation of a large number of factors compared to earlier approaches.

This study builds on earlier results that have shown the advantages of decomposable segregation indices. This study uses the entropy-based segregation index  $H$ , ranging from 0 (absence of segregation) to 100 (complete segregation). Segregation is calculated for each US metropolitan area and Census year using block-level data, which is the smallest spatial unit that the Census provides. To summarize the segregation patterns across the 224 metropolitan areas, a population-weighted average is used.

In line with earlier research<sup>5,19</sup>, segregation is decomposed into macro and micro components. Macro segregation refers to place-based segregation, that is segregation between the cities and suburbs of a metropolitan area, as well as incorporated communities in the suburbs and unincorporated areas outside of the principal city ('fringe' areas). Hence, block-level segregation in a metropolitan area is 'total segregation', segregation between places is 'macro segregation' and segregation within places is 'micro segregation'. The difference between total and micro segregation is that for total segregation the reference distribution is the racial group distribution of the entire metropolitan area, while for micro segregation the reference distribution is the racial group distribution of the respective place. The three quantities are related through the simple identity macro + micro = total segregation. Theoretically, it is possible that segregation is entirely between places or entirely within places. In the first case, places are internally racially homogeneous, but differ in their racial composition from each other (for instance, central cities could be all-Black, while suburbs could be all-white). In the second case, places are internally segregated, but all contain the same proportions of each racial group (for instance, each place contains some neighborhoods that are all-Black, and some that are all-white).

Changes in the population distribution are produced by two distinct processes: natural population changes (the sum of births and deaths) and net residential mobility (including international migration). Breakdowns of fertility and mortality rates by age and racial groups are produced by the Centers for Disease Control and Prevention, but only at the county level, and not for smaller geographies such as tracts or blocks. It is therefore not possible to break down the decomposition results further into components that separate natural population change from mobility. It is, however, likely that residential mobility is the most important of these factors. Many population increases are too large to be explained by natural population change alone. For instance, in blocks that were at least 90% white in 1990, the Black and Hispanic populations each increased more than tenfold between 1990 and 2020, which reflects an annual growth rate of over 8%. These increases produced strong declines in segregation, but—even assuming very high fertility and low mortality rates—are unlikely to be achieved without substantial in-mobility (for comparison, ref. 20 finds that the rate of natural increase for the entire United States between 2000 and 2006 was 0.6% per year; for Hispanic people, the total rate of natural increase was estimated to be about 2.3% annually between 1990 and 2005). The county-level analysis produced by ref. 13 also shows that changes due to natural processes are fairly small compared to changes produced by migration. Nonetheless, given the different age structures of the racial groups, and the likely possibility that fertility and mortality are related to neighborhood composition<sup>21,22</sup>, natural population changes may play a larger role in metropolitan areas where the age structure, fertility rate and mortality rate diverge strongly between racial groups and neighborhoods.

This study finds that most metropolitan areas are shaped by simultaneous and ongoing desegregation and resegregation. For macro segregation, Black integration and white resegregation offset each other to a large extent, leading to only small declines in macro segregation. This shows that the absence of change in the aggregate segregation index does not reflect the absence of neighborhood change. For micro segregation, declines in segregation produced by Black integration are not offset by white resegregation, leading to declines in total metropolitan area segregation. Hispanic-white and Asian-white segregation are similarly shaped by both resegregation and desegregation, even within the same racial groups.



**Fig. 2 | Population-weighted average micro and macro segregation across 224 metropolitan areas, 1990–2020.** Black–white (left), Hispanic–white (middle) and Asian–white (right) segregation for each Census year between 1990 and 2020. Total segregation is broken down into two broad components: macro

and micro segregation. Micro segregation is further broken down into separate components for the principal city, suburban places and fringe areas. Total segregation is printed in small font at the top of the bar.

Decomposition results are shown for Black–white, Hispanic–white and Asian–white segregation. Equivalent results for Black–Hispanic, Asian–Black and Asian–Hispanic segregation are contained in Supplementary Information.

## Results

### Micro and macro segregation trends

To assess overall trends, Fig. 2 reports population-weighted averages across the 224 metropolitan areas. Between 1990 and 2020, Black–white segregation decreased from 58 to 45. Given that the scale extends from 0 to 100, this is still an extremely high level of average segregation. Hispanic–white and Asian–white segregation have remained at about the same level throughout this period, but are at much lower levels compared to Black–white segregation.

Throughout this period, micro segregation is quantitatively more important than macro segregation. Hence, the majority of segregation is not due to differences in the distribution of racial groups between cities and suburbs, but due to segregation that occurs within cities, suburbs and fringe areas. Most of the micro segregation originates in the principal cities of each metropolitan area. This component also declined by the largest amount.

### Decomposing changes in segregation

Figure 2 shows a cross-sectional decomposition of total segregation into micro and macro components. Figure 3, instead, shows the results of decomposing changes in segregation between 1990 and 2020. The change in segregation is decomposed into four components: the contribution to macro segregation change for each of the two racial groups, and the contribution to micro segregation change for each of the two racial groups.

Across all metropolitan areas, Black–white segregation decreased on average by about 12 points. Figure 3 shows that a large majority of this decrease is due to changes in the population distribution of the Black population, which contributed –9 points to macro segregation change, and –8 points to micro segregation change. The changing population distribution of the white population acted in the opposite direction, contributing 5 points to increasing macro segregation. In total, the sum of these four factors led to a segregation decrease of 12 points. The counteracting forces of the white and Black populations are only revealed through this decomposition, which shows that it is the

Black population only that contributed toward desegregation, while the white population contributed to resegregation.

The decomposition results are similar for changes in Hispanic–white segregation, although the overall magnitude of the changes is smaller. There is one notable difference, however: for Hispanic–white segregation, population changes of the white population contribute to resegregation for both micro and macro segregation, while for Black–white segregation these movements contribute only to increases in macro segregation. This results in a much smaller decline in total Hispanic–white segregation compared to Black–white segregation.

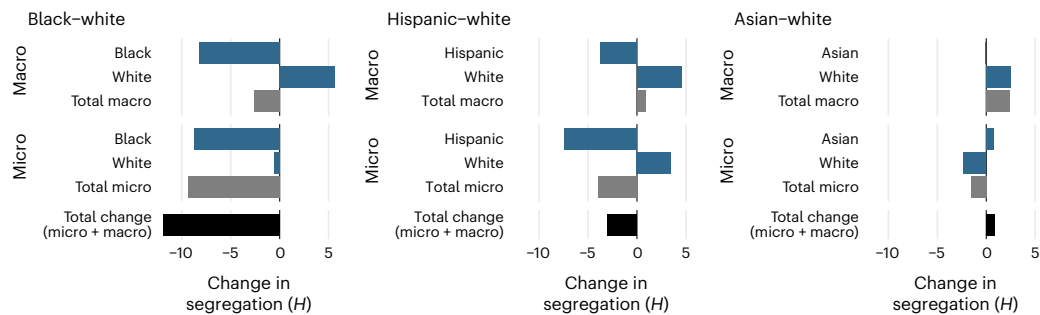
Asian–white segregation patterns deviate notably from both Black–white and Hispanic–white segregation. The magnitudes of change are much smaller here. Notably, while white population changes contribute to macro segregation, Asian population shifts in total are not associated with macro segregation changes.

Figure 3 reveals previously hidden patterns of segregation change. In broad strokes, mobility patterns and natural population changes of the Black and Hispanic populations contributed most to declines in segregation for these groups, while the net contribution of white people points to increases in segregation. For Black–white segregation, the results are especially stark: almost the entire decline in segregation is driven by the Black population, and, were it not for the increasing macro segregation of white people, the decline would have been even more pronounced. Generally, the white contributions run in opposite direction to those of the minority groups. Most metropolitan areas are shaped by simultaneous desegregation—produced mostly by minority groups—and resegregation—produced mostly by the white population.

### Detailed macro decomposition

The decompositions in Fig. 3 show some important patterns of segregation change. However, they do not address the specific spatial sources of change: is the white contribution toward increasing segregation due to suburbanization? Do the Black and Hispanic populations see growth in the suburbs as well, and thereby decrease segregation? Figure 4 shows a detailed decomposition of macro segregation into specific spatial patterns of change, where each component is characterized by a racial group, an increase or decrease, and a location. This decomposition reveals how large-scale geographical resorting between cities, suburbs and the metropolitan fringe contributes to changes in segregation.





**Fig. 3 | Decomposition of changes in segregation, 1990–2020.** Decomposition results for Black–white (left), Hispanic–white (middle) and Asian–white (right) changes in segregation between 1990 and 2020. These components are the population-weighted averages across separate decompositions for each of the 224 metropolitan areas. The simulation standard errors range from 0.0002 to 0.006, and are therefore too small to be visually depicted. For each combination of racial groups, the change in segregation is decomposed into four components: the contribution of white population changes to micro and macro segregation

The large contributions of the Black population toward decreasing segregation are indeed due to population growth in the suburbs and fringe areas. These effects are partially offset by population changes of the white population, which tend to increase segregation. The most important of these changes are white population declines in principal cities, which is indicative of continuing white suburbanization. However, the results also show that there is ongoing sorting within suburbs: for white people, both growth and decline in suburban places contributed toward increasing segregation. This shows that the white population declined in more racially integrated suburbs, while it increased in more racially homogeneous suburbs. It is possible that some of these patterns are reactive: Black suburbanization may have prompted white people to leave suburban places. If this interpretation is true, white flight would not be restricted to principal cities.

It is also interesting to consider effects where we might have expected a contribution, but where the contribution is small. This is most apparent for the effect of Black decline in principal cities. If there is Black suburbanization, would we not expect that this leads to declining segregation in the principal city? One possible explanation for this is that the fringe and suburban effects reflect a reshuffling of population within suburban and fringe areas. For instance, if Black people move from majority Black suburbs to majority white suburbs, this reduces segregation in the suburbs, but leaves segregation in the principal city unchanged. Another possible explanation is cross-metropolitan area mobility patterns, where the segregation-reducing effects in the suburbs are brought about by in-mobility from other metropolitan or rural areas. Lastly, some of the change may also be attributable to natural population growth in the suburbs.

For Hispanic–white segregation, the overall picture is similar to Black–white segregation, with one important difference: the Hispanic population contributes to both decreases and increases in macro segregation. The most important of the factors that increase segregation is Hispanic growth in principal cities. For Asian–white segregation, the effects are smaller, but the patterns are similar. As in the Hispanic–white case, a large segregation-increasing effect is due to Asian population growth in principal cities.

The countervailing trends for both Hispanic people and Asian people may point to the role of immigration. Newly arrived immigrants often settle in more ethnically homogeneous neighborhoods, where family and friendship networks are present. This group may therefore be distinct from the group that migrates to the suburbs: probably the latter are native-born or later-generation immigrants, who have acquired the necessary resources. Given that the Hispanic

change, respectively, and the contribution of Black/Hispanic/Asian population changes to micro and macro segregation change, respectively. The sum of the white macro and Black/Hispanic/Asian macro changes add to the ‘Total macro’ component (and similarly for the micro component). The sum of the ‘Total macro’ change and the ‘Total micro’ change equal the ‘Total change’ depicted at the bottom. The ‘Total Macro’ component is further decomposed in Fig. 4 below, and the ‘Total micro’ component is further decomposed in Fig. 5 below.

and Asian populations skew notably younger compared to the white population, differences in natural population growth may also play a role. Importantly, we again find simultaneous desegregation and resegregation—this time even within racial groups.

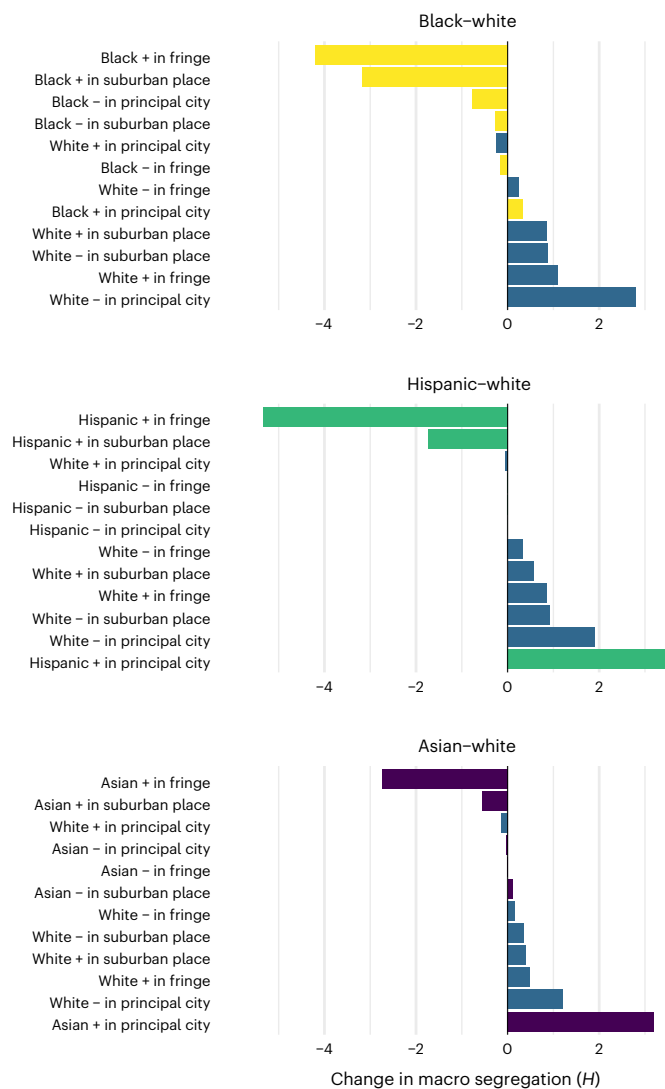
### Detailed micro decomposition

Figure 5 shows results for the detailed decomposition of micro segregation change. To understand where the important changes in micro segregation are occurring, the decomposition distinguishes between changes within principal cities, suburbs and fringe areas. Each component is characterized by a racial group, an increase or decrease and a location (either a majority block or a ‘mixed’ block). Consistent with Fig. 2, most of the changes in micro segregation occur in principal cities. This decomposition reveals how small-scale changes in the population distribution at the neighborhood level within cities and suburbs contribute to changes in segregation.

For Black–white segregation, the largest factor affecting micro segregation across all three geographical subdivisions is Black population growth in majority white areas. One might have suspected that this is mostly a suburban and fringe effect, but the effect is the largest in principal cities. The decline in micro segregation that we observe is therefore mostly due to Black growth in majority white areas. Factors that increase segregation are smaller, but there are quite a few factors that partially offset the overall decline. These include the classic indicators of ‘white flight’ (white population declines in majority Black and mixed neighborhoods, and white population growth in majority white areas). Black population decline in majority white areas is the largest of the offsetting factors, pointing again to substantial heterogeneity: while some majority white neighborhoods experience minority population growth, others experience a decline.

The defining factor that led to the decrease in Hispanic–white micro segregation is Hispanic population growth in majority white areas. These effects are substantial not only in the principal city, but also in the suburbs. Similarly to Black–white segregation, factors that have partially offset these decreases are indicators of ‘white flight’, as well as Hispanic population decline in majority white areas.

Figure 3 shows that declines in Asian–white micro segregation are small, and due entirely to white population changes. The detailed decomposition reveals that there is, however, substantial resorting also for Asian people, but that these effects cancel each other out. Asian growth in majority white areas led to decreases in segregation, but the simultaneous growth of the Asian population in mixed neighborhoods and declines in majority white areas have offset these decreases.



**Fig. 4 | Detailed decomposition of changes in macro segregation, 1990–2020.** Black–white (top), Hispanic–white (middle) and Asian–white (bottom) macro segregation are decomposed separately. As before, the figure shows population-weighted averages. The simulation standard errors range from  $-0$  to  $0.004$ , and are therefore too small to be visually depicted. In each panel, the 12 bars sum up to the average change in macro segregation (‘Total macro’ change in Fig. 3). To interpret the individual factors, consider the example of ‘Black + in fringe’ for Black–white segregation, where the ‘+’ stands for population growth. To compute this factor, each fringe area is classified in terms of either Black population growth or decline. Then, the counterfactual is computed only for those areas where Black population growth occurred. The factor can then be interpreted as follows: if the Black population grew in those fringe areas where we observed growth between 1990 and 2020, net of other population changes, how would segregation have changed?

## Discussion

Why did residential segregation in the United States decrease between 1990 and 2020? The short answer is that minority groups contributed a large amount toward segregation decline, while the white population contributed a smaller amount toward segregation increase. This study has shown these countervailing trends for the first time, and shows that declines in segregation were achieved despite white resegregation. These findings lend support to both existing theories of segregation, but with nuances by both racial groups and spatial scale: over the course of the past 30 years, many members of minority groups moved toward integration, especially in the suburbs, supporting spatial assimilation theory. Simultaneously, the white population, as well as parts of

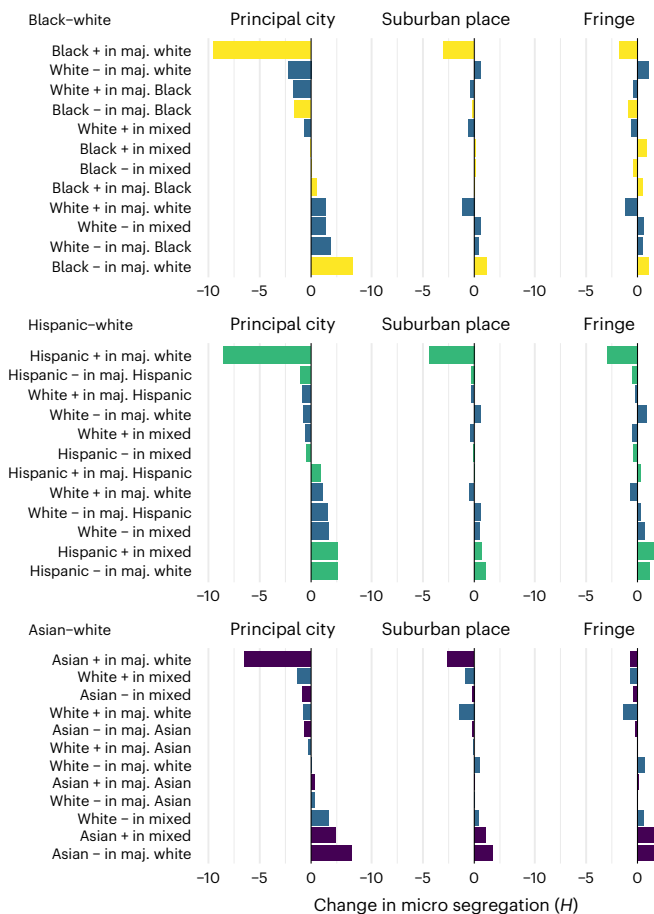
the Asian and Hispanic populations, resegregated, supporting place stratification theory. While the white population offset some of the declines in segregation, it is important to acknowledge the large trend in desegregation that has been brought about by Black suburbanization. Much of the literature on segregation change focuses on the fact that segregation persists, not investigating the factors that led to segregation decline<sup>41</sup>. To accelerate reductions in segregation, positive factors, such as the increasing presence of minorities in the suburbs, have to be strengthened, while offsetting factors, such as the resegregation of white people, need to be limited. Policies that aim to reduce segregation will be effective only if they are integrating all groups simultaneously.

Similar to earlier results<sup>5</sup>, this study has found that segregation declines are mostly produced by declining micro segregation, that is, the segregation within cities and suburbs. This means that the scale of segregation shifts to the macro level: As neighborhoods in cities and suburbs become more integrated, it matters more and more in which exact city or suburb within the metropolitan area a racial group resides. In terms of macro segregation, changes in the population distributions of minority populations are typically segregation reducing, while the white population typically contributes to increasing segregation. These countervailing trends produce a trend in macro segregation that is slightly decreasing (for Black–white) or slightly increasing (for Hispanic–white and Asian–white segregation). The results confirm the thesis of a constant ‘racialized reshuffling’ at the macro level, indicating that there was an enormous amount of change in the population distributions of the different racial groups—without, however, impacting segregation. Such offsetting trends show that one should not mistake the stability of any aggregate measure of segregation for the absence of change.

The relative stability of macro segregation contrasts with declining micro segregation (especially for Black–white segregation), which has mostly occurred in principal cities. Why did micro segregation decline by such a large amount, while macro segregation remained stable? The reason is that there was no substantial countervailing contribution of the white population. The Black and Hispanic populations contributed toward integration, and white people did not, in large numbers, contribute toward resegregation. In fact, the detailed decompositions show that white population changes matter less for micro segregation; the largest effects are all produced by the minority groups (except for Asian–white segregation).

Despite ongoing suburbanization, the principal cities of metropolitan areas continue to be central to the production and reduction of segregation. In 1990, segregation within principal cities alone accounted for more than 40% of total segregation. This study has shown that the largest factors that contributed to decreases in segregation have all occurred within principal cities, where minority groups reduced segregation by growing in majority white areas. In 2020, micro segregation within cities still accounts for more than 30% of segregation, which is now on par with the contribution of macro segregation. In comparison, segregation within suburbs has remained stable, despite drastic changes in suburban racial diversity. The results show that there is substantial ‘racialized reshuffling’ within suburbs as well, but these changes contribute to a stable segregation trend. The suburbs hence become more important in understanding why segregation overall remains at high levels<sup>23–25</sup>. These findings show the importance of understanding segregation not as a ‘city-only’ phenomenon, but to include the entire metropolitan area in the analysis of segregation.

The results of this study also need to be contextualized by other large-scale demographic trends that have accompanied changes in segregation. First among these is the increasing diversity of the US population in terms of racial and ethnic groups<sup>26</sup> (see also Fig. 1), and the emergence of substantial ‘new minorities’, such as Hispanic and Asian people. At the same time, the white population is rapidly aging, and its growth rate has stalled. In addition, metropolitan areas in the South and West of the United States experience substantial population



**Fig. 5 | Detailed decomposition of changes in micro segregation, 1990–2020.** Black–white (top), Hispanic–white (middle) and Asian–white (bottom) micro segregation are decomposed separately. As before, the figure shows population-weighted averages. The simulation standard errors range from  $-0$  to  $0.003$ , and are therefore too small to be visually depicted. The sum of all bars within a panel sums up to the average change in micro segregation (‘Total Micro’ change in Fig. 3). Blocks are classified as either majority white, majority Hispanic or majority Asian blocks. Within each pairing of racial groups, we only consider the relevant majority blocks (for example, for Black–white segregation, we consider only majority white and majority Black blocks), and classify all other blocks as ‘mixed’. To interpret the individual factors, consider the example of ‘Black + in maj. white’ for Black–white segregation, where the ‘+’ stands for population growth. To compute this factor, each Census block where the white population has a share of at least 50% is classified in terms of either Black population growth or decline. Then, the counterfactual is computed only for those blocks where Black population growth occurred. The factor can then be interpreted as follows: if the Black population grew in majority white areas where we observed growth between 1990 and 2020, net of other population changes, how would segregation have changed?

growth. While changes in diversity and racial composition do not have to lead directly to changes in segregation<sup>27</sup>, they may have a number of indirect effects, such as the declining importance of racial identification or changing rates of marital homogamy. Future research should aim to incorporate these trends into the study of segregation.

To better understand the nature of changes in the population distribution, it would be desirable to distinguish between changes that are due to fertility and mortality from those that are produced by residential mobility. Ideally, the mobility component would be further broken down by mobility within the metropolitan area, longer-distance moves and international migration. At the small geographical scale, such data are not available, but it may be possible to model such estimates and then use them in decompositions. Similarly, it would be

desirable to incorporate breakdowns by socioeconomic status and age into the decomposition. The fact that the white population has a neutral effect on micro segregation suggests that gentrification does not have a large impact on reductions in segregation, but this may be an effect of lifecycle: potentially, white people move into minority neighborhoods when they are younger, but move out to majority-white neighborhoods as they get older. We also know that US metropolitan areas are shaped by large income and wealth inequalities<sup>28</sup>, which limits residential mobility and impacts mortality and fertility.

This study has highlighted the importance of understanding segregation as a process that is causally shaped by the presence or absence of population growth and decline in different parts of the metropolitan area. This implies that studies of segregation change need to be much more closely connected to the mechanistic demographic processes that cause segregation change<sup>29</sup>. From a theoretical perspective, then, what needs to be explained are not changes in segregation itself, but differences in rates of natural population growth between racial groups, the prevalence of residential mobility, the extent of international migration and so on. Counterfactual decomposition methods, such as the one developed in this paper, can then be used to understand the downstream consequences of these demographic changes on segregation.

## Methods

### Data

In studies of segregation, two research designs are commonly used: cross-sectional and harmonized studies. In a cross-sectional study, data are collected (usually for different Census years), and the geographical boundaries that are valid in each year are used. This is relevant for both blocks and places: blocks are redrawn for each Census (although many are stable), and places may have expanded or contracted. This cross-sectional approach was used, for instance, by ref. 5. In this paper, however, I am interested in decomposing changes in segregation by quantifying the impact of population growth and decline in certain, fixed areas. This design requires stable geographical areas, and I am therefore using a harmonized design, where 2010 blocks definitions are applied to the 1990, 2000 and 2020 files. The crosswalks are provided by the Integrated Public Use Microdata Series National Historical Geographic Information System (IPUMS NHGIS)<sup>30</sup>. For details on the procedure, see ref. 31. The crosswalk will introduce uncertainty into the estimates. I compare results from the cross-sectional and harmonized files and find that the uncertainty is small.

The full dataset is constructed as follows. I obtain block-level racial group counts from the Census datasets for 1990, 2000, 2010 and 2020 through IPUMS NHGIS<sup>30</sup>. I apply metropolitan area definitions from the Office of Management and Budget to these files. The February 2013 files are used as these definitions apply to the 2010 Census, which is the focal year to which the other years are crosswalked to. I remove metropolitan areas that contain less than 1,000 people of any of the four largest racial groups (Asian and Pacific Islander, Black or African American, Hispanic or Latino, and white). Consistent with prior research<sup>4,5</sup>, these four racial groups are coded such that Hispanic refers to Hispanic ethnicity of any race, and the other racial groups are non-Hispanic, single race (that is, ‘white’ refers to ‘non-Hispanic, single-race white’). This leaves 224 metropolitan areas. Blocks nest perfectly in places, and each block can be attributed uniquely toward a place or a nonplace area. In line with earlier research, I refer to nonplace areas, including Census-designated places, as ‘fringe areas’ (I am not aware that other studies that distinguish between place and nonplace areas include Census-designated places in the nonplace category. I argue that this approach is more sensible, as the interest in studying places is usually prompted by their legal status. As the name implies, Census-designated places do not coincide with any governmental or legal function, and I therefore choose to include them as ‘fringe areas’). To construct the harmonized file, I crosswalk 1990, 2000 and 2020 block definitions toward 2010 block definitions, and then apply place



definitions as of 2010. Therefore, the number of blocks and places is stable in the harmonized data set, as well as the definition of places and metropolitan areas.

The comparability of racial groups over time is complicated by two factors. The first of these factors is the increase in the population that identifies as two or more races. In most studies of segregation, this population is not included in the calculation of segregation indices, and the same method is adopted here. However, the share of this population has steadily risen: in 2000 (the first Census that allowed multiracial classification), the share was 1.8%, in 2000 it was 2% and in 2020 it had doubled to 4.1%. While the share of the multiracial population is still relatively low compared to the major racial groups, it would be desirable to incorporate this population in future research. The second complicating factor is the implementation of differential privacy (DP) for the 2020 Census. Simulation studies indicate that segregation measures might become more noisy<sup>32</sup>, without a clear bias in the lower or upper direction. Conclusive results for the effect on 2020 measures have not been reached, however.

As a sensitivity analysis that partially addresses both the large increase of the multiracial population in 2020, and the implementation of DP, the online appendix contains two additional figures: a version of Fig. 3 that breaks down the decomposition results by periods (that is, 1990–2000, 2000–2010 and 2010–2020), and a version of the same figure where the decomposition is not done at the detailed block level, but at the level of the Census tract. Tract-level estimates are far less affected by DP compared to Census blocks, although this analysis also injects some noise because Census tracts do not perfectly nest into places. A comparison of the results over time indicates that the period 2010–2020 does not deviate systematically from the earlier periods. Hence, if the analysis would only include the years 1990–2010, the interpretation of the results would be similar (or, at times, even more pronounced, given that the contributions of the white population toward increasing segregation have attenuated in the most recent period). The correlation between the block-level and tract-level decomposition results is around 0.93 for the periods 1990–2000 and 2000–2010, but has decreased to 0.85 for the period 2010–2020. A major part of this decrease is driven by comparisons involving the Asian population, which would be consistent with the fact that DP should especially affect smaller racial groups. Some caution should therefore be exercised when interpreting results that involve the Asian population for the period 2010–2020.

## Segregation measure

Throughout this study, I use the entropy-based segregation index  $H$ . To define  $H$  in general terms, let  $T$  be a matrix with  $U$  rows (spatial units) and  $G$  columns (racial groups). This study focuses on pairwise indices, so  $G = 2$  in this case. The rows  $U$  represent spatial units within a metropolitan area or a subarea, for example, blocks or places. Let the entries of  $T$  be  $t_{ug}$ , the number of people of race  $g$  in spatial unit  $u$ , and let  $t$  be the total population of  $T$ , that is  $t = \sum_{u=1}^U \sum_{g=1}^G t_{ug}$ . The joint probability of being in spatial unit  $u$  and racial group  $g$  is  $p_{ug} = t_{ug}/t$ . Also define  $p_{u.} = \sum_{g=1}^G t_{ug}/t$  and  $p_{.g} = \sum_{u=1}^U t_{ug}/t$  as the marginal probabilities of spatial units and racial groups, respectively. The index  $H$  is then defined as

$$H(T) = \frac{100}{E(T)} \sum_u \sum_g p_{ug} \log \frac{p_{ug}}{p_{u.} p_{.g}},$$

where  $E(T) = -\sum_{g=1}^G p_{.g} \log p_{.g}$  is the entropy of the racial group marginal distribution of  $T$ . In this formulation, the index ranges from 0 (absence of segregation) to 100 (complete segregation).

Using this general formulation, it is possible to calculate a number of  $H$  indices for a given metropolitan area. To quantify segregation in the entire metro area, define the matrix  $B$ , which contains as rows all Census blocks that belong to the metro area. The result of  $H(B)$  will

then quantify block-level segregation in the entire metropolitan area, referred to here as ‘total segregation’.

To make use of the micro–macro decomposition, also define a matrix  $P$  that aggregates Census blocks to places. This is possible because each block uniquely belongs to a place or a nonplace area. Note that matrices  $B$  and  $P$  describe the same population, but that  $P$  contains many fewer rows than  $B$ . Further, define a matrix  $B_s$ , which contains the subset of blocks that belong to place  $s$ . If we stack all matrices  $B_s$  for a given metropolitan area, we obtain matrix  $B$  again. The decomposition of  $H(B)$  is then given by

$$\begin{aligned} H(B) &= H_{\text{macro}}(B) + H_{\text{micro}}(B) \text{ where} \\ H_{\text{macro}}(B) &= H(P) \\ H_{\text{micro}}(B) &= \sum_{s=1}^S \frac{E(B_s)}{E(B)} p_s H(B_s), \end{aligned}$$

and where  $p_s$  is the population proportion of place  $s$  among the total metropolitan area population<sup>33</sup>. These two components are referred to as the macro (between-place segregation) and micro (a weighted sum of within-place segregation scores) components of total segregation.  $H_{\text{micro}}$  is not an  $H$  index, but a weighted sum of  $H$  indices.

For each combination of metropolitan area and year, we calculate Black–white, Hispanic–white and Asian–white segregation using the  $H$  index, each of which can then be decomposed into macro and micro components.

## Shapley decomposition

One core contribution of this paper lies in the development and application of a flexible methodology to understand changes in segregation. The basic idea is to use counterfactuals to study how factors affect the outcome, segregation, and then use a decomposition approach to quantify the importance of each factor. The method is first developed in general terms, and an example is given in the next section.

The decomposition procedure was first studied by ref. 16 in the context of game theory. To define the Shapley decomposition formally (the notation loosely follows ref. 34), let  $I$  be the outcome of interest,  $N = \{1, 2, \dots, m\}$  be the set of factors of interest, and  $\nu(\cdot)$  be a set function whose inputs jointly determine the outcome. The outcome of interest  $I$  can then be written as

$$I = \nu(N) - \nu(\emptyset) = \nu(\{1, 2, \dots, m\}) - \nu(\emptyset).$$

For this paper,  $\nu(N)$  returns segregation at time point 2, and  $\nu(\emptyset)$  returns segregation at time point 1. Hence,  $I$  is the outcome of interest, the change in segregation between time points 1 and 2.

The goal of any additive decomposition procedure is to find appropriate values for the contributions  $\varphi_1, \varphi_2, \dots, \varphi_m$  that satisfy

$$I = \varphi_1 + \varphi_2 + \dots + \varphi_m. \quad (1)$$

A ‘naive’ version that fulfills equation (1) is a decomposition that enters all factors sequentially, but the contributions will then depend on the order in which the factors are entered. This is often called the path-dependency problem in decomposition analysis<sup>35</sup>. The solution of the Shapley decomposition is to consider all the sequences in which the factors could be entered, arriving at the following decomposition rule for a set of factors  $N$  and a value function  $\nu$ :

$$\varphi_i(N, \nu) = \frac{1}{m!} \sum_{j_j \in N \setminus \{i\}} |j_j|! (m - 1 - |j_j|)! [\nu(j_j \cup \{i\}) - \nu(j_j)] \quad (2)$$

where the summation extends over all possible subsets  $j_j$  of  $N \setminus \{i\}$ , and  $|\cdot|$  denotes the cardinality of a set. The complexity of the formula should not distract from the fact that the idea is very simple. As ref. 34 writes,



'In broad terms, the proposed solution considers the marginal effect on  $I$  of eliminating each of the contributory factors in sequence, and then assigns to each factor the average of its marginal contributions in all possible elimination sequences. This procedure yields an exact additive decomposition of  $I$  into  $m$  contributions.' (p. 101)

As an example, consider an outcome  $I$  that we would like to decompose into two factors:

$$I = v(\{1, 2\})$$

At this point, we define two  $v(1)$  and  $v(2)$ , which reflect the situation if only factor 1 (2) is included in the calculation. Then, there are two possible elimination sequences for each factor. To eliminate factor 1, we can compare  $v(\{1, 2\})$  to  $v(2)$ , as well as  $v(1)$  to  $v(\emptyset)$ . To quantify the contribution  $\phi_1$  of factor 1, the average over the two possible elimination sequences is taken:

$$\phi_1 = \frac{1}{2} [v(\{1, 2\}) - v(2)] + \frac{1}{2} [v(1) - v(\emptyset)]$$

The elimination of factor 2 proceeds in the same way:

$$\phi_2 = \frac{1}{2} [v(\{2, 1\}) - v(1)] + \frac{1}{2} [v(2) - v(\emptyset)]$$

Because  $v(\{1, 2\}) = v(\{2, 1\})$ , it is immediately evident that

$$\phi_1 + \phi_2 = v(\{1, 2\}) - v(\emptyset).$$

If we define  $v(\emptyset) = 0$ , then  $I = \phi_1 + \phi_2$ . This shows that we have achieved an additive decomposition of  $I$  into two components, each of which depends only on the inclusion of the factor of interest.

### Simulation algorithm

A major downside of the Shapley decomposition is its computational complexity. Given that all subsets of  $N$  have to be computed (including  $v(N)$  and  $v(\emptyset)$ ), there are  $2^m$  computations necessary. Depending on the complexity of  $v$ , then, computing the Shapley decomposition for a large number of factors is often not feasible. I therefore implement an algorithm to approximate the solution of the Shapley decomposition by sampling randomly from the  $m!$  elimination sequences.

Key to the algorithm is the fact that the weighting factor in equation (2),  $|y_j|!(m-1-|y_j|)!$ , ensures that subsets of different sizes are given equal weight. If there are  $m$  factors in total, there are  $2^{m-1}$  subsets that exclude  $i$ . Of these subsets, there is  $\binom{m-1}{0} = 1$  subset of size 0, there are  $\binom{m-1}{1} = m-1$  subsets of size 1, there are  $\binom{m-1}{2}$  subsets of size 2, and so on. It follows then, that for a subset of size  $w$ , the total weight given to the subsets of this size is

$$\binom{m-1}{w} w!(m-1-w)! = (m-1)!.$$

As this number does not depend on  $w$ , all subsets of different sizes are given equal weight. This fact suggests the use of a two stage algorithm: first, randomly choose  $w$  from  $(0, 1, \dots, m-1)$  with equal probability; second, randomly choose a subset of size  $w$  with equal probability.

The algorithm requires two parameters:  $t$  is the minimum number of iterations, and  $s$  is the desired simulation error. The algorithm to approximate  $\phi_i$  for a factor  $i$  is as follows:

1. Repeat the following steps for  $j = 1, 2, \dots$ 
  - (a) Sample an integer between 0 and  $m-1$ ; call it  $w$ .

- (b) Sample  $w$  elements from  $N \setminus \{i\}$  without replacement; call the resulting set  $R$ .
- (c) Calculate  $\phi_i^j = v(R \cup \{i\}) - v(R)$ .
- (d) If  $j > t$ , estimate the simulation error  $\hat{s}$  as the standard error of the  $\phi_i^j$ 's, and stop if  $s < \hat{s}$ .
2. Let  $\phi_i \equiv \frac{1}{M} \sum_{j=1}^M \phi_i^j$  where  $M$  is the number of  $\phi_i^j$  sampled.

The number of contributions that are sampled is determined by  $t$ , the minimum number of contributions that are sampled for each factor, and  $s$ , the minimum simulation error that is desired. After some experimentation, I set  $t = 25$  and  $s = 0.01$  for all decompositions shown in this paper. By increasing  $M$ , the standard error could be further reduced and will eventually be indistinguishable from zero.

### Decomposition of segregation change

In this paper, I apply the Shapley decomposition to study segregation change. The general setup is as follows. Let  $T_1$  be the relevant contingency table at time point 1, and  $T_2$  be the contingency table at time point 2. Write  $H(T_1)$  for the segregation index at time point 1, and  $H(T_2)$  for the segregation index at time point 2. The outcome of interest is then

$$I = H(T_2) - H(T_1),$$

such that  $H(T_2) \equiv v(N)$  and  $H(T_1) \equiv v(\emptyset)$ .

One simple decomposition is to attribute changes in segregation to each of the racial groups. For the case of two racial groups, let the set of factors be  $N = \{A, B\}$ , where  $A = 1$  stands for the impact of population  $A$  on segregation change, and  $B = 2$  stands for the impact of the population  $B$  on segregation change. When both factors are included, the  $H$  value at time point 2 is obtained, and if no factors are included, the  $H$  at time point 1 is obtained. The remaining counterfactuals,  $v(A)$  and  $v(B)$ , are defined as follows: For  $v(A)$ , I calculate the  $H$  index for a matrix where the racial group counts for population  $A$  come from  $T_2$ , but the racial group counts for population  $B$  come from  $T_1$ . Thus, the index obtained through  $v(A)$  reflects a hypothetical situation where population  $B$  remains in place, while the population  $A$  is distributed as in  $T_2$ . The counterfactual  $v(B)$  is defined equivalently.

As a simple example, consider the following two contingency tables, where the first column contains counts for population  $A$ , and the second column contains counts for population  $B$ . This very small metropolitan area contains just three spatial units (for example, neighborhoods):

$$T_1 = \begin{bmatrix} 10 & 40 \\ 10 & 40 \\ 80 & 20 \end{bmatrix} \quad T_2 = \begin{bmatrix} 20 & 35 \\ 20 & 35 \\ 60 & 30 \end{bmatrix}$$

The  $H$  indices for these tables are  $H(T_1) = 28$  and  $H(T_2) = 6$ , for a decline of  $I = -22$ . Just from visual inspection of the matrices, it seems that population  $A$  redistributed more thoroughly. The first two neighborhoods gained 10 members of population  $A$  and lost 5 members of population  $B$ , while the third neighborhood lost 20 members of population  $A$ , and gained 10 members of population  $B$ . Both groups are now more evenly distributed across units, but how much impact did each racial group have on the total reduction in segregation? Let

$$v(A) = H \left( \begin{bmatrix} 20 & 40 \\ 20 & 40 \\ 60 & 20 \end{bmatrix} \right) = 12,$$

where the first column is taken from  $T_2$ , and the second column from  $T_1$ . This reflects the counterfactual situation that only population  $A$  has redistributed. Also,

$$v(B) = H \left( \begin{bmatrix} 10 & 35 \\ 10 & 35 \\ 80 & 30 \end{bmatrix} \right) = 19,$$

where the second column is taken from  $T_2$ , and the first column from  $T_1$ . Then compute

$$\begin{aligned}\varphi_A &= \frac{1}{2} [\nu(\{A, B\}) - \nu(B)] + \frac{1}{2} [\nu(A) - \nu(\emptyset)] \\ &= \frac{1}{2} [6 - 19] + \frac{1}{2} [12 - 28] = -14.5\end{aligned}$$

and

$$\begin{aligned}\varphi_B &= \frac{1}{2} [\nu(\{B, A\}) - \nu(A)] + \frac{1}{2} [\nu(B) - \nu(\emptyset)] \\ &= \frac{1}{2} [6 - 12] + \frac{1}{2} [19 - 28] = -7.5\end{aligned}$$

For this simple example, we would therefore conclude that about 65% of the decline can be attributed to changes in the distribution of population  $A$ . Note that the decomposition results are based on absolute numbers, and not on proportions. Hence, if the counts for one racial group stay identical between two time periods, none of the observed segregation change would be attributed to that group.

The racial-group decomposition attributes changes in segregation to only two factors. The Shapley decomposition allows us to define arbitrarily complex counterfactuals, with an (in theory) infinite number of factors. In this paper, the interplay between the different spatial units is of special interest. For instance, are segregation dynamics different in central cities, suburbs and fringe areas? To understand these dynamics, I construct counterfactuals that take into account the type of spatial unit, and whether the racial group in question grew or declined in that unit. Again, I define these decompositions separately for macro and micro decompositions.

For macro segregation, I distinguish between principal cities, suburban places and fringe areas. By combining these with the racial groups and their growth/decline trajectory, 12 factors of interest are obtained. For instance, the factors for Black–white macro segregation change are: ‘Black growth in suburban places’, ‘white decline in fringe areas’, ‘white growth in principal cities’ and so on. The distinction between growing and declining populations is important, because, at least in theory, both components are independent: for instance, the white population could grow in some suburbs (probably increasing segregation), but could decline in some other suburbs (possibly decreasing segregation).

For micro segregation, each block is classified by its racial composition in 1990, following precedent in the literature<sup>27</sup>. Blocks that are more than 50% Asian, Black, Hispanic or white are classified as ‘majority Asian’, ‘majority Black’, ‘majority Hispanic’ and ‘majority white’, respectively. Because majority Hispanic and majority Asian blocks should not be of much interest when decomposing changes in Black–white segregation, I distinguish only majority white and majority Black blocks, while all other blocks are classified as mixed. The equivalent procedure is used for the Hispanic–white and Asian–white decompositions. In total, 36 factors are obtained for each decomposition, with factors such as ‘Black growth in majority Black blocks in principal cities’, or ‘white decline in majority white blocks in suburban places’ and so on.

This paper focuses on the decomposition of two-group indices. Technically, it is also possible to extend the decomposition to multi-group indices of segregation. In the current study, this would result in a large number of terms that make the decomposition results difficult to interpret. In future research, it would be desirable to incorporate multigroup indices explicitly. This could be achieved, for instance, by developing procedures that select appropriate factors automatically, to make the decompositions results more interpretable.

## Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

## Data availability

All necessary datasets can be freely obtained through IPUMS NHGIS after registration. The following datasets are required: (1) 1990 Census:

STF 1 – 100% Data, variable: Hispanic Origin by Race; (2) 2000 Census: SF 1b – 100% Data [Blocks & Block Groups], variable: Population by Hispanic or Latino and Not Hispanic or Latino by Race; (3) 2010 Census: SF 1a – P & H Tables [Blocks & Larger Areas], variable: Hispanic or Latino Origin by Race; (4) 2020 Census: P.L. 94-171 Redistricting Data Summary File, variable: Hispanic or Latino, and Not Hispanic or Latino by Race. The crosswalk files are also obtained from IPUMS NHGIS. Metropolitan area definitions (2013) are obtained from the Office of Management and Budget. See <https://doi.org/10.17605/OSF.IO/YKJ4R> for a full replication package.

## Code availability

See <https://doi.org/10.17605/OSF.IO/YKJ4R> for a full replication package.

## References

- Farley, R. & Frey, W. H. Changes in the segregation of whites from Blacks during the 1980s: small steps toward a more integrated society. *Am. Sociol. Rev.* **59**, 23–45 (1994).
- Charles, CamilleZubrinisky The dynamics of racial residential segregation. *Annu. Rev. Sociol.* **29**, 167–207 (2003).
- Iceland, J. Beyond black and white. *Soc. Sci. Res.* **33**, 248–271 (2004).
- Reardon, S. F. et al. Race and space in the 1990s: changes in the geographic scale of racial residential segregation, 1990–2000. *Soc. Sci. Res.* **38**, 55–70 (2009).
- Lichter, D. T., Parisi, D. & Taquino, M. C. Toward a new macro-segregation? Decomposing segregation within and between metropolitan cities and suburbs. *Am. Soc. Rev.* **80**, 843–873 (2015).
- Massey, D. S. & Denton, N. A. Spatial assimilation as a socioeconomic outcome. *Am. Sociol. Rev.* **50**, 94 (1985).
- Massey, D. S. & Tannen, J. Suburbanization and segregation in the United States: 1970–2010. *Ethnic Racial Stud.* **41**, 1594–1611 (2018).
- Korver-Glenn, E. *Race Brokers: Housing Markets and Segregation in 21st-Century Urban America* (Oxford Univ. Press, 2021).
- Crowder, K., Hall, M. & Tolnay, S. E. Neighborhood immigration and native out-migration. *Am. Sociol. Rev.* **76**, 25–47 (2011).
- Krysan, M. & Farley, R. The residential preferences of Blacks: do they explain persistent segregation? *Soc. Forces* **80**, 937–980 (2002).
- Hwang, J. & McDaniel, T. W. Racialized reshuffling: urban change and the persistence of segregation in the twenty-first century. *Annu. Rev. Sociol.* **48**, 397–419 (2022).
- Hall, M. Residential integration on the new frontier: immigrant segregation in established and new destinations. *Demography* **50**, 1873–1896 (2013).
- Winkler, R. L. & Johnson, K. M. Moving toward integration? Effects of migration on ethnoracial segregation across the rural–urban continuum. *Demography* **53**, 1027–1049 (2016).
- Bader, M. & Warkentien, S. The fragmented evolution of racial integration since the Civil Rights Movement. *Sociol. Sci.* **3**, 135–164 (2016).
- Fischer, C. S., Stockmayer, G., Stiles, J. & Hout, M. Distinguishing the geographic levels and social dimensions of U.S. metropolitan segregation, 1960–2000. *Demography* **41**, 37–59 (2004).
- Shapley, Lloyd. S. in *Contributions to the Theory of Games (AM-28)* Vol. 2 (eds Kuhn, H. W. & Tucker, A. W.) 307–317 (Princeton Univ. Press, 1953).
- Rey, SergioJoseph, Cortes, R. & Knaap, E. Comparative spatial segregation analytics. *Spatial Demogr.* **9**, 31–56 (2021).
- Elbers, B. A method for studying differences in segregation across time and space. *Sociol. Methods Res.* <https://doi.org/10.1177/0049124121986204> (2021).
- Farrell, C. R. Bifurcation, fragmentation or integration? The racial and geographical structure of US metropolitan segregation, 1990–2000. *Urban Stud.* **45**, 467–499 (2008).

20. Johnson, K. M. & Lichter, D. T. Natural increase: a new source of population growth in emerging Hispanic destinations in the United States. *Popul. Dev. Rev.* **34**, 327–346 (2008).
21. Johnson, K. M. Subnational non-Hispanic white natural decrease in the United States. *Popul. Dev. Rev.* **46**, 7–31 (2020).
22. Nelson, C. The joint distribution of age and race in racially integrated neighbourhoods. *Popul. Space Place* **28**, e2511 (2022).
23. Logan, J. R. & Stults, B. J. *Metropolitan Segregation: No Breakthrough in Sight. Working Papers 22-14* (Center for Economic Studies, US Census Bureau, 2022).
24. Timberlake, J. M. & Howell, A. J. Trends in the ‘ecological distance’ of ethnoracial group suburbanization in U.S. metropolitan areas, 1970–2019. *Spat. Demogr.* **10**, 413–446 (2022).
25. Lichter, D. T., Thiede, B. C. & Brooks, M. M. Racial diversity and segregation: comparing principal cities, inner-ring suburbs, outlying suburbs, and the suburban fringe. *RSF* **9**, 26–51 (2023).
26. Frey, W. H. *Diversity Explosion: How New Racial Demographics Are Remaking America* (Brookings Institution Press, 2018).
27. Fowler, C. S., Lee, B. A. & Matthews, S. A. The contributions of places to metropolitan ethnoracial diversity and segregation: decomposing change across space and time. *Demography* **53**, 1955–1977 (2016).
28. Reardon, S. F. & Bischoff, K. Income inequality and income segregation. *Am. J. Sociol.* **116**, 1092–1153 (2011).
29. Bhrolcháin, M. áireN. í & Dyson, T. On causation in demography: issues and illustrations. *Popul. Dev. Rev.* **33**, 1–36 (2007).
30. Manson, S., Schroeder, J., Van Riper, D., Kugler, T. & Ruggles, S. IPUMS National Historical Geographic Information System: Version 16.0 *IPUMS* <https://doi.org/10.18128/D050.V16.0> (2021).
31. Geographic crosswalks. *NHGIS* <https://www.nhgis.org/user-resources/geographic-crosswalks> (2023).
32. Asquith, B. et al. Assessing the impact of differential privacy on measures of population and racial residential segregation. *Harvard Data Sci. Rev.* <https://doi.org/10.1162/99608f92.5cd8024e> (2022).
33. Mora, R. & Ruiz-Castillo, J. Entropy-based segregation indices. *Sociol. Methodol.* **41**, 159–194 (2011).
34. Shorrocks, A. F. Decomposition procedures for distributional analysis: a unified framework based on the Shapley value. *J. Econ. Inequal.* **11**, 99–126 (2013).
35. Fortin, N., Lemieux, T. & Firpo, S. in *Handbook of Labor Economics* Vol. 4 (eds Ashenfelter, O. & Card, D.) 1–102 (Elsevier, 2011).

## Acknowledgements

I thank the participants of seminars at Nuffield College and the Leverhulme Centre for Demographic Science for valuable comments. I also thank J. M. Aburto, T. DiPrete, Z. Doering, R. Gruijters and S. Mutgan for their detailed input on the paper at various stages of this project. This research was supported by a grant from the American Educational Research Association which receives funds for its AERA Grants Program from the National Science Foundation under NSF award NSF-DRL #1749275. Opinions reflect those of the author and do not necessarily reflect those of AERA or NSF. This research was also supported by the Horowitz Foundation for Social Policy.

## Author contributions

B.E. has undertaken all research described in this paper.

## Competing interests

The author declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s44284-024-00032-w>.

**Correspondence and requests for materials** should be addressed to Benjamin Elbers.

**Peer review information** *Nature Human Behaviour* thanks Ricardo Iglesias-Pascual, Richelle Winkler, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2024

## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- |                                     |                                     |  |
|-------------------------------------|-------------------------------------|--|
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | The statistical test(s) used AND whether they are one- or two-sided<br><i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | A description of all covariates tested   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | For null hypothesis testing, the test statistic (e.g. $F$ , $t$ , $r$ ) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted<br><i>Give <math>P</math> values as exact values whenever suitable.</i>                            |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | Estimates of effect sizes (e.g. Cohen's $d$ , Pearson's $r$ ), indicating how they were calculated   |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

### Software and code

Policy information about [availability of computer code](#)

Data collection No software was used for data collection.

Data analysis All analyses were done using R 4.2.0, including a number of open-source packages. Please see replication package for details.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

All necessary datasets are obtained through IPUMS NHGIS (<https://www.nhgis.org>). The following dataset are required:

- 1990 Census: STF 1 – 100% Data, variable: Hispanic Origin by Race



- 2000 Census: SF 1b – 100% Data [Blocks & Block Groups], variable: Population by Hispanic or Latino and Not Hispanic or Latino by Race
- 2010 Census: SF 1a – P & H Tables [Blocks & Larger Areas], variable: Hispanic or Latino Origin by Race
- 2020 Census: P.L. 94-171 Redistricting Data Summary File, variable: Hispanic or Latino, and Not Hispanic or Latino by Race

The crosswalk files are also obtained from IPUMS NHGIS (<https://www.nhgis.org/user-resources/geographic-crosswalks>).

Metropolitan area definitions are taken from the Office of Management and Budget (<https://www.census.gov/geographies/reference-files/time-series/demo/metro-micro/delineation-files.html>).

## Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender

n/a

Population characteristics

n/a

Recruitment

n/a

Ethics oversight

n/a

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description

quantitative, observational data

Research sample

full population in U.S. metropolitan areas - see datasets under "Data"

Sampling strategy

none (full population data)

Data collection

n/a - this is data collected by the U.S. Census Bureau

Timing

data for the years 1990, 2000, 2010, and 2020 was used

Data exclusions

no data was excluded

Non-participation

no participants were involved in the study

Randomization

n/a, as the dataset comprises the full population

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging