What's Behind a Racial Category? Uncovering Heterogeneity Among Asian Americans Through a Data-Driven Typology

LUCAS G. DROUHOT AND FILIZ GARIP

Despite emphasis on the importance of intragroup heterogeneity in much theoretically inclined migration and race scholarship, quantitative research routinely relies on split sample approaches in which ethnoracial groups are the categories of analysis. This cumulatively contributes to the reification of groups under study when research findings are assessed and groups compared side by side. In this paper, we ask: How are Asian Americans internally differentiated, and how does this heterogeneity matter for broader patterns of immigrant inclusion? Using latent class analysis, we produce a typology at the intersection of class, gender, regional location, and immigrant generation, pointing to vulnerable, ordinary, hyper-selected, rooted, *and* achieving *Asian Americans. These subgroups reveal differentiation in the experience of race and suggest that racialization and inclusion dynamics are jointly occurring social forces among Asian Americans. Our approach offers a blueprint for inductive analyses of immigrant-origin groups emphasizing heterogeneity and reflexivity vis-à-vis racial and national-origin categories.*

Contemporary sociological theories of immigrant incorporation and migration-driven diversity share a central concern for population heterogeneity. In the United States, segmented assimilation emphasizes the diversity of incorporation pathways among the second generation (Portes and Zhou 1993), while neoassimilation theory revamps Milton Gordon's classic assimilation model (1964) through a character-ization of the mainstream as diverse and flexible rather than a White, middle-class reference point (Alba and Nee 2003). In Europe, meanwhile, Steven Vertovec's (2007) influential theorizing points to the importance of heterogeneity occurring at the intersection of multiple social categories of difference within "superdiverse" immigrant groups.

Despite a theoretical consensus on the im-

Lucas G. Drouhot is a postdoctoral research fellow at the Max Planck Institute for the Study of Religious and Ethnic Diversity, Germany. **Filiz Garip** is professor of sociology at Princeton University, United States.

portance of heterogeneity in studying immigrant incorporation, however, quantitative research routinely relies on split-sample approaches in which ethnic and racial groups are the categories of analysis. The "general linear reality" and average-case focus of regression-based approaches (Abbott 1988) invariably flattens the social structure of the ethnoracial groups under study and cumulatively contributes to their reification when research findings are assessed and groups compared side by side. A recent review of empirical trends across large, census-based racial and national-origin categories (Drouhot and Nee 2019) shows that Asian-origin immigrants are collectively thought to be successfully incorporating in the United States (Kasinitz et al. 2008; Sakamoto, Goyette, and Kim 2009; Nee and Holbrow 2013) relative to Hispanic-origin groups, thanks in particular to their high rates of academic achievement (Hsin and Xie 2014). This overall positive picture, however, has led to calls to avoid essentializing Asian Americans as a model minority with a set of values leading to success (Lee and Zhou 2015), and to instead pay attention to intragroup heterogeneity in terms of socioeconomic attainment and racial experiences (N. Kim 2007; Min 2002; Dhingra 2007). An analytic strategy attentive to within-group heterogeneity would produce an epistemological break from everyday categories of practice (Brubaker 2013) and popular discourse reifying Asians as a cohesive group endowed with culture and agency. It would be further justified if it showed that the analytic payoffs of analyses across multiple empirical subgroups outweigh the costs to parsimony and statistical power of traditional analyses.

This article proposes and implements such an approach to intragroup heterogeneity among Asian Americans. We pool the pre- and post-election waves of the National Asian American Survey (Ramakrishnan et al. 2017) and use latent class analysis to create a data-driven typology of Asian Americans pointing to five major, clearly differentiated subgroups occurring at the intersection of class, gender, regional location, and immigrant generation: vulnerable, ordinary, hyper-selected, rooted, and achieving Asian Americans. Together, these categories capture *patterned differentiation*—the complex

aggregation of class, gender, and other characteristics in non-intuitive, yet nonrandom ways that reflects the temporal and geographic heterogeneity of Asian migration streams to the United States—in a more inductive and parsimonious manner than approaches relying on national origins.

Our contribution is twofold. First, we show the empirical potential of data-driven classifications to reconcile divergent empirical findings. We switch the focus from the question of how much assimilation or racialization occurs among Asian Americans as a whole, to who experiences what in this diverse population. Our analyses suggest new patterns of racialized incorporation, and reveal heterogeneity in the subjective experience of race and perceived discrimination across the national-origin groups making up the Asian American category. Second, our approach showcases the theoretical and epistemological potential of data-driven classification methods such as latent class analysis to study immigrant groups without "groupism" (Brubaker 2004)—that is, without importing racial or national-origin categories from everyday life and reifying them by a priori assuming their analytical relevance.

HETEROGENEITY AS A THEORETICAL AND PRACTICAL ISSUE IN MIGRATION RESEARCH

The importance of heterogeneity among native and immigrant groups is a leitmotiv in theoretically inclined migration research. According to segmented assimilation theory (Portes and Zhou 1993), the unprecedented ethnic and racial diversity of post-1965 immigration flows results in distinct modes of incorporation. Specifically, race interacts with government reception and characteristics of the ethnic community to create myriad trajectories for different groups. Often construed as an intellectual rival, the neoassimilation model (Alba and Nee 2003) differs in important respects but shares a concern for heterogeneity among immigrant groups—particularly their differential endowment in various forms of capital. This model jettisons the ethnocentric and essentialist definition of the mainstream as a White, middle-class core in Gordon's (1964) canonical model, and instead emphasizes the mainstream as het-

erogeneous and accommodative of cultural difference. Meanwhile, in Europe, the importance of heterogeneity finds its clearest expression in Vertovec's (2007) concept of superdiversity, which captures the interaction of multiple categories such as gender, place, and legal status that internally differentiate immigrant groups.

These three approaches are theoretical touchstones for large bodies of literature on both sides of the Atlantic. In spite of their differences, intragroup heterogeneity—that of immigrant or native groups—is central and theoretically generative in each. The concern for heterogeneity is both analytical and normative. That is, researchers need to accommodate increasingly diverse immigrant groups in a single theoretical model and to avoid essentializing minority and majority groups as homogenous entities with static traits and culture. This latter concern animates much of qualitative research on immigrant and other racial minorities, where long-standing debates on the relationship between minority culture and poverty have turned intragroup heterogeneity into a recurrent empirical and rhetorical motif used to avoid reifying stereotypes (Small, Harding, and Lamont 2010). To take just one influential and well-regarded example, Philip Kasinitz and his colleagues (2008, 23) write a cautionary note in their study of the second generation in New York City:

> We further recognize it is possible to read group comparisons as stereotypes or even racist generalizations. Let us be clear: any reference to group differences makes groups appear more homogeneous than they actually are. Our young respondents belonged not only to ethnic groups but also to social classes, genders, social groups, and neighborhoods. Like all modern people, they had a multiplicity of interacting social roles and identities. Although a quick reading of a table comparing groups will not always make this apparent, we have tried to remain sensitive to individual variation without losing sight of the real differences that ethnicity makes.

A generalized wariness to pointing out intergroup differences per se also reflects the influence of intersectional approaches (McCall 2005), for which a neglect of intragroup variation is responsible for blind spots on the experience of those belonging to intersecting social categories, as well as recent waves of Bourdieu-influenced theorizing on ethnicity, which advocates treating ethnic categories as neither bounded nor internally cohesive (Brubaker 2004; Wimmer 2009).

Despite such strong intellectual currents, survey-based analyses still rely on broad ethnoracial categories in the form of dummy variables (or split samples) for immigrant groups (defined typically by national origins) in regression analyses, cumulatively amounting to an "ethnoracial Olympics" (FitzGerald 2014) in which groups are assessed side by side on various incorporation outcomes. This approach has been undeniably fruitful (Waters and Jiménez 2005; Drouhot and Nee 2019). Yet the average-case focus on regression approaches also presumes a "general linear reality" (Abbott 1988) that erases within-group variation. Many empirical findings, as a result, remain blind to the social structure of immigrant groups—the specific configurations of attributes making up distinct types of immigrants within broad racial or national-origin categories (Garip 2012; Brubaker 2004). One can discover such configurations with regression models that include interaction terms between indicators of immigrant groups and other social categories (such as gender or regional location), but one would quickly run into the untenable issue of data sparsity. Given these difficulties, much contemporary quantitative migration research has eschewed the issue of immigrant heterogeneity despite its salience in theoretical and qualitative work. Here, we propose and implement an analytic approach attentive to the "consolidated parameters of the social structure"—that is, the patterned but not intuitive interrelations of various forms of social differentiation (Blau 1974).

TENSIONS ACROSS RESEARCH COMMUNITIES

Existing empirical accounts suggest that Asian-origin immigrants are successfully incorporating into the United States (Drouhot and Nee 2019; Zhou and Gonzales 2019; Kasinitz et al.

2008; Sakamoto, Goyette, and Kim 2009; Nee and Holbrow 2013) relative to Hispanic-origin groups, in particular on the basis of their extraordinarily high rates of academic achievement (Hsin and Xie 2014; Lee and Zhou 2015). Specialized literatures document tell-tale signs of assimilation, such as high rates of intermarriage (Qian and Lichter 2011) and residential attainment in immigrant middle-class neighborhoods that form "ethnoburbs" (Li 2009; Matsumoto 2018) where Whites are relegated to being "just alright," as Tomás Jiménez and Adam Horowitz's (2013) ethnographic research vividly describes.

The aggregate trend of Asian American assimilation is a "stylized fact" (Hirschmann 2016)—an "empirical regularity in need of an explanation." Dominant explanations for the aggregate success of Asian American groups in the United States emphasize immigrant selectivity (Lee and Zhou 2015), legal status and immigration law privileging high-skilled migration (Nee and Holbrow 2013), and—directly related to selectivity patterns—a culture and community norms of academic and professional success (Hsin and Xie 2014; Kasinitz et al. 2008). Although sociologists have traditionally steered clear of depicting Asian Americans as a model minority, the stylized fact of Asian American assimilation remains vivid when contrasted with the trajectories of Hispanic-origin migrants, whose legal status and endowment in various forms of capital are far more precarious (Drouhot and Nee 2019; Zhou and Gonzales 2019; Pew Research Center 2012).

In the assimilation narrative, the law shapes the selectivity of migration flows and resulting assimilation trajectories, not race. As Victor Nee and Hilary Holbrow write, "the mainstream success of so many Asian American immigrants suggests that race may not be such a decisive factor in shaping socioeconomic attainment as it was in the American past, and that assimilation still is as characteristic of the course of contemporary immigration as it was for earlier immigration from Europe. In an increasingly inclusive mainstream, the significance of race has declined considerably. Instead, patterns of legal and illegal entry are more consistently determinative of immigrant access to mainstream opportunities" (2013, 72).

Other sociologists have criticized this interpretation, pointing instead to the persisting influence of race in shaping the social experience of Asian Americans and treating the model minority stereotype as a myth (Chou and Feagin 2008). These perspectives—which we refer to broadly as *critical race*—hold that Asian Americans have yet to reach equality with the White majority. The overeducation thesis suggests that Asian Americans compensate for their racial disadvantage by being more educated than Whites to secure similar incomes or jobs (Takaki 1998). Others point to a "bamboo ceiling" blocking high-achieving Asian Americans from leadership positions in the workplace as evidence of persisting discrimination (Varma 2004), and document racial stereotypes specific to this group, such as being asocial, subservient, and shy in creativity (Lin et al. 2005; Sue et al. 2007). The perception of racial stigma and discrimination has been linked to health outcomes (Paradies et al. 2015; Gee et al. 2009), political participation (Lien 2001; Wong et al. 2011), and panethnic identity (Masuoka 2006; Kibria 1998; Okamoto 2003) among Asian Americans in past work. We revisit this point later.

A second critique of the assimilation perspective is concerned with the homogenization of Asian Americans as a racial group endowed with a specific culture of success (Lee and Zhou 2015). This work typically focuses on educational and socioeconomic attainment, and sets out to "unravel" (Lee 1996), "complicate" (Ngo and Lee 2007) or "deconstruct" (Museus and Kiang 2009) the model minority narrative. Empirical studies focus on subgroups whose experience diverges from the assimilation narrative—for instance, academic low-achievers (Lee 1996) or severely disadvantaged groups, such as the Hmong and Laotians who largely came to the United States as refugees (Ngo and Lee 2007; Museus and Kiang 2009)—and seek to mitigate the optimism induced by aggregate census trends. As reviewed, the emphasis on heterogeneity has an *analytical* rationale (depicting a group's complex reality) as well as a *normative* one (avoiding the essentialization of Asian Americans as a successful group, implicitly blaming other immigrants for their lower attainment).

MOTIVATION OF THIS STUDY

An overview of the work on Asian Americans shows different strands of research in significant tensions with one another. Although the assimilation literature depicts an optimistic trend of progress toward the mainstream, other lines of work highlight the importance of race in shaping incorporation, and drawing attention to less advantaged groups, such as those originating from Southeast Asia. Our goal is to implement an empirical approach to describe heterogeneity among Asian Americans and productively reconcile these divergent perspectives. We also seek to put into practice long-standing theoretical concerns about intragroup heterogeneity discussed earlier (Alba and Nee 2003; Portes and Zhou 1993; Vertovec 2007).

Our analysis of within-group heterogeneity switches the focus from how much Asian Americans are experiencing either assimilation or racialization, to who is undergoing which processes in this diverse population. We expect that incorporation processes might work differently across subgroups. We seek to identify patterned differentiation among Asian Americans and to assess the relative prevalence of each subgroup, and productively circumvent the intellectual stalemate between the assimilation perspective highlighting aggregate trends and more recent work focused on small but unrepresentative subgroups. Further, we study how the experience and perception of race varies across the subgroups making up Asian Americans. Last, we illustrate the differential impact of racial discrimination across subgroups on three key outcomes: health, political participation, and panethnic identity.

This study is not the first to consider heterogeneity among Asian Americans. Instead, it is something many scholars have called for in their research. For example, Nadia Kim writes, "Socioeconomic data on Asian Americans need also be *disaggregated*. Asian Americans—consisting of Bangladeshi, Cambodian, Chinese, Filipino, Indian, Japanese, Korean, Laotian, Pakistani, Vietnamese, Pacific Islander ethnics, and so on—are among the most diverse of the racialized groups and are internally stratified in profound ways. Yet social scientists tend to lump all of these groups together, not differentiating between ethnic/national groups that are

highly dissimilar" (2007, 565, emphasis in the original).

Nadia Kim's call for disaggregation echoes historical perspectives emphasizing the political nature of the Asian American panethnic category—one born out of long-standing patterns of racial exclusion as well as political struggles for recognition among otherwise dissimilar migrant groups (see Le Espiritu 1992; S. Kim 2020, 4–9; Takaki 1998). Quantitative studies indeed often disaggregate Asian Americans by national origins (Srinivasan and Guillermo 2000; Hsin and Xie 2014; Sakamoto, Goyette, and Kim 2009; Kim and Sakamoto 2010), to "problematize the model minority image of Asian Americans" (Zhou and Xiong 2005, 1). Yet national origins constitute only one possible dimension of differentiation among many competing ones (such as social class, immigrant generation, region of settlement), and its use leads to methodological nationalism—the naturalization of nation-states as categories of analysis by social scientists (Wimmer and Glick-Schiller 2003). In our case, assuming that Asian Americans are primarily made up of national-origin groups would replace a much-criticized racial essentialism with an ethnic one whereby national-origin groups are implicitly assumed to have a common fate and culture. Therefore, we consider the salience of ethnic differences among Asian Americans as an open question, and one to be answered empirically rather than methodologically "baked in" (Brubaker 2004, chapter 1).

EMPIRICAL APPROACH AND DATA

Our goal is to characterize social heterogeneity among Asian Americans. Researchers use split samples, interaction terms, or hierarchical models to show the heterogeneity in factors relevant to immigrant integration outcomes. Studies on Asian Americans, for example, highlight differences among men and women (Min and Kim 2009), among different national-origin groups (Qian, Blair, and Ruf 2001), or across geographic contexts (Okamoto 2007). But such approaches overlook the fact that different dimensions might work in conjunction to produce different outcomes for different individuals. This idea—captured most vividly in writings on intersectionality (McCall 2005) and super-

diversity (Vertovec 2007)—suggests a particular direction for empirical work: rather than focusing on a single dimension of social life, we need to think about configurations of multiple dimensions to understand heterogeneity in the social experiences of immigrant groups. This is difficult to do with regression analysis, where higher-order interactions needed to capture such patterns are often uninterpretable and hard to estimate because of data sparsity for certain categorical combinations.

Consider a simple case. We want to study the heterogeneity in integration outcomes across two binary attributes: gender (man or woman) and income (high and low). In combination, the two attributes yield four possible categories. It is easy to consider each category with a two-way interaction in a regression model, and to make sense of comparative evaluations across categories. Now, add two additional binary attributes: immigrant generation (first or subsequent) and education (high school or college). If we cross-classify all four attributes, we get two to the fourth power, or sixteen possible categories. It is difficult to interpret variation across all sixteen, let alone identify patterns with statistical analysis (which would require four-way interactions in regression models, and possibly lead to data sparsity issues if some of these categories are scarcely populated).

One way around this problem comes from recognizing that not all possible combinations are equally prevalent in the data. This is because many attributes (such as education, income, immigrant generation) are highly correlated with one another, which implies that individuals naturally cluster around a few distinct configurations. Therefore, in this study, rather than dissecting the data or including multiple interaction terms to consider different combinations of a few selected attributes, we focus on identifying these configurations.

Several methods identify the configurations that define distinct groups among Asian Americans. Many scholars turn to cluster analysis or latent class analysis (LCA). For example, Filiz Garip (2012, 2016) uses cluster analysis to identify four distinct groups among first-time Mexico-U.S. migrants, whereas Bart Bonikowski and Paul DiMaggio (2016) use latent class analysis to characterize four types of popular nationalism in the United States (see also Drouhot, forthcoming, for another example of cluster analysis). Both cluster analysis and LCA partition the data into groups and fall under the general umbrella of unsupervised machine learning—a suite of methods from computer science that search for representations of a set of attributes (X) that is more useful than X itself (Molina and Garip 2019). These methods are data driven or inductive because they use the data—not prenotions from the researcher—to identify a categorization scheme, that is, a model through which that data can be understood. These methods are an efficient way to describe the data and to study their inherent heterogeneity parsimoniously.

We use latent class analysis, which estimates a latent (unobserved) variable that accounts for the covariance between the observed attributes (see also S. Kim 2021, this issue). This variable is assumed to have a categorical distribution with each value corresponding to a "latent class" (group) in data. LCA is similar to cluster analysis in that it detects underlying groups. It differs from cluster analysis in that it uses a model to describe the distribution of the data, and is therefore probabilistic rather than deterministic.[1] LCA enables estimating the probability that a case belongs to a particular latent class (rather than rigidly assigning each case to a group, as often done in cluster analysis). When assigning a case to a latent class, we use posterior probabilities to create multinomial

1. Besides, mainstream clustering algorithms such as the popular *k-means* algorithm are not able to incorporate nominal variables—which feature prominently in our approach—without inducing significant distortion in the data partitioning process (Magidson and Vermunt 2002). K-means algorithm and nominal data are incompatible since distance functions on nominal data spaces are not meaningful. One workaround involves dichotomizing each category, but this leads to nominal variables taking excessive weights as multiple, highly correlated binary variables in the computation and final clustering results altogether. The integration of nominal data within cluster analysis remains an area of active research within data and computer science (see, for example, Roy and Sharma 2010).

distributions and assign cases at random based on these distributions so that our assignment is probabilistic. The appendix formally describes our empirical approach and provides further technical details.

Our analyses are based on the pooled pre- and post-election waves from the 2016 National Asian American Survey (Ramakrishnan et al. 2017). To discover groups in data, we chose a set of attributes shown to shape immigrant experiences in the United States. These include indicators for education (less than high school, high school or some college, college degree or more), income (earning $100K or more), gender, immigrant generation (first, second, third, or more), residence (California, West excluding California, East, Midwest, South, or Pacific), and are intended to capture both the selectivity of the migrant flows (with respect to gender, education, and income) and their context of reception (as proxied by immigrant generation and residential region in the United States).[2] Gender, for instance, is key to shaping migration reasons and strategies, as well as integration outcomes in destination (Donato et al. 2006). Socioeconomic class of respondents captures both the reasons underlying migration among first-generation migrants and the trajectories of integration among the second or third generation (Waters and Jiménez 2005; Drouhot and Nee 2019). Similarly, region of residence captures both the likely pull of existing social networks in place (Sue, Riosmena, and Lepree 2019) as well as varying levels of receptivity to immigrants in the United States (Waters and Eschbach 1995). In keeping with our intention to avoid methodological nationalism, we do not include national origins in the latent class modeling.[3] Likewise, we do not include cultural or attitudinal variables because our fo-

cus is on the sociodemographic dimensions reflecting social structural differentiation among Asian Americans.[4]

LCA requires researchers to specify a priori the number of classes in the data. To avoid obtaining artificial groupings, researchers use goodness-of-fit measures to choose the optimal number of classes. This process resembles model selection in regression analysis, where researchers rely on a criterion like the likelihood ratio to choose the best and most parsimonious model for the data. Here, we used two measures—the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC)—to select the optimal number of classes. Both measures capture the trade-off between goodness-of-fit (which is improved by making the model more complex, for example, by specifying more classes to describe the data) and parsimony (where simpler models are presumed to be better). BIC gives stronger weight to parsimony.

Figure 1 shows AIC and BIC values when we estimate the LCA model with different numbers of classes, ranging from three to eight. For both measures, smaller values indicate better model fit. For BIC, models with three, four, and five classes yield the lowest values of the index. For AIC, models with five classes or higher minimize the index. We select the five-class model for two reasons. First, both AIC and BIC indicate it as nearly optimal. Second, our inspection of alternative models suggests this solution to be ideal for identifying substantively meaningful heterogeneity while retaining interpretability.

As an alternative evaluation of the selected model, we rely on national-origin indicators—which were deliberately left out from the attribute set used in the discovery of classes—to

---

2. Because of data constraints, we consider all foreign-born respondents as first generation, including those who migrated at an early age. The post-election survey, which we merge with the pre-election survey for our analyses, does not feature a question on year of arrival that would allow us to create a 1.5 generational status (indicating arrival at or before adolescence). Additionally, the binary age variables (younger or older than thirty-five) are not precise enough for that purpose.

3. We repeated our analyses with national-origin categories included and obtained substantively similar groupings (results available on request).

4. See Sunmin Kim's article in this issue for an analogous application of latent class modeling to attitudinal items from the same data (2021).

Figure 1. Model Fit Indices for Latent Class Analysis
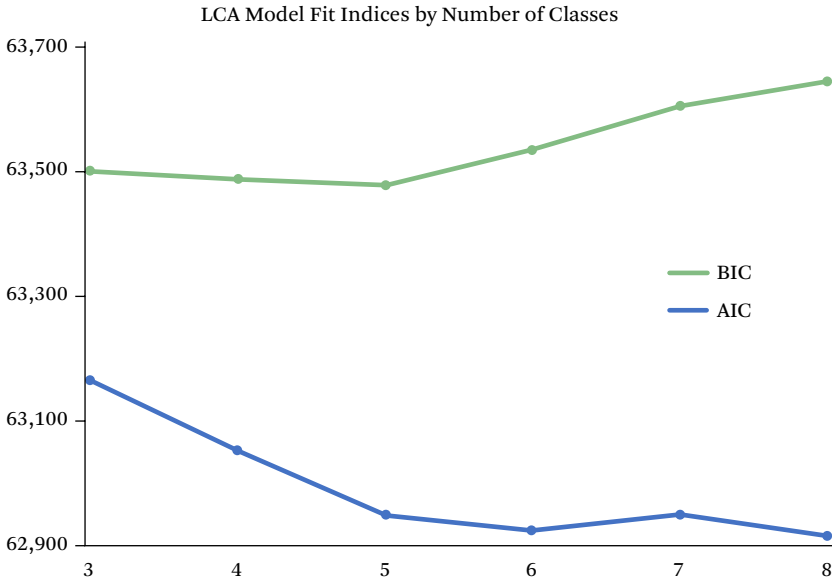
LCA Model Fit Indices by Number of Classes

Source: Authors' tabulation based on National Asian American Survey data (Ramakrishnan et al. 2017).
Note: For both the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), lower values indicate better model fit.

perform substantive validity checks. Earlier work relied on similar strategies to validate findings from data-driven methods. For example, Garip (2016) confirms that four migrant types in her data—obtained by clustering survey responses alone—related differently to macro-level economic and political indicators. Similarly, Bonikowski and DiMaggio (2016) test whether four varieties of nationalism in their data correlated with social and policy attitudes that were not used in the identification of the typology. In the same spirit, we confirm that the distributions of national-origin groups across our five classes conform to the general patterns for these groups identified in empirical work. We first describe the five classes and then examine how class membership is associated with the subjective experience of race across subgroups. In a third step, we examine heterogeneity in the statistical relationship of experienced discrimination with self-reported health, political participation, panethnic identity—three outcomes on which discrimination is presumed to have a causal impact in past work.

RESULTS

Identifying Heterogeneity Among Asian Americans

The five columns in table 1 show the mean values of characteristics used in latent class analysis for each of the five classes. The last row shows the number of respondents in each class. The characteristics include binary indicators for education (less than high school and college degree or more), income (earning $100K or more per year), gender (female), immigrant generation (first, second, and third or higher), region (California, West excluding California, East, Midwest, South, and Pacific). For each characteristic, bold values are significantly higher than the next highest value across five classes ($p<.05$, one-tailed test).

The first class contains the highest share of respondents with less than a high school degree (0.69), as well as the highest share of women (0.68) and the first generation (0.99) across all groups. The class includes the smallest share (0.03) of respondents earning a high income. Close to half of respondents in this

Table 1. Patterned Differentiation Among Asian Americans

	Class 1 Vulnerable	Class 2 Ordinary	Class 3 Hyper-selected	Class 4 Rooted	Class 5 Achieving	Grand Mean
Nativity (proportion)						
First generation	**0.99**	**0.89**	0.91	**0.17**	**0.51**	0.76
Second generation	0.00	**0.11**	**0.06**	**0.22**	**0.46**	0.15
Third generation and above	**0.01**	0.00	0.03	**0.61**	**0.02**	0.08
Socioeconomic status (proportion)						
Less than a high school diploma	**0.69**	**0.18**	**0.02**	0.00	0.00	0.20
College degree or more	0.07	**0.12**	**0.67**	**0.42**	**0.53**	0.36
Earning more than $100K	0.03	**0.08**	**0.45**	**0.31**	**0.38**	0.25
Region (proportion)						
California	**0.44**	**0.30**	0.24	0.32	**0.62**	0.38
Western United States other than California	**0.08**	**0.14**	0.10	0.10	0.05	0.10
Eastern United States	0.19	0.19	**0.18**	0.00	**0.14**	0.16
Midwestern United States	0.19	**0.11**	0.07	0.05	**0.17**	0.12
Southern United States	**0.08**	**0.19**	**0.41**	0.03	0.02	0.17
Pacific United States	**0.02**	**0.07**	0.00	**0.50**	0.00	0.08
Gender (proportion female)	**0.68**	0.47	0.32	0.48	**0.44**	0.48
Observations	1,372	1,227	1,425	663	1,171	5,858

Source: Authors' tabulation based on National Asian American Survey data (Ramakrishnan et al. 2017).
*Note:* Table shows the composition of each class (columns) in terms of various characteristics (rows) included in the analysis. Bolded values are significantly higher than the next lowest value ($p < .05$, one-tailed test).

class live in California (0.44); about one-fifth (0.19) each reside in the East and the Midwest. Respondents here are by far the most disadvantaged among all groups given their low education, low income, and first-generation status. We therefore call them the vulnerable.

The majority of respondents in the second class hold a high school degree or have some college experience (0.70, not shown). About one in five (0.18) has not completed high school, and about one in ten (0.12) has a college degree or more. Given the educational distribution, it is not surprising that less than one in ten (0.08) earns an income of $100K or more. This group is gender balanced and includes a large majority (0.89) of the first generation. The most likely destination is California (0.30), followed by the East (0.19) and the South (0.19). The respondents in the class are not as disadvantaged as those in the first group, but they are also not

nearly as educated or high earning as those in the next three groups. Given these characteristics in a population whose depictions are polarized as either extremely successful or disadvantaged, we call them the ordinary.

The third class contains the best-educated and highest-earning respondents across all groups. Two-thirds (0.67) hold a college degree or higher and about a half (0.45) earn more than $100K. The group is predominantly male (0.68) and first generation (0.91). A majority lives in the South (0.41) or in California (0.24). This class echoes the literature on upward mobility and professional attainment among Asian Americans emphasizing "hyper" selectivity in the composition of Asian migration flows in terms of human capital, and the role of such selectivity in producing the stereotypical, high-achieving Asian American groups (Lee and Zhou 2015; Nee and Holbrow 2013). In recogni-

Figure 2. Relative Proportions of Five Classes Among Asian Americans
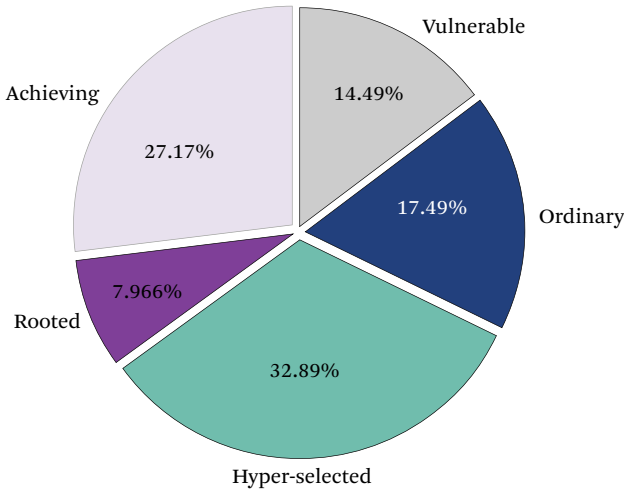
Source: Authors' tabulation based on National Asian American Survey data (Ramakrishnan et al. 2017).
Note: Proportions are adjusted by sample weights.

tion of its high socioeconomic standing and large majority of foreign born, we label this group the hyper-selected.[5]

The fourth group is also educated and high earning, though not quite at the level of the hyper-selected. About 40 percent of respondents hold a bachelor's degree or higher, and close to one-third (0.31) earn $100K or more. The most likely destination (0.50) is the Pacific region, followed by California (0.32). Unlike the first three groups, which contain mostly the first generation (the immigrants), most respondents here are third generation (0.61). To highlight this distinctive trait, we call this class the rooted.

The fifth and final class is between the rooted and the hyper-selected in terms of education and income. About half of the respondents (0.53) in this group are college educated or hold graduate degrees, and more than one-third (0.38) earn $100K or more. The group is almost equally split between the first generation (0.51) and the second (0.46). Most of the

group (0.62) are settled in California, and the remainder mostly in the East (0.14) and Midwest (0.17). Following ethnographic description of Asian-origin families in California emphasizing academic success among their children (Jiménez and Horowitz 2013), and to recognize its relatively high socioeconomic attainment as well as its mixed generational status, we call this group the achieving.

Figure 2 displays the distribution of the sample across the five classes.[6] The hyper-selected make up one-third (33 percent) of the adjusted sample, the achieving account for over a fourth (27 percent), and the rooted less than one-tenth (8 percent). Together, these three groups, which are closest to the model minority stereotype, account for more than two-thirds of Asian Americans in the sample. This statistical dominance helps explain the popularity of this narrative as well the stylized fact of Asian American immigrant success in the literature based on average-case analyses and samples aggregated by racial groups (Sakamoto, Goyette, and Kim

5. Supplementary analyses (not shown) indicate that a large majority of respondents in this class (more than 70 percent) completed their high level of education before migrating to the United States, which supports our interpretation emphasizing selectivity.

6. The shares shown on the chart adjust for sampling weights, and therefore slightly differ from the shares implied by the class sizes in the last row of table 1.

2009; Kao and Thompson 2003; Hsin and Xie 2014). By contrast, the vulnerable cluster (15 percent) is more reminiscent of Southeast Asian subgroups who mainly came to the United States as refugees with little human capital (Ngo and Lee 2007; Gordon 1987).

The gender composition of each class is noteworthy. Only three (ordinary, rooted, and achieving) of our five groups are gender balanced. The vulnerable are predominantly female, the hyper-selected predominantly male. This gendered pattern—where women occupy disadvantaged positions relative to men—is not attributable to national origins alone, given that both the vulnerable and hyper-selected include all the nationalities in the sample (figure 3). Our categorization suggests complex linkages between gender, national origins, and socioeconomic status that are underappreciated in existing research.

Finally, our five classes reveal an increasing diversity among the first and second generation than among subsequent generations. Whereas the first generation is distributed across the vulnerable, ordinary, hyper-selected, and achieving categories, members of the third generation are mainly in the rooted group. This pattern suggests that earlier arrivals to the United States are relatively similar to one another in terms of education, income, and state of residence. By contrast, the later arrivals (mostly in the post-1965 era) show greater variation.

Figure 3 shows bar charts with the distribution of national origins among respondents in each of the five classes (panel A), and vice versa (panel B). Two general patterns are evident. First, the composition of each group is well aligned with descriptions of national-origin groups in the literature. For example, the hyper-selected and achieving groups both include a large share of Indians and Koreans, two countries often noted for the selectivity of their U.S.-bound migrants (Krogstad and Radford 2018). The vulnerable group, by contrast, includes large shares of Cambodian and Hmong migrants, two groups with less education than the native-born population and who largely entered the United States as refugees (Gordon 1987). The rooted category has the most overlap with a single national-origin group—namely, the

Japanese, whose economic integration across three immigrant generations has been the object of past work in economic sociology (Bonacich and Modell 1980). The comparison of our five classes against known patterns on national-origin groups thus offers a substantive validation of our categorization.

Second, each of our five classes contains all national origins but in varying compositions. For example, although the hyper-selected identifies Indians and Koreans as its largest constituent national groups, it also includes Cambodians and Hmong. Likewise, Chinese immigrant groups are present in all latent classes. The diversity of each group with respect to national origins confirms a point often made but rarely implemented in quantitative analyses. Treating nations as a taken-for-granted category of analysis, or methodological nationalism (Wimmer and Glick-Schiller 2003), blinds us to important differences within each national group. It also does not allow us to consider potential similarities among individuals across national groups. For instance, national-origin groups often heralded for their academic success and upward mobility, such as the Chinese and the Koreans, are present in latent classes characterized by social disadvantage (for example, the vulnerable and to a lesser extent the ordinary). National-origin groups (panel B) appear more diverse than commonly depicted in the literature, where ethnic origins often proxy certain dominant traits—such as the high human capital of Indian-origin immigrants or the socially disadvantaged character of the Hmong.

Linking Patterned Differentiation to Heterogeneity in the Subjective Experience of Race

In their review, Jennifer Lee and Samuel Kye (2016, 255) argue for "examin[ing] more critically the processes, not just the outcomes" of Asian American assimilation. Our approach seeks to do just that by, first, identifying the distinct configurations of characteristics that define different groups among Asian Americans, and then investigating the orientation of each group to its racial identity as well as its experience with the American mainstream. We, in other words, characterize the diversity in the

Figure 3. Distribution of National Origins
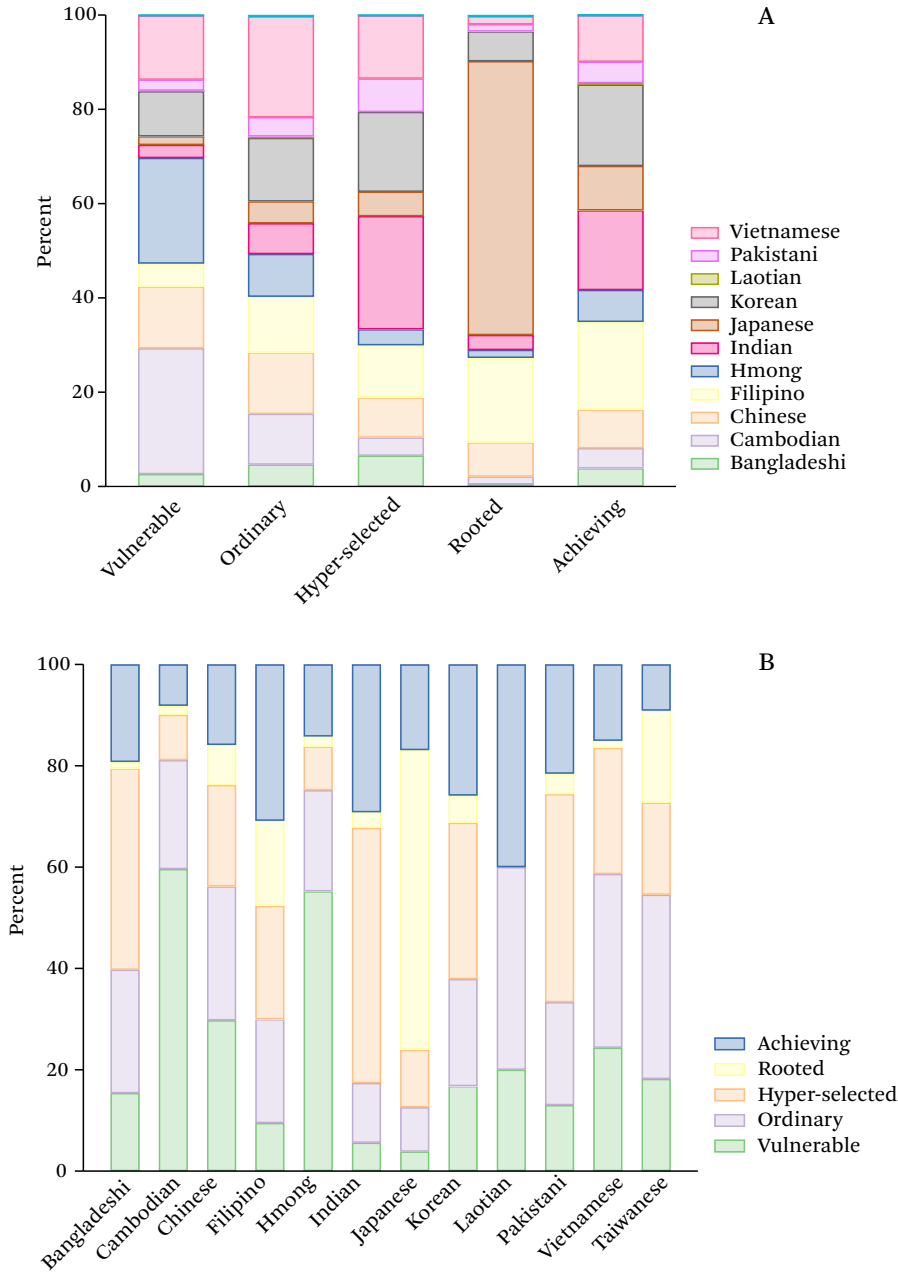
Source: Authors' tabulation based on National Asian American Survey data (Ramakrishnan et al. 2017).
Note: Distribution of national origins within five classes of Asian Americans (panel A) and distribution of class membership by national origins (panel B). Groups are stacked according to the order of the legend to facilitate reading. The optimal way to view this figure is in color. We refer readers of the print edition of this article to https://www.rsfjournal.org/content/7/2/22 to view the color version.

Figure 4. Differences Across Five Classes, Subjective Experience of Race
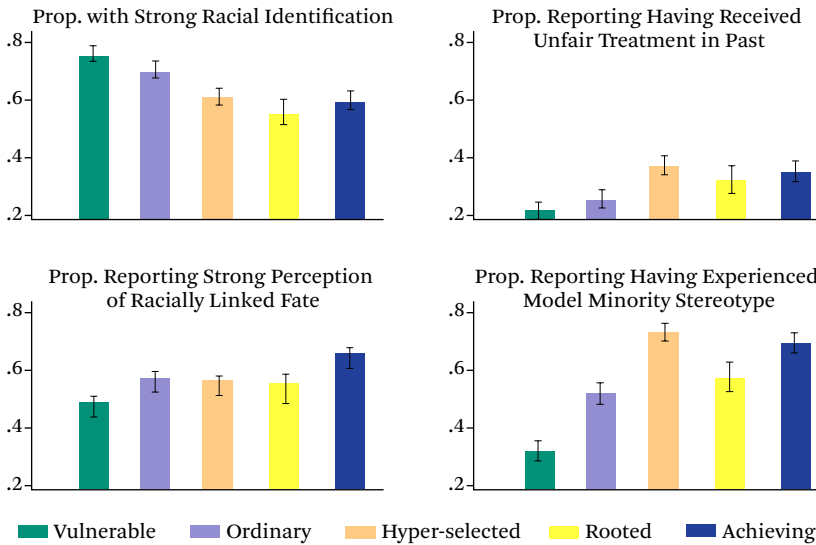
Source: Authors' tabulation based on National Asian American Survey data (Ramakrishnan et al. 2017).
Note: Color-coded bars show the average proportions within each group that strongly identify as Asian American (upper-left panel), report having received unfair treatment in the past (upper-right panel), report strong perceptions of racially linked fate (lower-left panel), and report having experienced stereotyping as "model minority" (lower-right panel). The brackets over each bar indicate 95 percent confidence intervals around the average value.

population to allow for diversity in various outcomes, and to better understand the mechanisms underlying it.

Figure 4 compares the share of respondents in each class that report experiences with four outcomes related to the subjective experience of race.[7] The upper left panel shows that the proportion reporting a strong racial identity is significantly higher in the vulnerable and ordinary groups (>70 percent) than the hyper-selected, rooted, and achieving (~60 percent). Although racial identification is prevalent in all groups, it seems to weaken across immigrant generations, or through achievement of higher socioeconomic status. The share of respondents strongly identifying as Asian American is higher among groups dominated by the first generation (vulnerable and ordinary) than among those dominated by the second and third generation (rooted and achieving). The hyper-selected offer a notable exception. Al-

though this group is predominantly first generation, it displays levels of racial identification comparable to more established migrant groups, presumably due to its high socioeconomic status.

The upper right panel displays group-specific shares experiencing unfair treatment. About 25 percent of the vulnerable and ordinary report discrimination, versus nearly 40 percent of the hyper-selected, rooted, and achieving. The vulnerable contain mostly women; the ordinary include a majority of respondents without a college degree.

In line with the growing literature providing evidence for discrimination against Asian Americans, our findings question the seamless assimilation story often implied for this group (Ancheta 2006; Chou and Feagin 2008; C. Kim 1999; N. Kim 2007). Our categorization shows unfair treatment to be more common among the relatively advantaged groups, which is in

7. For the survey items used in the construction of the race-related variables, see online appendix C (https://www.rsfjournal.org/content/7/2/22/tab-supplemental).

line with critical race work exploring the specific experiences of racism among highly educated, (upper) middle-class Asians (Chou and Feagin 2008). Across our subgroups, reports of past discrimination among Asian Americans appear to increase among more settled and educated groups. This makes sense, as both factors are associated with increased expectations for equal treatment and perceptions of discrimination in past work (Banerjee 2008).

Relatedly, the lower-right panel shows that the hyper-selected and achieving—the most educated groups in our data—contain significantly higher shares of individuals (~70 percent) who have been stereotyped as model minority relative to the remaining three groups. The rooted and ordinary in turn include a significantly higher proportion (~50 percent) reporting this stereotype than the vulnerable—the group in which this stereotype is least common (~30 percent). We interpret this pattern as reflecting both emergent racialization as well as differential endowments in the social resources associated with the model minority stereotypes. That stereotype reporting is still common for the vulnerable and ordinary subgroups suggests that these stereotypes are generically applied to most people that can be identified as Asian in interpersonal encounters.

We find that experiences with unfair treatment do not go hand in hand with strong racial identification. The relatively advantaged groups (the hyper-selected, achieving, and rooted), who report experiencing discrimination at greater rates, are less likely to identify strongly with their racial category than the less advantaged (the vulnerable and ordinary). This contrasts with social psychological findings of "reactive racial identity" among African Americans in the United States—that is, higher in-group attachment as a result of experiencing stigma of the in-group (Branscombe, Schmitt, and Harvey 1999).

The lower left panel in figure 4 shows group-specific perceptions of a racially linked fate. These perceptions seem least prevalent among the vulnerable (~50 percent), moderately present in the ordinary, hyper-selected, and rooted groups (~57 percent), and slightly higher among the achieving (~64 percent). The relative ordering of the five groups on perceptions of a ra-

cially linked fate seems to be the mirror image of their ordering on strong racial identification. As the more settled groups report more discrimination, and weaker racial identity, they also solidify their sense of racially shared fate.

Our categorization confirms some existing observations (for example, prevalence of discrimination and the model minority stereotype among the highly educated), but also show new patterns (such as contradictory stances on racial identification and racially linked fates). These latter results are important: the decreasing significance of race on life chances—the key signal of assimilation into the mainstream as conceived in neoassimilation theory (Alba and Nee 2003, 12)—does not appear to be accompanied by a decreasing salience in the perceived weight of race in one's personal life. This decoupling between structural and subjective significance of race can be understood as part of "racialized assimilation" among Asian Americans (Lee and Kye 2016), in which increasing cultural-political embeddedness into the United States is not signaled by the strength of racial identification as much as the emergent sense of "us" as a racial group endowed with a collective destiny.

One might ask how our inductive classes fare relative to national-origin categories in explaining these race-related outcomes. That is, do we gain any explanatory power with our groupings? To answer this question, we undertake a simple exercise. For each of the four outcomes related to racial experience in figure 4, we estimate two analyses of variance models. These models help determine whether the means of included groups in the outcome are truly different. An F-test compares the variation between group means to the variation within groups to statistically test the equality of means (the null hypothesis).

The first set of models includes indicators for our five classes. For each of the four outcomes, the F-test allows us to reject the null hypothesis ($p < .05$). The second set of models includes indicators for national origins. Again, for each outcome, the F-test leads us to reject the equality-of-means hypothesis ($p < .05$). But, the important question for us is how the former models compare with the latter ones. Lacking formal statistical tests for this purpose, we rely

Figure 5. Comparison of Results from Analyses of Variance for Latent Classes and National Origins for Selected Outcomes
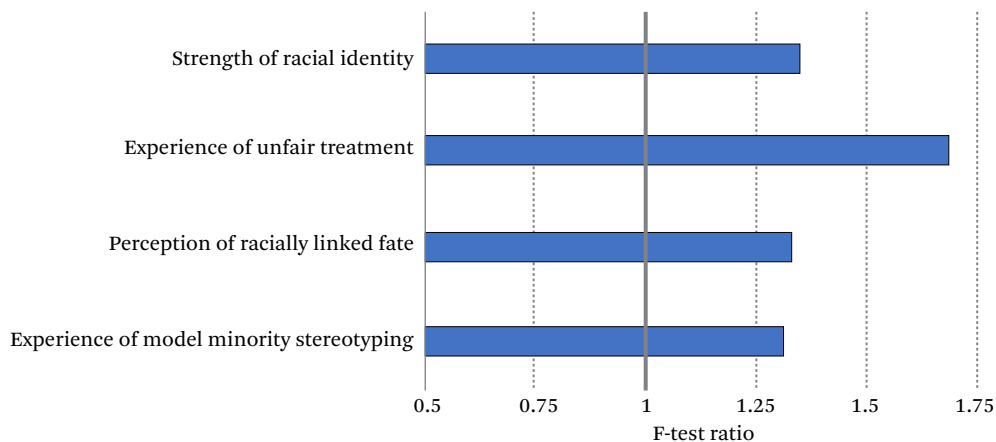
Source: Authors' tabulation based on National Asian American Survey data (Ramakrishnan et al. 2017).

on an F-test ratio, where we divide the F-test value from the first model by the same value from the second model. Figure 5 shows the results. For each outcome, the F-test ratio is substantially greater than 1, meaning that, in each case, the null hypothesis is more strongly rejected for the five classes than it is for the national-origin indicators. Although this is not strictly a statistical test (as we cannot assign confidence intervals to the F-test ratio), the results still suggest that the inductive categorization creates better differentiated groupings of Asian Americans in terms of their racial experiences in the United States.

Another plausible related criticism of our approach may be that we recover groups based on compositional variation and that such variation can be better accounted for with a more parsimonious linear model in which differences in the experience of race follow from differences in socioeconomic sources—that is, income and education. It is easy to counter that argument because our regression models establish an association between class membership and four race-related outcomes above and beyond compositional variation in socioeconomic resources between latent classes—as indicated by the statistical significance of the main terms for latent class membership. Ad-

ditionally, interaction terms show that the relationship of socioeconomic indicators to race-related outcomes vary by latent class. Substantively, this establishes that differences in the subjective experience of race among latent classes exist above and beyond compositional differences between classes in terms of socioeconomic resources.[8]

Intragroup Heterogeneity Matters for the Effect of Discrimination on Health, Political Behavior, and Panethnic Identity

The analysis established that the experiences of race among Asian Americans follows axes of patterned differentiation. We now illustrate the analytical benefits of subgroup-specific modeling relative to a pooled sample approach. We revisit the statistical relationships in the literature (documented or theorized) regarding the effect of discrimination on three outcomes. A large body of work—to which we cannot do justice given space constraints—finds a negative relationship between the experience of discrimination and self-reported health (Paradies et al. 2015). Studies also document a positive relationship between perceptions of unfair treatment and political participation that is presumed to work through increased group consciousness (Lien 2001; Ramakrishnan and

---

8. These analyses are available in online appendix B (https://www.rsfjournal.org/content/7/2/22/tab -supplemental).

Figure 6. Effect Sizes of the Experience of Discrimination on Three Outcomes
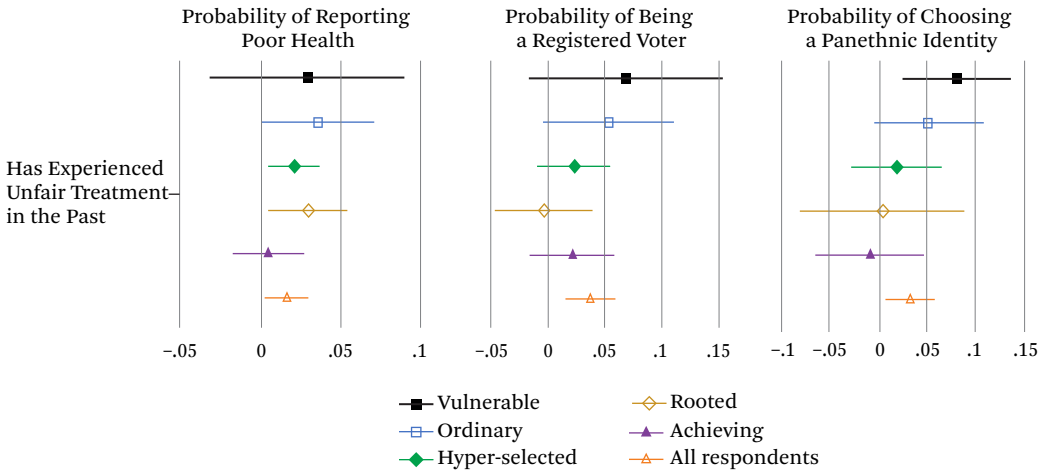
Source: Authors' tabulation based on National Asian American Survey data (Ramakrishnan et al. 2017).
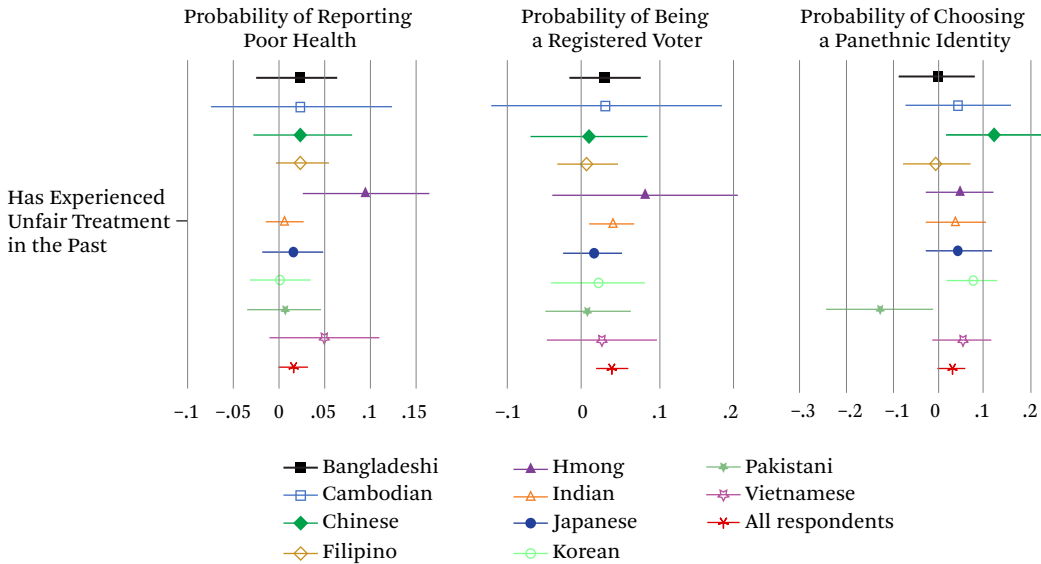Note: Sizes in percentage change.

Espenshade 2001). Finally, scholars theorize and observe a positive relationship between discrimination and panethnic identity (Masuoka 2006; Kibria 1998) among Asian Americans. We now test whether the link between discrimination and these outcomes varies across our five classes.

The three panels in figure 6 show results from logistic regressions of three outcomes—whether the respondent reported poor health, whether the respondent registered as a voter, and whether the respondent claimed a panethnic identity (such as Asian American rather than Chinese American in the case of a respondent of Chinese origins)—on the respondent's experience with unfair treatment. All models include controls for respondent's education, income, and gender. Each panel presents coefficient estimates (dots) with confidence intervals (lines) for the key variable (experience with discrimination) estimated on the pooled sample (bottom coefficient) as well as on samples containing each of our five classes.

The impact of discrimination on reporting poor health (left panel) is positive for all groups, but not statistically significant for the vulnerable and achieving. One temptation is to attribute the null effect of discrimination for the achieving to the group's high education and income. The hyper-selected and rooted, how-

ever, experience a negative impact of discrimination on health despite being similarly advantaged. The null effect for the vulnerable is also surprising, given the low status of this group and its presumed fragility. A key takeaway is the differential impact of discrimination across subgroups. The pooled coefficient reflects the situation of the ordinary, hyper-selected, and rooted respondents, but hides the absent (yet expected based on the literature) relationship for the other two groups.

The middle panel similarly suggests the potential for the pooled analysis to mask heterogeneous effects of discrimination on political participation. Although the effect appears positive in the overall sample, it is driven by the more recently settled groups. Given that the vast majority of Asian Americans are foreign born, this pattern indicates that the politicizing influence of discrimination decays with immigrant generation. Finally, similar observations apply to the right side panel. The pooled analysis suggests that unfair treatment pushes respondents to claim a panethnic identity, but group-specific models reveal that this positive effect is driven by the more recent and disadvantaged groups in the data, namely, the vulnerable and ordinary. For the remaining four groups, discrimination has no significant association with panethnic identification. It is

Figure 7. Effect Sizes of the Experience of Discrimination on Three Outcomes

Source: Authors' tabulation based on National Asian American Survey data (Ramakrishnan et al. 2017).
Note: Sizes in percentage change.

unlikely for the differences in significance between pooled and split-sample coefficients to follow from differences in statistical power. The negative effect of discrimination on health remains significant for the rooted, for example, which is much smaller than the other groups; meanwhile, the confidence intervals are large for the vulnerable (and overlapping with zero) despite the large size of the group. These differences therefore likely reflect the differentiated effect of discrimination across subgroups.

This analysis suggests the utility of our inductive categorization for clarifying the complex relationship of racial experiences (captured via discrimination here) to key integration outcomes (health and panethnic identification) and behaviors (registering as voters). The relationship—which is often presumed uniform across Asian Americans in comparative analyses with other racial groups like African Americans and Hispanics—varies substantially across groups occupying different structural positions. Inferences from parameter estimates from pooled samples can lead to misleading conclusions for the majority of Asian Americans. Conversely, inferences from highly visible groups such as those closest to the model mi-

nority stereotype can lead to similarly incorrect conclusions for the racial group as a whole— such as the hyper-selected subgroup for voting registration and panethnicity.

An analytical benefit of our typology (which relies on a relatively small number of inductively identified groups) is the preservation of statistical power for within-subgroup analyses and between-subgroup comparison. To illustrate this benefit, we replicated the analyses in figure 6, replacing our subgroups with national-origin categories.

Figure 7 shows the results. Many coefficients appear positive but lack significance. The high number of national-origin categories and relatively low number of respondents per category make it difficult to adjudicate if non-effects are truly non-effects, or if non-effects are a result of low statistical power. This issue, known as type 2 error, is much less salient with our approach, which creates larger categories cutting across national origins. Thus, in addition to theoretical concerns about methodological nationalism (Wimmer and Glick-Schiller 2003) and calls for reflexivity on ethnic and racial categories of analyses (Brubaker 2004), data-driven classifications facilitate subgroup comparison

and allow researchers to remain attentive to heterogeneity while retaining statistical power.

DISCUSSION: ASIAN AMERICANS—A DIVERSE GROUP BETWEEN ASSIMILATION AND RACIALIZATION

The characterization of the incorporation trajectories among Asian Americans has led to tensions among sociologists of migration— among those identifying the Asian American experience as one of rapid entrance into the American mainstream (Nee and Holbrow 2013; Drouhot and Nee 2019), others focusing on the continuing significance of race and characterizing the model minority narrative as a myth (N. Kim 2007; Chou and Feagin 2008), and yet others seeking a balanced view through the articulation of "racialized assimilation" (Lee and Kye 2016; see also Golash-Bosa 2006). We contend that these tensions map onto different methodological cultures. Much quantitative work evaluates assimilation among Asian Americans with large, aggregate samples and side-by-side comparisons with other racial groups, thus ruling out within-group heterogeneity a priori. Qualitative and ethnographic work attends to intragroup variation but does not offer generalizations beyond its rich descriptions of the case at hand. In this article, we seek to resolve the theoretical and methodological tensions by switching the focus asking how much assimilation or racialization occurs among Asian Americans as a whole, to who experiences what in this diverse population. We do so while avoiding methodological nationalism and thereby rejecting the assumption that heterogeneity among Asian Americans necessarily follows ethnic lines.[9]

Our first empirical contribution is to identify patterned differentiation—the nonintuitive, yet nonrandom configurations of social attributes making up distinct types—among Asian Americans. Using latent class analysis, we uncovered five major, clearly differentiated subgroups occurring at the intersection of class, gender, regional location, and immigrant generation: vulnerable, ordinary, hyperselected, rooted, and achieving Asian Americans, and showed that this typology captures heterogeneity in the experience of race in a manner that is more inductive and parsimonious than a grouping based on national origins.

In turn, comparing the subjective experience of race among these subgroups leads to a second empirical contribution. Rather than forcing a choice between assimilation and the opposing perspective emphasizing the enduring weight of race, our approach allowed us to combine both perspectives, and thus pointed to a productive way forward. The numeric dominance of three groups (hyper-selected, rooted, and achieving) in our data, which make up more than two-thirds of our sample and conform most closely to the model minority stereotype, helps explain the empirical grounding of the assimilation narrative in past work documenting average trends for Asian Americans in samples aggregated by race (Sakamoto, Goyette, and Kim 2009; Drouhot and Nee 2019; Nee and Holbrow 2013). Yet we saw that these groups are also the ones reporting higher levels of discrimination, and a higher sense of racially linked fate, without necessarily reporting stronger attachment to their racial identities. These findings are in line with past work from a critical race tradition, which investigates the subjective experience of race and belonging among upwardly mobile Asian Americans who need to negotiate social spaces dominated by Whites (Chou and Feagin 2008; Dhingra 2007). Among our subgroups, those closer to the assimilation pattern are in fact those most affected by a sense of subjective racialization. The assimilation and critical race perspectives may therefore not be as contradictory as typically thought; rather, it might be precisely because many Asian American are structurally (in terms of socioeconomic status) successful—as emphasized in the neoassimilation perspective—that they experience a subjective sense of racializa-

9. In doing so, we do not claim to invalidate national origins as a way to apprehend within-group diversity among Asian Americans. National origins surely capture a substantial amount of differentiation within Asian Americans (for an empirical study of ethnic heterogeneity in public opinion among Asian Americans, see S. Kim 2021, this issue). However, we argue that national origins is only one way to capture intragroup heterogeneity and not a particularly reflexive and parsimonious one.

tion—in the form of higher perception of discrimination and a sense of racially linked fate—as they come into social spaces dominated by the White majority group. This interpretation is also in line with ethnographic and interview-based studies of upwardly mobile Black immigrants (Vickerman 1999). Supplementary analyses reveal that subgroup membership is significantly associated with such race-related outcomes (perceptions of discrimination), even after we control for socioeconomic indicators at the individual level (see online appendix B). In other words, these subgroups capture distinct aspects of the subjective experience of race that are not attributable to socioeconomic status differences alone.

Although Asian Americans' life chances are not structurally shaped by their racial category in the post–civil rights era (Drouhot and Nee 2019; Nee and Holbrow 2013), their subjective racial experience and sense of belonging are. Our results indicate a decoupling between the structural aspects of assimilation (relating to upward mobility) and symbolic aspects such as perceptions of racially linked fate. These findings counter canonical accounts, such as Gordon's (1964) multistep model of assimilation, which conceive of identification and a shared sense of "peoplehood" as proceeding from immigrant economic integration. We show that immigrants and their children may continue to experience subjective forms of stigma and racialized group consciousness in spite of high socioeconomic attainment. This reconciliation of assimilation as socioeconomic attainment and the subjective salience of race complements recent descriptions of "racialized assimilation" among Asian Americans (Lee and Kye 2016). Together, these findings suggest avenues for renewed theoretical work on the relationships between assimilation, immigrant upward mobility, and the subjective experience of race.

Bridging Theory and Empirics with Data-Driven Classification Methods

In this article, we implemented an empirical approach attentive to long-standing concerns for within-group heterogeneity in several strands of migration-focused theorizing, such as those on segmented assimilation, and superdiversity (Portes and Zhou 1993; Vertovec 2007).

Although empirical descriptions of heterogeneity and within-group differentiation are a mainstay of qualitative work, they are not as common in quantitative work. The general linear reality and average-case focus of regression-based approaches (Abbott 1988) flattens the social structure of the ethnoracial groups under study, and cumulatively contributes to their reification when research findings are assessed and groups compared side by side. The polarization of research between generalized findings from aggregate data and average case–based method, on the one hand, and thicker, qualitative description from smaller and ungeneralizable samples, on the other, contributes to maintaining an undesirable intellectual stalemate between potentially complementary approaches. Data-driven classification methods provide an avenue for bringing together the search for broad patterns while remaining attentive to subgroup-specific processes and intragroup heterogeneity.

In the Asian American case, we identify five rather large subgroups that cross-cut ten national-origin groups and therefore preserve enough statistical power for meaningful subgroup comparison, as well as within-subgroup modeling. Similar comparisons across national-origin groups were not fruitful given small sample sizes for each group. Beyond addressing a theoretical concern for heterogeneity, our analyses demonstrated the analytical benefit of comparing statistical relationships across subgroups. Specifically, we show that the effect of reported experience of unfair treatment has a differential impact on various outcomes across the five subgroups under the Asian American category. This finding stands in stark contrast to work estimating a single coefficient (assumed to apply uniformly to the entire sample at hand) for the effect of unfair treatment on outcomes such as reported health and panethnic identity. For the latter outcome, our analyses shows that a positive and significant coefficient obtained for a category-level (pooling all Asian Americans) analysis is entirely carried by a very large coefficient for one subgroup—the "vulnerable"—and is not replicated for the other four subgroups, for whom perceived unfair treatment has no sizable effect on the probability of reporting a strong panethnic identity.

Being rooted in historical patterns of exclusion and inclusion and being imprinted on the law, racial and national-origin categories are powerful, naturalized, and cognitively salient—among the general public and scholars alike. Quite simply, the "groupness" of Asian Americans is real in part because ordinary individuals, taking cues from venerable institutions such as the U.S. Census, think it is real. To be clear, we do not interpret our approach and results as suggesting that racial categories such as Asian American are imaginary or artificial. What motivated our analyses is precisely the opposite: the cognitive pull of racial categories is such that it warrants reflexivity and attention to heterogeneity in scholarly analyses involving their application. In that regard, data-driven classification methods such as LCA offer migration researchers possibilities to produce an epistemological break from "categories of practice"—categories from everyday life such as those serving bureaucratic purposes (for example, in the census)—and create inductive and reflexive "categories of analysis" (Brubaker 2004). As Rogers Brubaker (2013, 2) notes, "the heavy traffic between the two, in both directions, means that we risk using pre-constructed categories of journalistic, political or religious common sense as our categories of analysis." Simply put, dividing a sample of immigrants by national origins or racial categories is already assuming that national origins or race matter and should organize one's look at the data. In our case, this amounts to naturalizing the analytical and empirical relevance of racial (Asian American) and national-origin (such as Chinese) categories when taking them as categories of analysis. However, both racial and national-origin categories are primarily categories of practice intimately linked to state bureaucracies, minority activist politics, and national projects whose very interests reside in the naturalization of such categories as principles for the "vision and division of the world" (Brubaker 2004, 2013; Wimmer 2009; Wimmer and Glick-Schiller 2003).

In this article, we implement a novel approach to studying Asian Americans, a particular immigrant group, "without groupism" (Brubaker 2004)—that is, without reifying neither race as a natural boundary to the processes affecting those self-identifying as Asian Americans, nor the ethnic categories existing beneath this racial label. Data-driven classification methods offer a promising avenue to implement such broad theoretical concerns for the adoption of sound categories of analysis without unwittingly baking racial or ethnic groupism in our research designs (Wimmer 2013). We hope our study offers a blueprint for future analyses attentive to such analytical and epistemological problems. Future research could go further than what has been presented here by pooling samples across racial groups, and letting racial or ethnic differentiation emerge from the data rather than imposing it a priori in the research design and research question formulation stage.

APPENDIX: TECHNICAL ASPECTS OF LATENT CLASS ANALYSIS WITH STOCHASTIC ASSIGNMENT

LCA estimates a latent (that is, unobserved) variable that accounts for the covariance between the observed attributes. This latent variable is assumed to have a categorical distribution with each value corresponding to a "latent class" or a group in data. An LCA model with observed response items u has a categorical latent (unobserved) variable c with k classes. Formally:

$$P(u_j = 1) = \Sigma_{K=1}^{K} P(c = k) \, P(u_j = 1 \mid c = k). \quad (1)$$

Equation (1) yields the marginal item probability for item u_j. For s number of response items to be clustered upon, the class probabilities of each individual respondent—the so-called posterior probabilities—are given by:

$$P_{ik} = P(c = k \mid u_1, u_2, \dots, u_s) = \frac{P(c = k) \, P(u_1 \mid c = k) \, P(u_2 \mid c = k) \dots P(u_s \mid c = k)}{P(u_1, u_2, \dots, u_s)}. \quad (2)$$

A common practice in LCA applications is to use a so-called modal assignment—that is, to assign cases to classes for which they have the highest posterior probability of belonging. In our view, this remains too close to the deterministic assignment issue associated with hard clustering, in which class membership is treated as an exact, observed variable. Because

modal assignment effectively erases assignment ambiguity for cases that have large probabilities of belonging to multiple classes (such as a set of posterior probabilities of 0.43, 0.37 and 0.20 belonging to three classes), standard errors obtained in post-assignment analysis can be deceptively small, and inferences from such analysis potentially erroneous (Clark and Muthén 2009). Rather than a statistical or analytical nuisance, we regard assignment uncertainty as meaningful as it reflects the blurry boundaries between ideal-typical subgroups making up the Asian American category. The so-called three-step approach to LCA calls for correcting for classification error as it can lead to underestimation of the relationships between obtained classes and other covariates in the third step (Vermunt 2010; Bakk et al. 2013). We, though, choose to rely on ambiguous assignment (without the correction for classification errors) to obtain a more conservative test in subsequent investigations of differences across subgroups.[10]

For our class assignment to be probabilistic and take membership uncertainty into account, we implement a stochastic assignment procedure in which a case's class is randomly drawn from the distribution of the posterior probabilities. Formally, respondent j's assignment A is thus given by:

$$A_i \sim Multinomial(\boldsymbol{p}_i), \text{ where } A_i \in \{1, 2, \dots, k\}, \quad (3)$$

where

$$\boldsymbol{p}_i : \{p_{i1}, p_{i2}, \dots, p_{ik}\},$$

$\boldsymbol{p}$ being respondent i's distribution of posterior probabilities. We estimated latent classes with the *lclass* command in Stata 15. In a second step, we used the posterior probabilities to assign individuals to a class based on a random draw using the *Hmisc::rMultinom* command in *R*. We analyzed the resulting classes back in Stata 15.

REFERENCES

Abbott, Andrew. 1988. "Transcending General Linear Reality." *Sociological Theory* 6(2): 169–86.

Alba, Richard, and Victor Nee. 2003. *Remaking the American Mainstream: Assimilation and Contemporary Immigration.* Cambridge, Mass.: Harvard University Press.

Ancheta, Angelo N. 2006. *Race, Rights, and the Asian American Experience.* New Brunswick, N.J.: Rutgers University Press.

Bakk, Zsuzsa, Fetene B .Tekle, Jeroen K. Vermunt. 2013. "Estimating the Association Between Latent Class Membership and External Variables Using Bias-Adjusted Three-Step Approaches." *Sociological Methodology* 43(1): 272–311.

Banerjee, Rupa. 2008. "An Examination of Factors Affecting Perception of Workplace Discrimination." *Journal of Labor Research* 29: Article 380. DOI: 10.1007/s12122-008-9047-0.

Blau, Peter M. 1974. "Parameters of Social Structure." *American Sociological Review* 39(5): 615–35.

Bonacich, Edna, and John Modell. 1980. *The Economic Basis of Ethnic Solidarity: Small Business in the Japanese American Community.* Berkeley: University of California Press.

Bonikowski, Bart, and Paul DiMaggio. 2016. "Varieties of American Popular Nationalism." *American Sociological Review* 81(5): 949–80.

Branscombe, Nyla R., Michael T. Schmitt, and Richard D. Harvey. 1999. "Perceiving Pervasive Discrimination Among African Americans: Implications for Group Identification and Well-Being." *Journal of Personality and Social Psychology* 77(1): 135–49.

Brubaker, Rogers. 2004. *Ethnicity Without Groups.* Cambridge, Mass.: Harvard University Press.

———. 2013. "Categories of Analysis and Categories

---

10. Other extensions of LCA allow researchers to consider the local dependence between the survey responses (see S. Kim 2021, this issue). This approach is useful when researchers consider many highly correlated variables (for example, based on attitudinal questions trying to get at the same construct), but is not necessary in our case given the few (five) dimensions we include in analysis. Another extension allows researchers to include additional variables (which are not included in the identification of latent classes) during the phase of assigning individual cases to classes. As we employ all variables of interest in the first phase (identification of latent classes), this extension is not relevant for our purposes.

of Practice: A Note on the Study of Muslims in European Countries of Immigration." *Ethnic and Racial Studies* 36(1): 1–8.

Chou, Rosalind S., and Joe R. Feagin. 2008. *The Myth of the Model Minority: Asian Americans Facing Racism*. Boulder, Colo.: Paradigm.

Clark, Shaunna L., and Bengt Muthén. 2009. "Relating Latent Class Analysis Results to Variables Not Included in the Analysis." Unpublished paper. Accessed October 14, 2020. http://www.statmodel.com/download/relatinglca.pdf.

Dhingra, Pawan. 2007. *Managing Multicultural Lives: Asian American Professionals and the Challenges of Multiple Identities*. Stanford, Calif.: Stanford University Press.

Donato, Katherine, Donna Gabaccia, Jennifer Holdaway, Martin Manalasan IV, and Patricia Pessar. 2006. "A Glass Half Full? Gender in Migration Studies." *International Migration Review* 40(1): 3–26.

Drouhot, Lucas G. Forthcoming. "Cracks in the Melting Pot? Religiosity and Assimilation Among the Diverse Muslim Population in France." *American Journal of Sociology*.

Drouhot, Lucas G., and Victor Nee. 2019. "Assimilation and the Second Generation in Europe and America: Blending and Segregating Social Dynamics Between Immigrants and Natives." *Annual Review of Sociology* 45(1): 177–99.

FitzGerald, David Scott. 2014. "The Sociology of International Migration." In *Migration Theory: Theory Across Disciplines*, edited by Caroline B. Bretell and James F. Hollifield. New York: Routledge.

Garip, Filiz. 2012. "Discovering Diverse Mechanisms of Migration: The Mexico-US Stream 1970–2000." *Population and Development Review* 38(3): 393–433.

———. 2016. *On the Move: Changing Mechanisms of Mexico-U.S. Migration*. Princeton, N.J.: Princeton University Press.

Gee, Gilbert C., Annie Ro, Salma Shariff-Marco, and David Chae. 2009. "Racial Discrimination and Health Among Asian Americans: Evidence, Assessment, and Directions for Future Research." *Epidemiologic Reviews* 31(1): 130–51.

Golash-Bosa, Tanya. 2006. "Dropping the Hyphen? Becoming Latino(a)-American Through Racialized Assimilation" *Social Forces* 85(1): 27–55.

Gordon, Linda W. 1987. "Southeast Asian Migration to the United States." In *Pacific Bridges: The New Immigration from Asia and the Pacific Islands*, edited by James T. Faucett and Benjamin V. Cariño. Staten Island, N.Y.: Center for Migration Studies.

Gordon, Milton. 1964. *Assimilation in American Life: The Role of Race, Religion and National Origins.* New York: Oxford University Press.

Hirschmann, Daniel. 2016. "Stylized Facts in the Social Sciences." *Sociological Science* 3(30): 604–26.

Hsin, Amy, and Yu Xie. 2014. "Explaining Asian Americans' Academic Advantage over Whites." *Proceedings of the National Academy of Sciences* 111(23): 8416–21.

Jiménez, Tomás R., and Adam L. Horowitz. 2013. "When White Is Just Alright: How Immigrants Redefine Achievement and Reconfigure the Ethnoracial Hierarchy." *American Sociological Review* 78(5): 849–71.

Kao, Grace, and Jennifer S. Thompson. 2003. "Racial and Ethnic Stratification in Educational Achievement and Attainment." *Annual Review of Sociology* 23(1): 417–42.

Kasinitz, Philip, Mary Waters, John H. Mollenkopf, and Jennifer Holdaway. 2008. *Inheriting the City: The Children of Immigrants Come of Age*. New York: Russell Sage Foundation.

Kibria, Mazli. 1998. "The Contested Meanings of 'Asian American': Racial Dilemmas in the Contemporary US." *Ethnic and Racial Studies* 21(5): 939–58.

Kim, ChangHwan, and Arthur Sakamoto. 2010. "Have Asian American Men Achieved Labor Market Parity with White Men?" *American Sociological Review* 75(6): 934–57.

Kim, Claire J. 1999. "The Racial Triangulation of Asian Americans." *Politics & Society* 27(1): 105–38.

Kim, Nadia Y. 2007. "Critical Thoughts on Asian American Assimilation in the Whitening Literature." *Social Forces* 86(2): 561–74.

Kim, Sunmin. 2020. "Fault Lines Among Asian Americans: Pan-Ethnic Identity, Policy Opinions, and Potential for Political Mobilization." Unpublished manuscript. Dartmouth College, Department of Sociology.

———. 2021. "Fault Lines Among Asian Americans: Convergence and Divergence in Policy Opinion." *RSF: The Russell Sage Foundation Journal of the*

Social Sciences 7(2): 46–67. DOI: 10.7758/RSF .2021.7.2.03.

Krogstad, Jens Manuel, and Jynnah Radford. 2018. "Education Levels of U.S. Immigrants Are on the Rise." *Fact Tank* (Pew Research Center blog), September 14. Accessed January 25, 2021. https://www.pewresearch.org/fact-tank/2018 /09/14/education-levels-of-u-s-immigrants-are -on-the-rise/.

Le Espiritu, Yen. 1992. *Asian American Panethnicity: Bridging Institutions and Identities.* Philadelphia, Pa.: Temple University Press.

Lee, Jennifer C., and Samuel Kye. 2016. "Racialized Assimilation of Asian Americans." *Annual Review of Sociology* 42(1): 253–73. DOI:10.1146/annurev -soc-081715-074310.

Lee, Jennifer, and Min Zhou. 2015. *The Asian American Achievement Paradox.* New York: Russell Sage Foundation.

Lee, Stacey J. 1996. *Unraveling the Model Minority Stereotype: Listening to Asian American Youth.* New York: Teachers College Press.

Li, Wei. 2009. *Ethnoburb: The New Ethnic Community in America.* Honolulu: University of Hawai'i Press.

Lien, Pei-te. 2001. *The Making of Asian America Through Political Participation.* Philadelphia, Pa.: Temple University Press.

Lin, Monica H., Virginia S.Y. Kwan, Anna Cheung, and Susan T. Fiske. 2005. "Stereotype Content Model Explains Prejudice for an Envied Outgroup: Scale of Anti-Asian American Stereotypes." *Personality and Social Psychology Bulletin* 31(1): 34–47.

Magidson, Jay, and Jeroen K. Vermunt. 2002. "Latent Class Models for Clustering: A Comparison with K-Means." *Canadian Journal of Marketing Research* 20(1): 37–44.

Masuoka, Nathalie. 2006. "Together They Become One: Examining the Predictors of Panethnic Group Consciousness Among Asian Americans and Latinos." *Social Science Quarterly* 87(5): 993–1011.

Matsumoto, Noriko. 2018. *Beyond the City and the Bridge: East Asian Immigration in a New Jersey Suburb.* New Brunswick, N.J.: Rutgers University Press.

McCall, Leslie. 2005. "The Complexity of Intersectionality." *Signs: Journal of Women in Culture and Society* 30(3): 1771–800.

Min, Pyong Gap, ed. 2002. *The Second Generation:*

Ethnic Identity Among Asian Americans. Walnut Creek, Calif.: AltaMira Press.

Min, Pyong Gap, and Chigon Kim. 2009. "Patterns of Intermarriages and Cross-Generational In-Marriages Among Native-Born Asian Americans." *International Migration Review* 43(3): 447–70.

Molina, Mario, and Filiz Garip. 2019. "Machine Learning for Sociology." *Annual Review of Sociology* 45(1): 27–45.

Museus, Samuel D., and Peter Kiang. 2009. "Deconstructing the Model Minority Myth and How It Contributes to the Invisible Minority Reality in Higher Education Research." *New Directions for Institutional Research* 142(1): 5–15.

Nee, Victor, and Hilary J. Holbrow. 2013. "Why Asian Americans Are Becoming Mainstream." *Daedalus* 142(3): 65–75.

Ngo, Bic, and Stacey J. Lee. 2007. "Complicating the Image of Model Minority Success: A Review of Southeast Asian American Education." *Review of Educational Research* 77(4): 415–53.

Okamoto, Dina G. 2003. "Toward a Theory of Panethnicity: Explaining Asian American Collective Action." *American Sociological Review* 68(6): 811–42.

———. 2007. "Marrying Out: A Boundary Approach to Understanding the Marital Integration of Asian Americans." *Social Science Research* 36(4): 1391–414.

Paradies, Yin, Jehonathan Ben, Nida Denson, Amanuel Elias, Naomi Priest, Alex Pieterse, Arpana Gupta, Margaret Kelaher, and Gilbert Gee. 2015. "Racism as a Determinant of Health: A Systematic Review and Meta-Analysis." *PLoS ONE* 10(9): e0138511.

Pew Research Center. 2012. *The Rise of Asian Americans.* Washington, D.C.: Pew Research Center. Accessed October 14, 2020. https://www .pewsocialtrends.org/2012/06/19/the-rise-of -asian-americans.

Portes, Alejandro, and Min Zhou. 1993. "The New Second Generation: Segmented Assimilation and Its Variants." *Annals of the American Academy of Political and Social Science* 530(1): 74–96.

Qian, Zenchao, and Daniel T. Lichter. 2011. "Changing Patterns of Interracial Marriage in a Multiracial Society." *Journal of Marriage and Family* 73(5): 1065–84.

Qian, Zhenchao, Sampson Lee Blair, and Stacey D. Ruf. 2001. "Asian American Interracial and Inter-

ethnic Marriages: Differences by Education and Nativity." *International Migration Review* 35(2): 557–86.

Ramakrishnan, Karthick, and Thomas J. Espenshade. 2001. "Immigrant Incorporation and Political Participation in the United States." *International Migration Review* 35(3): 870–909.

Ramakrishnan, Karthick, Jennifer Lee, Taeku Lee, and Janelle Wong. 2017. "National Asian American Survey (NAAS) 2016 Pre-Election Survey." Riverside, Calif.: National Asian American Survey. 2017-12-05.

Roy, Dharmendra K., and Lokesh K. Sharma. 2010. "Genetic k-Means Clustering Algorithm For Mixed Numeric and Categorical Data Sets." *International Journal of Artificial Intelligence & Applications* 1(2): 23–28.

Sakamoto, Arthur, Kimberly A. Goyette, and Chang Hwan Kim. 2009. "Socioeconomic Attainments of Asian Americans." *Annual Review of Sociology* 35(1): 255–76.

Small, Mario Luis, David J. Harding, and Michèle Lamont. 2010. "Reconsidering Culture and Poverty." *Annals of the American Academy of Political and Social Science* 629(1): 6–27.

Srinivasan, Shobha, and Tessie Guillermo. 2000. "Toward Improved Health: Disaggregating Asian American and Native Hawaiian/ Pacific Islander Data." *American Journal of Public Health* 90(11): 1731–34.

Sue, Christina A., Fernando Riosmena, and Joshua Lepree. 2019. "The Influence of Social Networks, Social Capital, and the Ethnic Community on the U.S. Destination Choices of Mexican Migrant Men." *Journal of Ethnic and Migration Studies* 45(13): 2468–88.

Sue, Derald Wing, Jennifer Bucceri, Annie I. Lin, Kevin L. Nadal, and Gina C. Torino. 2007. "Racial Microaggressions and the Asian American Experience." *Cultural Diversity and Ethnic Minority Psychology* 13(1): 72–81.

Takaki, Ronald. 1998. *Strangers from a Different Shore: A History of Asian Americans*. New York: Little, Brown.

Varma, Roli. 2004. "Asian Americans: Achievements Mask Challenges." *Asian Journal of Social Science* 32(2): 290–307.

Vickerman, Milton. 1999. *Crosscurrents: West Indian Immigrants and Race*. New York: Oxford University Press.

Vermunt, Jeroen. 2010. "Latent Class Modeling with Covariates: Two Improved Three-Step Approaches." *Political Analysis* 18(4): 450–69.

Vertovec, Steven. 2007. "Super-Diversity and Its Implications." *Ethnic and Racial Studies* 30(6): 1024–54.

Waters, Mary C., and Karl Eschbach. 1995. "Immigration and Ethnic and Racial Inequality in the United States." *Annual Review of Sociology* 21(1): 419–46.

Waters, Mary C., and Tomás R. Jiménez. 2005. "Assessing Immigrant Assimilation: New Empirical and Theoretical Challenges." *Annual Review of Sociology* 31(1): 105–25.

Wimmer, Andreas. 2009. "Herder's Heritage and the Boundary-Making Approach: Studying Ethnicity in Immigrant Societies." *Sociological Theory* 27(3): 244–70.

———. 2013. *Ethnic Boundary Making: Institutions, Power, Networks.* Oxford: Oxford University Press.

Wimmer, Andreas, and Nina Glick-Schiller. 2003. "Methodological Nationalism, the Social Sciences, and the Study of Migration: An Essay in Historical Epistemology." *International Migration Review* 37(3): 576–610.

Zhou, Min, and Roberto Gonzales. 2019. "Divergent Destinies: Children of Immigrants Growing Up in the United States." *Annual Review of Sociology* 45(1): 383–99.

Zhou, Min, and Yang Sao Xiong. 2005. "The Multifaceted American Experiences of the Children of Asian Immigrants: Lessons for Segmented Assimilation." *Ethnic and Racial Studies* 28(6): 1119–152.