The Many Dimensions of Race:
Capturing Complexity with Latent Variables*

Aliya Saperstein
University of California, Berkeley
asaper@berkeley.edu

Abstract

Typically, race is measured through self-identification and included in quantitative research as a set of "dummy" variables that serve as controls in regression-type analyses. This implies several strong assumptions, including that race is an intrinsic characteristic of individuals and that it can be described by mutually exclusive and exhaustive categories. I propose a new approach to analyzing race in quantitative research in the hopes of bringing research practice closer to current social science theories of race. Using a statistical technique called latent class analysis, I combine measures of observed race, self-reported race and/or self-reported ethnic origins into a single variable that better captures the complexity of an individual's experience of race and ethnicity in the United States. The latent variable consists of multiple racial classes, or categories, and can be used in subsequent studies as either a dependent or independent variable.

*Paper prepared for the 2007 Population Association of America Annual Meeting in New York, March 29-31. *Please do not quote or cite without permission.* Many thanks to Mike Hout and the members of the Spring 2007 Berkeley Demography Dissertation Seminar for their helpful comments and suggestions.

For the past 15 years, disciplines from medicine to sociology have debated the proper way to collect and use racial data for the purposes of social research (e.g., Kaplan and Bennett 2003; Stephan and Stephan 2000; Yee et al. 1993). During the same period, public debate has revolved around how to count multiracial Americans, and whether or not Americans should be counted by race at all. Though official position papers that affirm the importance of collecting data on race or ethnicity have been published by the American Sociological Association (2003) and the American Anthropological Association (1997), the only real consensus about research practices across disciplines is that using race in social research is a tricky business.[1]

At the same time, the growing ease of computing regression analyses and the rise in the number of national surveys with data on multiple racial groups has prompted a dramatic increase in the use of race in U.S. sociological studies (Martin and Yeung 2003). However, this trend does not necessarily represent a shift toward tracking the racial dimensions of inequality – which is the usual explanation for why government entities in the United States collect data on race. Martin and Yeung (2003) find that race is "taken into account" in most studies in a very cursory way: by including dummy variables for racial groups as controls. They note that these variables are almost always coded with "White" as the reference group, and argue that this use of race is often an attempt to "deracialize" findings -- cleansing them of inconvenient variation to better describe the process or relationships of interest.

The phrase "race is a social construction" has become a mantra in social science classrooms across the country, but its implications for research practice remain undeveloped. Few alternative strategies for including race in quantitative research have been proposed beyond

[1] In this paper, I generally use the terms "race" and "ethnicity" interchangeably. I acknowledge there are wide-ranging debates about distinctions between the two terms, however I link them here, in part, due to constraints in the data I analyzed. I also argue that for studies of inequality making an analytical distinction between race and ethnicity is largely semantic. Both race and ethnicity are maintained through processes of identification and ascription, and both have conditioned the distribution of societal resources, in different places and eras.

superficial changes to question wordings and category options, despite the periodic calls for more "deliberate and reflexive" methods (Martin and Yeung 2003: 540), and a wealth of studies demonstrating what is lost in current methods of measuring race (e.g., Harris and Sim 2002, Telles 2002, Brown et al. 2006, Saperstein 2006). Even the highly touted data from Census 2000, which includes multiple mentions of self-reported race for the first time, is for the most part being analyzed using the standard mutually exclusive dummy variable method.

I propose a new approach to modeling race, using latent variables and measures of race from multiple sources, in the hopes of both bringing research practice closer to social science theories of race and advancing research on persistent racial inequality in the United States. Typical single-measure methods continue to treat race as if it were solely an intrinsic, biological characteristic that is causally prior to an individual's life outcomes instead of a complex and contested status hierarchy that varies over time and across contexts. If racialization occurs through interaction between individuals, organizations and the state—as social science theories of race suggest (e.g., Nagel 1994; Omi and Winant 1994)—then I argue research attempting to untangle the processes behind persistent racial inequality requires multiple measures of race from multiple and potentially contradictory perspectives, such as known ancestry, self-identity and observed classification.

The goal of this paper is to demonstrate concretely what a latent variable approach to race entails and what the U.S. racial landscape looks like once we take into account the interaction between observed and self-reported measures of race. I touch briefly on the theoretical and methodological justifications for my approach below; for a more in-depth discussion of my argument see Saperstein (2007). Here, I use latent class analysis and previously unanalyzed data from the National Survey of Family Growth (NSFG) to combine the respondent's race as observed by the survey interviewer and her self-reported "origin" or "racial background" into a single racial latent variable. I estimate a series of latent class measurement models, from the most basic two-class model to models with as many as eight classes, in order to determine the "best-fitting" model for each of the four survey years, which range from 1973 to 1988. I find that in no year can the race of NSFG respondents be described by fewer than three latent classes, and that all models with acceptable fit statistics include at least one class whose members have completely inconsistent racial identities and classifications. These results support my argument that including observed racial classifications when using race in quantitative research helps to identify meaningful and previously hidden variation in the American racial landscape.

Why we can't study identity alone

Two previous studies have taken a similar latent variable approach to modeling race or ethnicity in the United States (Johnson 1990, Macmillan and Liebler 2005), but both contain the same limitation: they rely solely on self-reported measures of race and ethnicity. Both papers argue, as I do, that there are many dimensions to race that conventional methods do not take into account, and each uses latent class analysis to make new and interesting contributions to our understanding of race and how the reporting of racial identities varies across the U.S. population.[2] But neither allows for those dimensions of race to be defined by anyone other than the individual in question. This is a serious limitation given recent research showing that it really

[2] Macmillan and Liebler (2005) reveal interesting patterns of racial identification in trying to make sense of the 128 possible combinations provided by the multiple mention race data from Census 2000. Johnson (1990) demonstrates that the complexity in repeated measures of Hispanic ancestry, ethnicity and identity belie Census Bureau statistical assumptions of simple response bias and independent and identically distributed measures.

matters who answers the race question (as well as where, when, and in front of whom they answer it).[3]

Numerous studies over the past 15 years document how racial classifications made by hospital personnel, coroners and other similar administrative sources can differ dramatically both among themselves and from self reports or the reports of family members (e.g., Hahn et al. 1992, Sugarman et al. 1993). Other studies of the identification of multiracial children show that parents of different races do not necessarily report their biological children as the exact combination of their parents' races, even when explicitly given the option to do so (Roth 2005, Tafoya, Johnson and Hill 2005, Paret and Saperstein 2006).[4] And when forced to choose a single race for their multiracial children, Xie and Goyette (1997) find that the parents' choice is not a simple function of surname or the gender of the minority parent, but is partly explained by personal characteristics (e.g., immigrant status) and partly by contextual ones (e.g., the racial composition of their neighborhood). Further, one of the first quantitative studies to look at how multiracial teens describe themselves finds that the self-reports vary, sometimes quite dramatically, by whether the children were asked at home or at school, as well as by how old they were and the racial composition of their neighborhood (Harris and Sim 2002).

Few studies have taken the next step to examine whether using these different measures of race lead us to equally different research conclusions about racial inequality, but those that do find that the two sets of results support substantively different conclusions about the degree of inequality in the outcome of interest. For example, Saperstein (2006) uses two different measures of race (self-report vs. interviewer classification) in otherwise similar multivariate regressions

[3] Reliance on self-reported measures of race is, of course, a limitation of using post-1960 U.S. census data in general.
[4] Lieberson and Waters (1993) find similar discrepancies with respect to reporting mixed ethnic ancestry among whites.

estimating income inequality in the United States. Using self-reports, the resulting economic rankings look like an even progression from black, to Hispanic and Asian, to white. However, using interviewer classifications, the picture of economic inequality in the contemporary U.S. better fits a black-nonblack divide first forecast by Gans (1999). Telles also reports discrepant findings regarding the severity of income inequality when comparing observed and self-reported measures of race in Brazil (Telles and Lim 1998; Telles 2002).

Even without such evidence, simple causal logic suggests that if the purpose of measuring race is to track racial inequality, then self-report data is only one aspect of the phenomenon in which we are interested. Using the typical single measure of self-reported race and including it in a regression analysis as a set of binary "dummy" variables is subject to the now-typical critique that such a method assumes the racial categories (or identities) are mutually exclusive, exhaustive, fixed and objective. However, using a self-reported measure of race in a study of racial inequality also implies that group differences in, say, earned income, can be explained by either which racial identity an individual felt closest to at the time of the survey or what one knows about one's ancestry. The tautology in such an argument would be readily apparent were it made explicit.[5]

Of course, we also should not revert back to the time when imposed racial classifications modeled on the taxonomy of species in the biological sciences drove pseudo-scientific findings about racial differences in everything from fertility to intelligence (see Gossett 1997 for a detailed history). The political decision to allow individuals to define themselves racially in the U.S. census and other government documents after 1960 was an appropriate response to a sordid

[5] As an example for explaining income equality or socioeconomic status more generally, the tautology stems from the positive association between educational attainment and reporting multiple ethnic (Lieberson and Waters 1993) or racial (Goldstein and Morning 2000) ancestries.

history of using official racial classifications to segregate and subjugate minority groups;[6] however, it does not follow that using self-identification alone is the best option for research into racial inequality in the United States. Individual aspirations and behavior may help explain some of the persistent racial disparities in income, political participation and health, but other factors such as how an individual is perceived by teachers, employers, doctors and other authority figures also play a role (e.g., Ferguson 1998, Kirschenman and Neckerman 1991). Current single-measure methods confound choices and constraints by failing to separate individual agency from institutional discrimination.

Multiple measures of race, including observed classifications or proxies for them, already exist in several national survey datasets in addition to the NSFG data I use here.[7] However, their full potential is not being exploited. To date, survey researchers generally have treated multiple measures of race as interchangeable and many consider the inconsistencies that result from comparing them to be an issue of "measurement error:" a problem that might affect the standard errors of one's statistical analysis, but would not call into question one's broader research conclusions. I argue instead that the inconsistencies between various measures of race contain useful information about the relative status of racial groups or identities, and that previous research conclusions based on single-measure methods likely misrepresent the state of racial stratification in the United States. By assuming (implicitly or otherwise) that each individual has a single "true" race, current methods obscure potential explanations for the perpetuation of racial inequality. For example, studies of group differences in status that rely only on self-reported racial identity will overstate inequality and immobility to the extent that individuals or groups

[6] Of course, the change also dovetailed conveniently with the Census Bureau's fiscally motivated switch from employing enumerators to send out census forms by mail.
[7] Multiple measures of race that include observed classifications (or a proxy for them) are also publicly available in the 1996 and 2000 General Social Survey, the 2004 and 2005 Behavioral Risk Factor Surveillance System and all three waves of the National Longitudinal Study of Adolescent Health (1996-2003).

change their reported identities as they improve their status (or change their reported identities in pursuit of higher status). Without measures of race from different, and potentially contradictory, perspectives (e.g., known ancestry, self-identity, observed classification, and which box one checks on a job application or other official form), survey researchers cannot hope to move from simply describing racial inequality in America to actually explaining how it persists.

What is *latent class analysis*?

The primary assumption in latent class analysis is that unobserved variation among individuals or the underlying structure of a population can be inferred from the pattern of individuals' responses to a given set of questions in a survey. The survey items are considered manifest, or observable, indicators of the latent, or unobservable, characteristics. So, for example, answering affirmatively to a question about "feeling blue" might be an indicator of the unobservable construct "depression." The goal of a latent class analysis is to end up with meaningful groupings of individuals for whom you have multiple pieces of information that you think are related. The model estimates these "classes" as clusters of observations with similar conditional probabilities of giving the same pattern of responses to the chosen "indicators" (in this case, observed and self-reported race). To name the latent classes, researchers often rely on the modal responses given by individuals assigned to that cluster (e.g., "observed black, self whites"). The characteristics of these classes, including their size, can then be compared to the racial categories typically used in quantitative research.[8]

Latent class analysis is part of the family of models known collectively as latent structure models. The different types of latent structure models are distinguished by the scales of their

[8] For a more formal statistical discussion of the latent class model and estimation techniques, see Goodman (1974) and McCutcheon (1987). For a recent overview of the use of latent variables in the social sciences, see Bollen (2002).

observed indicator variables and the hypothesized scale of the latent variable(s). For example, factor analysis is appropriate for use with continuous observed variables when one believes that the latent variable is also continuous, such as using numeric scores on a set of tests to approximate a range of “intelligence.” Latent class analysis is the categorical equivalent to factor analysis; it uses categorical observed variables and relates them to a latent variable that is also believed to be categorical, or to have several discrete “classes.” In surveys, race and ancestry variables are typically categorical (i.e., the response categories are “white,” “black,” “Asian,” etc.) and because my goal is to capture the intersection of observed and self-reported race, I assume the racial latent variable I want to estimate is also categorical.

The formal equation for the latent class models I estimate in this paper includes four observed variables (A, B, C and D) that represent the various observed and self-reported race responses and one latent variable X composed of T latent classes. The parameters being estimated include T (the number of latent classes), the size of each latent class in terms of the proportions of respondents that fit within it, as well as the conditional probabilities that a member of a specific latent class will have a given response on the observed indicators. Using the notation proposed by Goodman (1974), the model is written as follows:

$$\pi_{ijklt}^{ABCDX} = \pi_t^X \pi_{it}^{\bar{A}X} \pi_{jt}^{\bar{B}X} \pi_{kt}^{\bar{C}X} \pi_{kt}^{\bar{D}X}, \text{ for } i = 1 \ldots \text{I}; j = 1 \ldots \text{J}; k = 1 \ldots \text{K}; l = 1 \ldots \text{L}; t = 1, \ldots, \text{T}.$$

The left-hand side of the equation represents the joint probability that an observation is in response category *i* on variable A, in response category *j* on variable B, in response category *k* on variable C, in response category *l* on variable D and in class *t* on variable X. That joint probability equals the probability that an observation is in class *t* on variable X (this is the parameter indicating the size of the latent class) multiplied by the conditional probabilities for each of the observed variables.

I use the statistical package Mplus (Muthen and Muthen 2006) to estimate all of the models I present below. In addition to the latent class parameters described above, the program estimates posterior probabilities that are then used to assign individual cases to their most likely latent class. It is this process of modal assignment that allows the latent variable to be used in subsequent studies as either a dependent or independent variable.[9] As the purpose of this paper is primarily descriptive, I use these computed class memberships only to aid in characterizing the different latent classes.

The Data: National Survey of Family Growth

The National Survey of Family Growth is based on in-person interviews with women aged 15-44 and is typically used for studies of pregnancy, childbearing, contraception, and related aspects of maternal and infant health.[10] However, the survey also includes detailed background information about the respondent and her husband (if relevant), such as education, religion, ethnic origin, occupation, and earnings. Further, the first four cycles of the survey include the interviewer's classification of the respondent's race. In each cycle, the NSFG oversampled black women in order to allow for meaningful comparisons between blacks and whites, and the observed race data was used primarily to calculate the post-stratification weights. Interestingly, none of the published studies I found that use NSFG data from these years note

[9] It is also possible, using structural equation modeling, to estimate the latent classes and include, for example, covariates that help predict class membership all in one step.
[10] For additional details on the survey, see the National Center for Health Statistics website (http://www.cdc.gov/nchs/products/elec_prods/subject/nsfg.htm) or the NSFG webpage from the Office of Population Research at Princeton University (http://opr.princeton.edu/archive/nsfg/).

that multiple measures of race or ethnicity exist in the survey, nor do the authors state explicitly which measure they use in their analyses.

There are some important differences in the survey across cycles. The four cycles I use were fielded in 1973 (N=9,797), 1976 (N=8,611), 1982 (N=7,969) and 1988 (N=8,450). In 1973 and 1976, the NSFG sampled only women of reproductive age who had children of their own living in their household. In the 1982 and 1988 cycles, the survey is designed to be representative of all women aged 15-44. Before 1988, the surveys do not include women living in Alaska or Hawaii. The NSFG also is not immune to the changes in question wording, category options and/or coding protocol that plague nearly all racial data from repeated cross-sectional secondary sources. In 1973 and 1976, the survey asks the respondent only for her "origin or descent." Respondents in 1982 and 1988 are asked for their "racial background" and their "national origin or ancestry" in two separate questions. (For details on the response categories available in each of the four survey cycles, see Table 1.)

However, a feature that sets the NSFG data apart from similar secondary sources, particularly from this era, is that it allows multiple mentions for each of these questions. In 1973, respondents could name as many origin categories as they wished (and two women chose to name as many as nine). In the remaining survey cycles, respondents were allowed two mentions – in 1982 and 1988 this was true for both the race and ancestry questions. These multiple mentions of self-reported race and ancestry makes the NSFG a unique resource for studies of race and ethnicity in the United States; it allows us to see how American women identified themselves long before being able to report one's full multiracial ancestry became a political issue, and in an era when the idea of self-definition as self-determination for racial minority groups was just beginning to take hold. If I find meaningful complexities and discrepancies in these data, it represents an even stronger test of my argument that multiple measures of race need to be included in quantitative research than if I demonstrated it using contemporary data when the complicated racial ancestries and identities of political figures, such as Barack Obama, and sport stars, such as Tiger Woods, have become popular topics of speculation and conversation.[11]

As a first cut at estimating racial latent variables for these data, I use only the “racial background” responses for 1982 and 1988. I also aggregate the self-reported response categories into to “Black,” “European/White,” and “Other” in order to make the models more comparable across surveys.[12] Though some detail is certainly lost in the process, these three categories also match the coding of observed race, which is consistent in each survey cycle.[13]

Tables 2 through 5 depict the cross-classifications of observed and self-reported race on which my latent class models are based. Observed race is coded as a three-category nominal variable. I code the respondent’s answers into three binary variables for self-reported black, self-reported European/white and self-reported “other.” [14] Because of the possibility of multiple mentions, individuals can have “yeses” on each of the three self-report variables (thus, the row and column totals in Tables 2-5 do not sum to the sample size). Respondents who refused to answer the race or ancestry question, or replied “don’t know” are coded as “nos” on up to three of the self-report variables. The number of “don’t know” responses decreases dramatically between the 1970s and 1980s data, coincident with the shift from asking an open-ended question about the respondent’s “origin or descent” to asking a closed-ended question about her “racial

[11] The fact that the NSFG only samples women is a limitation, but I don’t expect it to affect my conclusions about whether latent-class analysis is a useful method for measuring race. In previous work comparing observed and self-reported race, I did not find statistically significant gender differences in the probability of having an inconsistent racial classification (Saperstein 2006).

[12] In future work, I plan to estimate these models taking into account all possible race and ancestry responses.

[13] The interviewer codes her observation of the respondent’s race before asking any of the background questions in the survey.

[14] In reality, then, the cross-classification table on which my models are based is a 2x2x2x3 table. I condense that to the 3x3 tables depicted in Tables 2-5 for the sake of simplicity.

background." In 1988, an observed classification was "not ascertained" for 48 of the respondents. These cases were dropped from the analysis.

Because they were oversampled, black women make up approximately one-third of the sample in each year (though blacks were just 10 percent of the total U.S. population during this time period). However, I do not use weights to reapportion the sample in either the cross-classification tables or in my analyses below. My goal in this paper is not to provide nationally representative statistics but to identify meaningful groupings of the population by observed and self-reported race. This renders the weights more of a hindrance—in terms of understating the power of the data to distinguish differences between blacks and whites—than a help. The weights only affect the estimated sizes of the latent classes in my models; they do not change the overall patterns of racial classification and identity that are observed or the number of latent classes that are identified.[15]

Model results

Before I present the results of my latent class models, I want to highlight an important tension between my aim of revealing previously hidden complexities and discrepancies in racial data using multiple measures of race, and the methods one uses to select the best-fitting latent class model. There is no one measure or fit statistic that provides a definitive answer as to which model is the most appropriate for a given set of empirical data (McCutcheon 1987). Here, I use two statistics that are commonly presented in latent class analyses, the log likelihood and the Bayesian Information Criterion (BIC), as well as a bootstrap likelihood ratio test (BLRT) that has been shown to accurately detect the correct t-class model from the t-1 and t+1 class models,

[15] If nationally representative descriptive statistics are desired regarding the distribution of the population among the classes of the latent variable, then the weights can be introduced after each individual is assigned to her most likely latent class.

even in very small samples, using simulated data (Nylund et al. 2006). Generally speaking, all the fit statistics reward parsimony, which is an important feature considering that technically the best-fitting latent class model is one that includes a separate latent class for each observation in the data. However, this means qualitatively interesting classes are likely to be rejected if they contain only a few individuals from the total sample. This distinction between qualitatively interesting and quantitatively useful is a key one. Latent class analysis is not a good method for fully exploring the tails of a distribution. Since my primary purpose is to explicate an approach to modeling race that can help advance everyday quantitative research, I limit myself to discussing models that identify a few meaningful groups that can be incorporated into subsequent studies. Whenever possible, though, I try to note some of the qualitatively interesting discrepancies that are glossed over in the process of identifying to "best-fitting" model for a given survey year.

"Anti-LCA" baseline models

My contention in this paper and the reason I am advocating the use of latent class analysis for studying race and racial inequality is that I believe observed racial classifications tell us something meaningful, and substantively different, than racial self-identifications. If this primary assumption is, in fact, false—if observed racial classifications and self-identifications do measure the same thing and both are simply prone to random measurement error—then latent class analysis is unnecessary. Empirical evidence in support of this counter argument would be that in each cycle of the NSFG the model that best fits the observed data is a three-class model, with one class comprised of consistent whites (i.e., women who both said they were white and were observed to be white), one comprised of consistent blacks, and the remaining class that includes everybody else. So, I first estimate this "anti-LCA" model as a baseline from which to gauge improvement in the various model results that follow.

Table 6 shows the model fit statistics for the "anti-LCA" models in each survey year, compared to the fits statistics for the unrestricted latent class models I discuss below. To support my argument that multiple measures of race are necessary to describe racial variation in the United States, I want to see log likelihoods that are larger (i.e., less negative or closer to zero) than those for the "anti-LCA" models. In contrast, I want to see BIC statistics that are smaller.[16] As noted above, the BLRT tests the null hypothesis that the t-1 class model fits better than the currently estimated t-class model. If the p-value falls below the conventional value of .05, then the t-1 class model is rejected. The BLRT is not particularly helpful for comparing across models as I do here, but seeing a p-value larger than .05 does give a good indication of when to stop adding classes (because the new class does not explain enough variation to justify the number of parameters required to estimate it).

The process of building my latent class models

For each survey year, I ran one further check on my results. I estimated a model with only one latent class, which tests the null hypothesis that the observed indicators are completely independent of one another (and therefore do not require a latent variable to explain their association). In each case, the null hypothesis was rejected (results not shown) and I moved on to estimate a series of two, three and four-class unrestricted latent class models. Given the 24 possible response combinations on my observed indicators, I could estimate up to 23 nonredundant parameters before using up the available degrees of freedom for a specific model.

[16] A general rule of thumb for adjudicating between BIC statistics for non-nested models is that a difference of 10 represents strong evidence of an improved fit for the model with the smaller BIC.

To estimate models with more than four classes, which requires exactly 23 unique parameters,[17] I had to introduce restrictions on either the conditional probabilities (e.g., requiring that all members of a given class self-reported as black) or the latent class frequencies (e.g., specifying that one of the classes included 12 percent of the population). I discuss the restricted models in detail below.

The estimated conditional probabilities and latent class frequencies for the unrestricted latent class models can be found in Appendix Tables 1-4. A conditional probability of one means that all members of the given class exhibited the specified response. A conditional probability of zero means that no member of the class exhibited that response. Though I use more general descriptions of the latent classes in my discussion, my comments are based the estimated probabilities in the appendix tables and the assignment of individuals to their most likely latent class.

Unrestricted two-, three- and four-class models

Interestingly, the two-class unrestricted model is nearly identical for each of the four survey years. Generally speaking, it divides NSFG respondents into consistent blacks and everyone else. However, in each year, the black-nonblack models provide a poorer fit to the observed data than the "anti-LCA" models (see Table 6).

The unrestricted three-class models explain more of the variation in the observed data than both the black-nonblack models and the "anti-LCA" models in each of the four survey years. These three-class models generally try to fit the diagonal of the observed cross-classification table—one class each for consistent blacks, consistent whites and consistent others,

[17] The formula for calculating the required number of parameters in a latent class model (following the notation in the formal equation given above) is: (T-1) + T [(I-1) + (J-1) + (K-1) + (L-1)]. So, for a four-class model with one three-category nominal indicator and three binary indicators, the number of parameters is (4-1) + 4(2+1+1+1) = 23.

with multiple mention respondents and those with discrepant classifications distributed across the three classes. This is a slight twist on the “anti-LCA” model that confines all the inconsistent cases to one class. However, these “diagonal” three-class models are similar enough that if the fit statistics suggest we should stop adding latent classes here, it also would support an argument that observed and self-reported measures of race provide essentially the same information about how the population is distributed. This only appears to be the case for the 1982 data, a point I discuss further below.

In 1973, 1976 and 1988, the unrestricted four-class model is a clear improvement on the “diagonal” three-class model (as well as the respective “anti-LCA” models).[18] In each case, the fourth class is made up of an easily interpretable group defined by a complex combination of self-identities and observed classifications, which supports my contention that using multiple measures of race is crucial to understanding race in America. In 1973, the additional class is made up entirely of observed blacks who reported themselves either as black-other or a combination of black, other and European origins (2 percent of the population). In 1976, the fourth class is made up of observed whites who reported themselves as “other” either alone or in combination with a European origin group (19 percent), while the observed black, self-black-multiple-mention respondents remained in the same class with the consistent blacks. In 1982 and 1988, the fourth class again identifies a small group of observed blacks who reported multiple origin groups, one of which was black. However, in 1982, as noted above, all three fit statistics suggest the four-class model is a poor fit to the observed data—likely because the population of

[18] In 1976, the BIC statistic for the unrestricted three-class model is lower than that for the four-class model. However, the BLRT clearly rejects the three-class model ($p<.0128$). I favor the four-class model here because Nylund, Asparouhov and Muthen (2006) found the BLRT to be more reliable than the BIC in these situations. Further, the fourth class is made up of a distinct and substantively interesting group, which also is an argument in favor of the more complex model.

observed black, self-reported black multi-mention respondents that year (N=44) is significantly smaller than any of the other three survey years.

Given strong initial support for my latent-variable approach from the unrestricted models, I went on to estimate models with additional classes with the aim of revealing more groupings of women who have complex or inconsistent racial identities and classifications. Doing so requires moving away from simply "letting the data speak," as I did in the models above, because the four-class unrestricted model exhausts the available degrees of freedom. Instead, I switched to testing the fits of a series of restricted models where the definitions of the latent classes are specified in advance.

Restricted models of five or more classes

Details of the restricted latent class models I estimated for each survey year can be found in Tables 7-10. They include my characterization of the class (e.g., "observed black, black-others"), the estimated class sizes and the associated fit statistics. I do not present tables of conditional probabilities for these models because the probabilities are what get "restricted" to create the classes. So, for example, to create a class of "observed white, self-reported others" the conditional probability of being observed white is fixed at 1, while the probabilities of being observed to be either black or other are fixed at zero. Similarly, the conditional probability of self-reporting as other would be set to 1, and the probability of self-reporting as black to zero. If the class also includes respondents who self-reported as other in combination with a European origin, then the parameter for self-reporting as European can either be freely estimated, or set to a middle range value. I used results from the unrestricted four-class models to guide my selection of parameter restrictions. For the reasons of practicality noted above, I did not estimate models with more than eight latent classes for any survey year.

Based on the results from the 1976 four-class model, I hypothesized that an appropriate fifth class for the 1973 data would be one comprised of observed-white, self-reported "others." Though the log-likelihood for this five-class model is smaller than that of the unrestricted four-class model (and bigger log likelihoods are better), both the BIC statistic and the BLRT indicate adding the fifth class of observed-white, self-reported "others" improves the fit (see Table 7).[19] Further splitting this relatively large class of inconsistent cases (15 percent) into separate classes for observed whites who reported only an "other" origin group and observed whites who report both European and "other" origins passes muster according to BLRT, but not BIC. A seven-class model that retains the same six classes but also separates observed-nonwhite self-reported others (i.e., they were observed to be either black or other) from the residual discrepancies results in the best fit of all the previous models. Further separating out the 289 observed white, self-reported "don't know" respondents into their own class is acceptable according to the BLRT but results in a 20-point increase in the BIC statistic.[20] If these "don't know" respondents look qualitatively different from everyone else on some outcome of interest, then an argument can be made for keeping them separate. Otherwise, the 1973 data are best described by the seven-class model that reveals three consistent classes and three meaningful groupings of racial identities and

[19] The log-likelihoods in all of my models are smaller than one would normally prefer, and the BIC statistics are larger (which is not surprising given that BIC is calculated using the log-likelihood), because the cross-classification tables underlying these models are quite sparse. However, the difference in the log-likelihoods between two related models are robust to the sparseness of the observed tables (Haberman 1974), which allows me to say that one model is a better fit than a competing model even when neither appear to be the best-fitting model in some absolute sense.

[20] It is important to note, here, that these seven and eight class models are not nested for the purposes of the BLRT. In each of these cases, the BLRT is testing whether dropping the unrestricted residual "everyone else" class and randomly distributing those cases across the remaining classes is a better fit to the observed data. It should not be interpreted to say that my eight-class restricted model is preferred to my seven-class restricted model. Instead, it suggests that the eight classes I defined explain enough of the variation to justify the additional parameter required to estimate the additional class. (There is only one additional parameter between the seven and eight class models because the other five parameters that would otherwise be required to estimate the additional class have been fixed in order to define the class as "observed white, self-reported 'don't knows.'")

classifications previously hidden by single-measure methods: observed white, self-reported others; observed whites who report mixed European and "other" origins; and observed blacks who report mixed black and "other" origins.[21]

Though the sizes of the various latent classes are different in the 1976 data, the substantive conclusions about how many meaningful racial groups exist in the population remains essentially the same (see Table 8). The major difference is related to the placement of the observed-black, self-reported blacks with mixed origins. In 1976, the distinctness of the observed-black, self-reported black multiple-mention women remains in question even in the seven-class model—unlike for 1973, when the unrestricted four-class model clearly defines them as a separate class. There is some support from the BLRT for making the observed-black, self-reported black multiple-mention women a separate seventh class in 1976, but the BIC statistic that corresponded to making the observed white, self-reported "don't knows" into the seventh class suggests that alternative model would be a better fit.[22] Also, the fact that the sizes of the two observed-white self-reported other classes are so much larger in 1973 is likely due to allowing more than two mentions on the origin question that year.

For 1982, though the three-class model was the best fit of the unrestricted models, I estimated several 4-class and higher restricted models to test whether the results were being skewed by the unusually small number of observed black self-reported black multiple mention women that year. As a baseline for my restricted models, I estimated a modified "anti-LCA" four-class model that fit the diagonal of the observed cross-classification table, but also allowed

[21] One of the qualitatively interesting discrepant groups that is lost in this seven-category classification scheme is the observed-black, self-reported Europeans (N=25), who end up lumped together with a lone observed other, self-reported European, the observed black, self-reported "don't knows" (N=20) and the other residual respondents with quirky classification-identity combinations.

[22] When I tried to estimate an eight-class model that included separate classes for both groups, the observed-black, black multiracial class was estimated to be empty.

for an “everyone else” class (instead of distributing the complicated and inconsistent cases across the other classes). This proved to be a better fit than the unrestricted four-class model for 1982 that tried to separate out the observed black, black multiple mention respondents (see Tables 6 and 9). I then added the fifth class of observed white, self-reported other multiple-mention women, which is so clearly distinguishable in the 1973 and 1976 data. That improves the fit relative to the restricted four-class, and is not rejected by the BLRT. Adding back in the class of observed black, self-reported black multiple mentions (restricted six-class model) and then splitting the observed white self-reported others, as I had for the previous survey years, results in statistically indistinguishable fits (because the difference between the BICs is less than 10). Though neither of these models represents an improvement over the unrestricted three-class model, the fact that the BLRT does not reject any of the five- and higher class models suggests they may be comparable. Given the substantive value of including the additional classes, and the similarity of those classes with the findings for 1973 and 1976, I would argue that the 1982 data is not as discrepant as it initially appeared and still provides some support for my general argument about the necessity of taking multiple measure of race into account using a latent-variable approach.

In contrast to these slightly discrepant findings for 1982, the results for 1988 regarding the appropriate number of latent classes are surprisingly congruent with those for 1973 and 1976. The most important findings for 1988, relative to the earlier years, relates to whether a new and distinct class will replace the observed white, self-reported “don’t knows” (who, as I noted earlier, no longer make up a notable proportion of the sample, likely due to the changes in question wording). This new group is revealed for the first time in the six-class model for 1988: observed white, self-reported blacks (see Table 10). The number of observed white, self-reported blacks went from zero in 1973 to 41 in 1988 (doubling every year from the initial 11 in 1976)—putting them nearly on par with the number of observed white, self-reported others (N=53), which actually declines even more dramatically over the same period (from 1,013 in 1973). I will resist speculating about the reason for the dramatic increase in this fascinating group because to do it justice likely would require a whole paper unto itself. Suffice it to say, for now, that the sudden arrival of observed white, self-reported blacks on the American racial landscape would have gone entirely undetected by conventional methods of measuring race.

Scratching the surface

The utility of any new method comes from its ability to tell us something about society that we did not already know. In earlier work, I demonstrated that observed and self-reported race cannot be considered comparable measures for studies of racial inequality (Saperstein 2006). Here, I detail a method for including both observed and self-reported racial classifications in quantitative research. My latent variable approach clearly identifies two previously hidden populations—observed-white, self-reported others (in each survey year) and observed-white, self-reported blacks (in 1988)—in addition to two noteworthy multiracial populations.

In gauging the potential impact of identifying the observed white, self-reported others across survey years, it is important to recall that for the purposes of this paper, my self-reported “other” category is a combination of Hispanic, American Indian and Asian origin responses. Given current theories of the “whitening” of these groups—through assimilation, intermarriage and improved socioeconomic status across generations (e.g., Zhou 2003)—being able to track the progress of observed-white, self-reported others and compare it to the progress of their observed-other, self-reported other counterparts could provide important insight into the persistence of inequality for some groups and the paths to mobility for others in the United States. Similarly, documenting the appearance in the late 1980s of a small but significant number of observed white, self-reported blacks raises a whole host of questions about what might have changed in 15 short years that either allowed previously self-reported blacks to suddenly be seen as white, or previously self-reported whites to suddenly decide to report themselves as black. Answering these questions would be a fruitful direction for future research. Finally, by highlighting several significant multiracial populations in the 1970s and 80s, that come with their attendant observed racial classifications, my latent class analyses of race also provide important benchmarks with which to compare more recent data that includes multiple measures of race, such as that found in the National Longitudinal Study of Adolescent Health.

Demonstrating the use of racial latent variable classifications, such as those detailed here, in subsequent studies of the racial inequalities in income, political participation and health is the primary focus of my future work. However, I also plan to estimate latent class models of race for the General Social Survey, the National Longitudinal Study of Adolescent Health and the Center for Disease Control's Behavioral Risk Factor Surveillance System. By doing so, I hope to further our understanding of how latent classes of race vary over time, across subgroups and when using different measures of race.

References

American Anthropological Association. 1997. “Response to OMB Directive 15: Race and Ethnic Standards for Federal Statistics and Administrative Reporting.”

American Sociological Association. 2003. *The Importance of Collecting Data and Doing Social Scientific Research on Race.* Washington, DC: American Sociological Association.

Bollen, Ken. 2002. “Latent Variables in Psychology and the Social Sciences.” *Annual Review of Psychology* 53: 605-34.

Brown, J. Scott, Steven Hitlin and Glen Elder Jr. 2006. “The Greater Complexity of Lived Race: An Extension of Harris and Sim.” *Social Science Quarterly* 87: 411-31.

Ferguson, Ronald. 1998. “Teachers’ Perceptions and Expectations and The Black-White Test Score Gap,” Pp. 273-317 in *The Black-White Test Score Gap*, Jencks, Christopher and Meredith Phillips, eds. Washington D.C.: Brookings Institution Press.

Gans, Herbert. 1999. “The Possibility of a New Racial Hierarchy in the 21st Century United States.” Pp. 371-90 In *The Cultural Territories of Race: Black and White Boundaries,* ed. Michele Lamont. Chicago: University of Chicago Press.

Goldstein, Joshua R. and Ann J. Morning. 2000. "The Multiple-Race Population of the United States: Issues and Estimates." *Proceedings of the National Academy of Sciences of the United States of America* 97:6230-6235.

Goodman, Leo A. 1974. “The Analysis of Systems of Qualitative Variables When Some of the Variables Are Unobservable. Part I: A Modified Latent Structure Approach.” *American Journal of Sociology* 79: 1179-1259.

Gossett, Thomas F. 1997 [1963]. *Race: The History of an Idea in America. New York: Oxford University Press.*

Haberman, Shelby J. 1974. *The Analysis of Frequency Tables*. Chicago: University of Chicago Press.

Hahn, R.A., J. Mulinare, and S.M. Teutsch. 1992. “Inconsistencies in Coding of Race and Ethnicity between Birth and Death in US Infants: A New Look at Infant Mortality, 1983 through 1985.” *Journal of the American Medical Association* 267:259-63.

Harris, David R. and Jeremiah Joseph Sim. 2002. "Who is Multiracial?: Assessing the Complexity of Lived Race." *American Sociological Review* 67:614-27.

Johnson, Robert Alan. 1990. “Measurement of Hispanic Ethnicity in the U.S. Census: An Evaluation Based on Latent-Class Analysis.” *Journal of the American Statistical Association* 85, 409: 58-65.

Kaplan, Judith B. and Trude Bennett. 2003. “Use of Race and Ethnicity in Biomedical Publication.” *Journal of the American Medical Association* 289: 2709-2716.

Kirschenman, Jolene and Katherine Neckerman. 1991. “We’d Love to Hire Them, But …” In *The Urban Underclass,* eds. Christopher Jencks and Paul E. Peterson. Washington D.C.: The Brookings Institution.

Lieberson, Stanley and Mary C. Waters. 1993. "The Ethnic Responses of Whites - What Causes Their Instability, Simplification, and Inconsistency." *Social Forces* 72:421-450.

Macmillan, Ross and Carolyn Liebler. 2005. “The Structure of Race: A Latent Class Approach to Modeling Racial Formation in Contemporary America.” Paper presented at the Population Association of America Annual Meeting, Philadelphia.

Martin, John Levi and King-To Yeung. 2003. “The Use of the Conceptual Category of Race in American Sociology, 1937-99.” *Sociological Forum* 18:521-543.

McCutcheon, Allan L. 1987. *Latent class analysis*. Beverly Hills: Sage Publications.

Muthen, Linda and Bengt O. Muthen. 2006. *Mplus User’s Guide.* Available online, www.statmodel.com.

Nagel, Joanne. 1994. “Constructing Ethnicity: Creating and Recreating Ethnic Identity and Culture.” *Social Problems* 41: 152-176.

Nylund, K.L., Asparouhov, T., & Muthen, B. (2006). “Deciding on the number of classes in latent class analysis and growth mixture modeling. A Monte Carlo simulation study.” Forthcoming in *Structural Equation Modeling*.

Omi, Michael and Howard Winant. 1994. *Racial Formation in the United States: From the 1960s to the 1990s*. New York: Routledge.

Paret, Marcel and Aliya Saperstein. 2006. “Inheriting Race: The Classification of U.S. Newborns in the Early Childhood Longitudinal Study-Birth Cohort.” Presented at the Population Association of America Annual Meeting, Los Angeles, Calif., March 30.

Raftery, Adrian. 1995. “Bayesian Model Selection in Social Research.” *Sociological Methodology* 25:111-64.

Roth, Wendy. 2005. “The End of the One-Drop Rule? Labeling of Multiracial Children in Black Intermarriages.” *Sociological Forum* 20: 35-67.

Saperstein, Aliya. 2006. “Double-checking the Race Box: Examining Inconsistency between Survey Measures of Observed and Self-Reported Race.” *Social Forces* 85(1): 57-74.

______________ 2007. “(Re)Modeling Race: How Using Latent Variables Can Advance Research on Race and Racial Inequality.” Unpublished manuscript.

Stephan, Cookie W. and Walter G. Stephan. 2000. “The Measurement of Racial and Ethnic Identity.” *International Journal of Intercultural Relations* 24: 541-52.

Sugarman, Jonathan R., Robert Soderberg, Jane E. Gordon, and Frederick P. Rivara. 1993. "Racial misclassification of American Indians: Its effect on injury rates in Oregon, 1989 through 1990." *American Journal of Public Health* 83(5):681-684.

Tafoya, Sonya, Hans Johnson and Laura Hill. 2005. “Who Chooses to Choose Two?**”** Pp. 332-52 in *The American People: Census 2000*, eds. Reynolds Farley and John Haaga. New York: Russell Sage Foundation.

Telles, Edward E. 2002. “Racial Ambiguity among the Brazilian Population.” *Ethnic and Racial Studies* 25: 415-441.

______ and Nelson Lim. 1998. “Does it Matter Who Answers the Race Question? Racial Classification and Income Inequality in Brazil.” *Demography* 35: 465-74.

Xie, Yu and Kimberly Goyette. 1997. “The Racial identification of Biracial Children with One Asian Parent: Evidence from the 1990 Census.” *Social Forces* 76: 547-70.

Yee, A.H., H.H. Fairchild, F.Weizmann and G.E. Wyatt. 1993. “Addressing Psychology’s Problems with Race.” *American Psychologist* 48: 1132-40.

Zhou, Min. 2004. “Are Asian Americans Becoming ‘White’?” *Contexts 3, 1: 27-35.*

Table 1. Race and Ancestry Response Categories in the First Four Cycles of the NSFG

	1973	1976	1982	1988
Observed Race	Black White Other	Black White Other	Black White Other	Black White Other
Racial Background *(two mentions)*	Not asked	Not asked	American Indian Asian or Pacific Islander Black White Another Group Refused/Don't Know	Same as 1982
Origin or Ancestry *(at least two mentions)*	German Italian Irish French Polish Russian English, Scot, Welsh Mexican Puerto Rican Cuban Other Spanish Black American Indian Asian American Other Don't Know	Puerto Rican Cuban Mexican Other Spanish Black American Indian Asian European Other	32 categories … including all from 1973 Refused Not Ascertained	44 categories Refused Not Ascertained

Question wording:

	1973	1976	1982	1988
Racial background	Not asked	Not asked	"Which of the groups on this card best describe your racial background?"	Same as 1982
Origin or Ancestry	"What is your origin or descent?"	Same as 1973	"Which of the groups on this card best describe your national origin or ancestry?"	Same as 1982

Table 2: Cross-tabulation of Observed Race and Self-reported Origin in the 1973 NSFG

		Observed race			
		Black	**White**	**Other**	**Row total**
Self-reported origin or descent	**Black**	3696	0	1	3697
	European	91	4562	3	4656
	Other	287	2452	76	2815
	Column total	4074	7014	80	N= 9,797

Note: Unweighted counts. Respondents were allowed multiple mentions of origin or descent, so the row and column totals do not sum to the sample size. "European" is my aggregation of specific national origin responses (e.g., Irish). "Other" is my aggregation of Hispanic (e.g., Cuban), American Indian, Asian, "American" and other origin responses. Respondents who answered "don't know" do not appear in this table: 289 were observed to be white, 20 were observed to be black.

Table 3: Cross-tabulation of Observed Race and Self-reported Origin in the 1976 NSFG

		Observed race			
		Black	**White**	**Other**	**Row total**
Self-reported origin or descent	**Black**	2925	12	3	2940
	European	65	4562	7	4634
	Other	194	1271	109	1574
	Column total	3184	5845	119	N= 8,611

Note: Unweighted counts. Respondents were allowed two mentions of origin or descent, so the row and column totals do not sum to the sample size. "Other" is my aggregation of Hispanic (e.g., Cuban), American Indian, Asian, "American" and other origin responses. Respondents who answered "don't know" (N=301) or have completely missing data on origin (N=11) do not appear in this table: 306 were observed to be white, six were observed to be black.

Table 4: Cross-tabulation of Observed and Self-reported Race in the 1982 NSFG

Self-reported race		*Observed race*			
		Black	**White**	**Other**	**Row total**
	Black	3180	24	7	3211
	White	33	4531	30	4594
	Other	59	115	90	264
	Column total	3272	4670	127	N= 7,969

Note: Unweighted counts. Respondents were allowed two mentions of racial background, so the row and column totals do not sum to the sample size. "Other" is my aggregation of American Indian, Asian and other race responses. Hispanic origin was coded in a separate variable and is not included in the analyses.

Table 5: Cross-tabulation of Observed and Self-reported Race in the 1988 NSFG

Self-reported race		*Observed race*			
		Black	**White**	**Other**	**Row total**
	Black	2711	42	8	2761
	White	44	5227	34	5305
	Other	82	166	166	414
	Column total	2837	5435	208	N= 8,402

Note: Unweighted counts. Respondents were allowed two mentions of racial background, so the row totals do not sum to the sample size. "Other" is my aggregation of American Indian, Asian and other race responses. Hispanic origin was coded in a separate variable and is not included in the current analyses. Forty-eight respondents had missing data for observed race and were dropped from the analysis. The five respondents who refused to give a racial background or answered "don't know" and the 122 who have missing data on racial background are not included in this table: 14 were observed to be black, 53 were observed to be white and 60 were observed to be "other."

Table 6. Comparison of fit statistics for "Anti-LCA" models and unrestricted latent class models estimated on data from the first four cycles of the NSFG

"Anti-LCA" baseline models

	1973	1976	1982	1988
LL	-16098	-13038	-7577	-9045
BIC	32260	26139	15218	18152
BLRT	$p < 0.0000$	$p < 0.0000$	$p < 0.0000$	$p < 0.0000$

Unrestricted latent-class models

		1973	1976	1982	1988
Two-class model	LL	-16211	-13285	-8248	-9977
	BIC	32524	26676	16596	20054
	BLRT	$p < 0.0000$	$p < 0.0000$	$p < 0.0000$	$p < 0.0000$
Three-class model	LL	-15451	-12418	**-7355**	-8808
	BIC	31058	**24990**	**14863**	17770
	BLRT	$p < 0.0000$	$p < 0.0000$	**$p < 0.0000$**	$p < 0.0000$
Four-class model	LL	**-15391**	-12415	-7354	**-8775**
	BIC	**30993**	25038	14915	**17758**
	BLRT	**$p < 0.0000$**	**$p < 0.0128$**	$p < 0.2273$	**$p < 0.0000$**

Note: Larger log likelihoods, smaller BICs and BLRTs of $p < 0.05$ indicate models with better fits to the observed data. Fit statistics in bold indicate the "best-fitting" model for each year, according to the given statistic.

Table 7. Restricted latent class models estimated using the 1973 NSFG

	Characterization of latent class	Est. Frequency
Class 1	Observed white, self Europeans	32%
Class 2	Consistent blacks	36%
Class 3	Observed black, self black multis	2%
Class 4	Observed white, self other-Euro.	15%
Class 5	unrestricted residual	16%

Model fit statistics:

LL	-15438
BIC	30960
BLRT	$p < 0.0000$

	Characterization of latent class	Est. Frequency
Class 1	Observed white, self Europeans	32%
Class 2	Consistent blacks	36%
Class 3	Observed black, self black multis	2%
Class 4	Observed white, self other-Euro.	15%
Class 5	Observed white, self other	10%
Class 6	residual (incl. nonwhite self-others)	5%

Model fit statistics:

LL	-15527
BIC	31137
BLRT	$p < 0.0000$

	Characterization of latent class	Est. Frequency
Class 1	Observed white, self Europeans	32%
Class 2	Consistent blacks	36%
Class 3	Observed black, self black multis	2%
Class 4	Observed white, self other-Euro.	15%
Class 5	Observed white, self other	10%
Class 6	Observed nonwhite, self others	2%
Class 7	unrestricted residual	4%

Model fit statistics:

LL	-15392
BIC	**30875**
BLRT	$p < 0.0000$

	Characterization of latent class	Est. Frequency
Class 1	Observed white, self Europeans	32%
Class 2	Consistent blacks	36%
Class 3	Observed black, self black multis	2%
Class 4	Observed white, self other-Euro.	15%
Class 5	Observed white, self other	10%
Class 6	Observed nonwhite, self others	2%
Class 7	Observed white, "don't knows"	3%
Class 8	unrestricted residual	1%

Model fit statistics:

LL	-15391
BIC	30892
BLRT	**$p < 0.0000$**

Fit statistics indicated in bold represent the "best-fitting" model according to the given statistic

Table 8. Restricted latent class models estimated using the 1976 NSFG

	Characterization of latent class	Est. Frequency
Class 1	Observed white, self Europeans	45%
Class 2	Consistent blacks	32%
Class 3	Observed white, self other-Euro.	8%
Class 4	Observed white, self other	7%
Class 5	residual (incl. consistent others)	8%

Model fit statistics:

LL	-12874
BIC	25829
BLRT	$p < 0.0000$

	Characterization of latent class	Est. Frequency
Class 1	Observed white, self Europeans	45%
Class 2	Consistent blacks	32%
Class 3	Observed white, self other-Euro.	8%
Class 4	Observed white, self other	7%
Class 5	Consistent others	1%
Class 6	unrestricted residual	7%

Model fit statistics:

LL	-12789
BIC	25678
BLRT	$p < 0.0000$

	Characterization of latent class	Est. Frequency
Class 1	Observed white, self Europeans	45%
Class 2	Consistent blacks	32%
Class 3	Observed white, self other-Euro.	8%
Class 4	Observed white, self other	7%
Class 5	Consistent others	1%
Class 6	Observed black, self black multis	2%
Class 7	unrestricted residual	5%

Model fit statistics:

LL	-12539
BIC	25177
BLRT	$p < 0.0000$

	Characterization of latent class	Est. Frequency
Class 1	Observed white, self Europeans	45%
Class 2	Consistent blacks	32%
Class 3	Observed white, self other-Euro.	8%
Class 4	Observed white, self other	7%
Class 5	Consistent others	1%
Class 6	Observed white, "don't knows"	3%
Class 7	unrestricted residual	4%

Model fit statistics:

LL	**-12492**
BIC	**25093**
BLRT	**$p < 0.0000$**

Fit statistics indicated in bold represent the "best-fitting" model according to the given statistic

Table 9. Restricted latent class models estimated using the 1982 NSFG

	Characterization of latent class	Est. Frequency
Class 1	Consistent whites	56%
Class 2	Consistent blacks	39%
Class 3	Consistent others	1%
Class 4	unrestricted residual	3%

Model fit statistics:

LL	-7540
BIC	15151
BLRT	$p < 0.0000$

Fits better than the unrestricted four-class, not the unrestricted three-class

	Characterization of latent class	Est. Frequency
Class 1	Consistent whites	56%
Class 2	Consistent blacks	39%
Class 3	Consistent others	1%
Class 4	Observed white, self other multis	1.4%
Class 5	unrestricted residual	2%

Model fit statistics:

LL	-7487
BIC	15063
BLRT	$p < 0.0000$

	Characterization of latent class	Est. Frequency
Class 1	Consistent whites	56%
Class 2	Consistent blacks	39%
Class 3	Consistent others	1%
Class 4	Observed white, self others multis	1.4%
Class 5	Observed black, self black-others	0.5%
Class 6	unrestricted residual	1.4%

Model fit statistics:

LL	-7454
BIC	**14998** *
BLRT	$p < 0.0000$

	Characterization of latent class	Est. Frequency
Class 1	Consistent whites	56%
Class 2	Consistent blacks	39%
Class 3	Consistent others	1%
Class 4	Observed white, self others	0.8%
Class 5	Observed white, self white-others	0.7%
Class 6	Observed black, self black-others	0.5%
Class 7	unrestricted residual	1.4%

Model fit statistics:

LL	-7452
BIC	15003
BLRT	**$p < 0.0000$** *

**Neither of these models clearly fits better than the unrestricted three-class*

Table 10. Restricted latent class models estimated using the 1988 NSFG

	Characterization of latent class	Est. Frequency
Class 1	Consistent whites	61%
Class 2	Consistent blacks	31%
Class 3	Observed black, self black multi.	1%
Class 4	Consistent others	2%
Class 5	unrestricted residual	5%

Model fit statistics:

LL	-8880
BIC	17859
BLRT	$p < 0.0000$

	Characterization of latent class	Est. Frequency
Class 1	Consistent whites	61%
Class 2	Consistent blacks	31%
Class 3	Observed black, self black multi.	1%
Class 4	Consistent others	2%
Class 5	Observed white, self other-whites	1%
Class 6	unrestricted residual	4%

Model fit statistics:

LL	-8833
BIC	17784
BLRT	$p < 0.0000$

	Characterization of latent class	Est. Frequency
Class 1	Consistent whites	61%
Class 2	Consistent blacks	31%
Class 3	Observed black, self black multi.	1%
Class 4	Consistent others	2%
Class 5	Observed white, self other-whites	1%
Class 6	Observed white, self blacks	0.5%
Class 7	unrestricted residual	3%

Model fit statistics:

LL	-8800
BIC	17682
BLRT	$p < 0.0000$

	Characterization of latent class	Est. Frequency
Class 1	Consistent whites	61%
Class 2	Consistent blacks	31%
Class 3	Observed black, self black multi.	1%
Class 4	Consistent others	2%
Class 5	Observed white, self other-whites	1%
Class 6	Observed white, self blacks	0.5%
Class 7	Observed white, self others	0.7%
Class 8	unrestricted residual	2.4%

Model fit statistics:

LL	-8790
BIC	**17671**
BLRT	**$p < 0.0000$**

Fit statistics indicated in bold represent the "best-fitting" model according to the given statistic

Appendix Table 1. Conditional probabilities and latent class frequencies for unrestricted latent class models of race data from 1973 NSFG

	Class 1	Class 2
Obs. White	.97	.00
Obs. Other	.33	.00
Obs. Black	ref. cat.*	1.00
Self Euro.	.76	.02
Self Black	.00	1.00
Self Other	.43	.05
Est. Freq.	62%	38%

	Class 1	Class 2	Class 3
Obs. White	.99	.00	.90
Obs. Other	.04	.00	.42
Obs. Black		1.00	
Self Euro.	.92	.02	.02
Self Black	.00	1.00	.00
Self Other	.31	.05	1.00
Est. Freq.	50%	37%	12%

	Class 1	Class 2	Class 3	Class 4
Obs. White	.89	.00	.90	.00
Obs. Other	.04	.00	.42	.00
Obs. Black		1.00		1.00
Self Euro.	.92	.01	.02	.01
Self Black	.00	1.00	.00	1.00
Self Other	.31	.00	1.00	.00
Est. Freq.	50%	36%	12%	2%

Model fit statistics:

	2-class		3-class		4-class
LL	-16211	LL	-15451	LL	**-15391**
BIC	32524	BIC	31058	BIC	**30993**
BLRT	$p < 0.0000$	BLRT	$p < 0.0000$	BLRT	**$p < 0.0000$**

Appendix Table 2. Conditional probabilities and latent class frequencies for unrestricted latent class models of race data from 1976 NSFG

	Class 1	Class 2
Obs. White	.99	.00
Obs. Other	.57	.00
Obs. Black		1.00
Self Euro.	.81	.01
Self Black	.00	1.00
Self Other	.25	.05
Est. Freq.	66%	34%

	Class 1	Class 2	Class 3
Obs. White	.99	.00	.93
Obs. Other	.08	.00	.72
Obs. Black		1.00	
Self Euro.	.93	.01	.04
Self Black	.00	1.00	.00
Self Other	.01	.05	1.00
Est. Freq.	57%	34%	9%

	Class 1	Class 2	Class 3	Class 4
Obs. White	.99	.00	.99	.00
Obs. Other	.12	.00	.00	.77
Obs. Black		1.00		
Self Euro.	1.00	.01	.52	.04
Self Black	.00	1.00	.00	.01
Self Other	.00	.05	.67	1.00
Est. Freq.	46%	36%	19%	2%

Model fit statistics:

	2-class		3-class		4-class
LL	-13285	LL	-12418	LL	-12415
BIC	26676	BIC	**24990**	BIC	25038
BLRT	$p < 0.0000$	BLRT	$p < 0.0000$	BLRT	**$p < 0.0128$**

**The cond. probabilities for observed race are estimated in a multinomial logisitic regression equation; thus they are relative to the reference category*

Appendix Table 3. Conditional probabilities and latent class frequencies for unrestricted latent class models of race data from 1982 NSFG

	Class 1	Class 2
Obs. White	.99	.01
Obs. Other	.79	.00
Obs. Black		
Self White	.97	.00
Self Black	.00	.99
Self Other	.04	.02
Est. Freq.	59%	41%

	Class 1	Class 2	Class 3
Obs. White	.99	.01	.76
Obs. Other	.50	.00	.81
Obs. Black			
Self White	1.00	.00	.02
Self Black	.00	1.00	.01
Self Other	.01	.01	1.00
Est. Freq.	58%	40%	2%

	Class 1	Class 2	Class 3	Class 4
Obs. White	.99	.01	.76	.00
Obs. Other	.50	.00	.81	.05
Obs. Black				
Self White	1.00	.00	.02	.00
Self Black	.00	1.00	.00	1.00
Self Other	.01	.00	1.00	.29
Est. Freq.	58%	39%	2%	1%

Model fit statistics:

	2-class		3-class		4-class
LL	-8248	LL	-7355	LL	-7354
BIC	16596	BIC	**14863**	BIC	14915
BLRT	$p < 0.0000$	BLRT	**$p < 0.0000$**	BLRT	$p < 0.2273$

Appendix Table 4. Conditional probabilities and latent class frequencies for unrestricted latent class models of race data from 1988 NSFG

	Class 1	Class 2
Obs. White	1.00	.02
Obs. Other	.92	.00
Obs. Black		
Self White	.94	.01
Self Black	.00	.99
Self Other	.06	.03
Est. Freq.	67%	33%

	Class 1	Class 2	Class 3
Obs. White	1.00	.02	.79
Obs. Other	.59	.00	.92
Obs. Black			
Self Black	.99	.01	.02
Self White	.00	1.00	.01
Self Other	.02	.02	.73
Est. Freq.	63%	33%	4%

	Class 1	Class 2	Class 3	Class 4
Obs. White	1.00	.02	.81	.00
Obs. Other	.60	.00	.92	.04
Obs. Black				
Self Black	.99	.00	.02	.18
Self White	.00	1.00	.00	.97
Self Other	.02	.00	.73	.50
Est. Freq.	63%	32%	4%	1%

Model fit statistics:

	2-class		3-class		4-class
LL	-9977	LL	-8808	LL	**-8775**
BIC	20054	BIC	17770	BIC	**17758**
BLRT	$p < 0.0000$	BLRT	$p < 0.0000$	BLRT	**$p < 0.0000$**

**The cond. probabilities for observed race are estimated in a multinomial logisitic regression equation; thus they are relative to the reference category*