STILL SEARCHING FOR A TRUE RACE? REPLY TO KRAMER ET AL. AND ALBA ET AL.[1]

We wrote "Racial Fluidity and Inequality in the United States" (Saperstein and Penner 2012) with the aim of jump-starting a conversation about how race is best conceptualized in studies of stratification. Does assuming that people have static, often mutually exclusive, "races" help us understand disparities in the contemporary United States? Or are there inequality-sustaining mechanisms we might have missed by assuming a process of consensus through which racial categorizations are ascribed at birth and effectively fixed? That two teams of scholars took time to engage with our work is a positive sign that a much-needed conversation is happening. It is, of course, disheartening that Kramer et al. (2016) (hereafter KDH) and Alba et al. (2016) (hereafter AIL) come to the conclusion that we, at worst, misinterpreted our data and, at best, overstated our case. Nevertheless, we are grateful to have this opportunity to clarify our claims and intentions and to offer new evidence that returning to assumptions of rigid racial ascription is not the way forward.

Our reply addresses the three main points of empirical critique across the two comments: (1) that a relationship between social status and racial categorization of similar direction and magnitude could be produced by measurement error; (2) that our findings are neither as common nor as generalizable as we claimed; and (3) that we did not adequately demonstrate that stereotypes, operating through what the interviewer does or does not hear about the respondent, are a key causal mechanism. We either provide evidence that directly refutes each point, explain why it results from a misreading of our argument, or both. In the process, we underscore our earlier

[1] The authors would like to acknowledge support and hospitality from the Russell Sage Foundation and the Institute for Research in the Social Sciences and valuable research assistance from Galo Falchettore, Steve McClaskie, and Robert Pickett. The authors also are grateful to Marianne Bitler, Elizabeth Cohen, Shelley Correll, Greg Duncan, Paula England, Claude Fischer, David Grusky, Aaron Gullickson, Mike Hout, Judd Kessler, Sasha Killewald, Mona Lynch, Vida Maralani, Suzanne Nichols, Mary Pattillo, Trond Petersen, Sean Reardon, Cecilia Ridgeway, Matt Snipp, and Chris Wildeman for helpful comments and discussions. Direct correspondence to Aliya Saperstein, Department of Sociology, Stanford University, Stanford, California 94305. E-mail: asaper@stanford.edu

Correction: This reply was reposted on July 19, 2016. In footnote 8 (p. 266), the page numbers for the quotation from Kramer, DeFina, and Hannon's commentary were mistakenly not supplied. The cite should read (pp. 237–38). An erratum will be published in the September 2016 issue.

0002-9602/2016/12201-0009$10.00

findings with additional evidence of how social factors shape categorization: through selective processes of "ethnic attrition" as well as what interviewers knew about the respondents' use of crack cocaine.

First, a clarification on the scope of this debate: the details of our empirical analysis are not the only aspect of our article being subjected to scrutiny; also in question for both commenters is whether or not race and ethnicity should be conceptualized as stable individual characteristics. Given space constraints, we focus on the specific empirical critiques, but it is important not to lose sight of their theoretical foundations and implications. Although AIL and KDH give us credit for advancing a "sophisticated" and "provocative" thesis, we are not the first to conclude that recording these characteristics, and assigning individuals to categories, is a more complex task than racial commonsense and received practice would indicate. Thus, we encourage readers to weigh not only other empirical research that demonstrates the relationship between social status and racial categorization can run in both directions (Saperstein and Gullickson 2013; Saperstein, Penner and Kizer 2014; Young, Sanchez, and Wilton 2015; Simonovitz and Kezdi 2016) but also the broader body of theoretical work on the social construction of race and ethnicity.

CAN THE RESULTS BE EXPLAINED BY MEASUREMENT ERROR?

Both AIL and KDH raise concerns about how much of the racial fluidity we reported is "real" and how much is due to measurement error. AIL highlight this issue by dividing their descriptive statistics to show which respondents experienced one or two changes in racial classification compared to three or more. KDH claim that our original estimates for the effect of social status on racial categorization could result from introducing random variation alone.

To counter these critiques, we begin with our preferred approach to potential coding mistakes in the data: we show that when we remove cases with classification discrepancies that might be considered "errors," we get results similar to those we originally reported. We then consider the results presented by KDH (in their table 1) and demonstrate that there is nothing in their approach that suggests our core finding can be attributed solely to measurement error. However, we also note that most discussions of measurement error assume the existence of an objectively verifiable value, as is the case for individual attributes like height or weight. To make an analogous assumption for race is problematic.

Blips versus Shifts in Classification

As we noted in "Racial Fluidity and Inequality" (p. 689), roughly one-third of the respondents in the 1979 National Longitudinal Survey of Youth

(NLSY) who experience a change in their racial classification have only one discrepant classification over the course of the survey.[2] These seemingly idiosyncratic "blips" can be contrasted with racial classification trajectories that represent shifts from one classification to another (e.g., WWWWW-OOOOOOOOOOOO) or classification trajectories that repeatedly oscillate, back and forth, between two or more categories (e.g., BBBWBBBB-WWBWBWBWB).[3] To the degree that blips in classification represent actual errors (e.g., introduced during transcription), and not unconscious racial biases on the part of the interviewers, we might wish to ensure that our findings are not driven by such cases.[4]

For this reanalysis, we operationalize "blips" by considering strings of five adjacent classifications and identifying cases where the respondent has the same race in both the previous two classifications and the subsequent two classifications, but a different race in between. Thus, in the string WWOWW, the third classification would be considered a blip, but strings like WWOOO, WWOWO, or BBOBW would not be considered blips.[5] According to this definition, blips account for about 22% of the cases where current and prior survey year racial classifications differ.

Having identified these blips, we then reestimate the results using only the nonblip racial classifications. Given that we can only identify blips when we have two previous and two subsequent racial classifications, we also

[2] In their discussion, AIL group together individuals with one or two changes in classification (see AIL, table 1). This is likely because the classification history WWOWW has one discrepant classification, but two changes (W to O and O to W). We focus on discrepancies (defined as the number of classifications that deviate from the modal classification) because WWWWO and WWWOO both have one change but differ in the number of discrepant classifications.

[3] A systematic examination of shifts and oscillations is beyond the scope of this reply. However, we do find evidence for the presence of shifts in racial classification, insofar as respondents' previous classification predicts their current classification net of their average racial classification, suggesting that discrepant classifications are not distributed randomly over time.

[4] We do not believe simple keystroke errors are driving our results because such errors would produce a pattern that is not reflected in the data. Before 1994, data coders entered racial classifications by typing 1 for "white," 2 for "black," and 3 for "other." From 1994 on, interviewers recorded their classifications directly into a laptop computer by toggling down through the list from white to black to other and then pressing "enter" to register their selection. Thus one would expect keystroke errors to result in changes being more common between adjacent categories, but that is not the case.

[5] In order to maximize the number of cases in these analyses, we use the sequence of all available classifications to determine if a classification is a blip, regardless of whether there were surveys that the respondent missed in between racial classifications. Although we use the nearest nonmissing classifications to identify blips, we only estimate our models using cases where racial classifications are present in the current and previous survey years.

reestimate our original models (from table 4, p. 699) using only the relevant set of survey years.[6] The results from models including and excluding blips are strikingly similar (see table 1), suggesting that our findings are not driven by one-off classification changes or the types of measurement error that might produce them.

Measurement Error in Independent and Dependent Variables

In evaluating KDH's critique that the relationship between social status and racial classification presented in "Racial Fluidity and Inequality" is simply a product of measurement error, it is crucial to remember that measurement error in dependent and independent variables can have very different effects.[7] Much of the conventional wisdom about the effects of measurement error comes from studies of continuous independent and dependent variables in bivariate ordinary least squares regression models (OLS). In the bivariate OLS case with continuous variables and independent errors, it is well established that measurement error in an independent variable introduces bias, while measurement error in a dependent variable results instead in a loss of efficiency, but less is known about how one should expect measurement error to work in other contexts (see Bound et al. 2001 for a review).

With this understanding of measurement error in mind, we engage KDH's critique and approach to testing our results on its own terms. The logic behind their approach is that each individual has a true race and that fluidity is observed because measures of race vary randomly from this true value. According to this logic, every measure of race should include some measurement error, but KDH's procedure only makes changes to the race measure being used as the independent variable.[8] To better capture the logic

[6] Unless otherwise noted, we retain variable coding and other model specifications from Saperstein and Penner (2012). However, all estimates are presented as log odds, following AIL and KDH.

[7] Classical measurement error also suggests that errors should not vary in prevalence across different populations (e.g., Hispanic and non-Hispanic) and that such errors should be uncorrelated with other variables in the model (both dependent and independent).

[8] KDH explain that they "set each individual's racial identity in 1979 equal to the corresponding 2002 self-identity . . . then randomly change the 1979 racial self-identities of 1.6% of the sample, the same percentage of individuals who actually changed in the NLSY . . . then replicate this procedure 100 times" (pp. 237–38). Expressed simply, KDH estimate a version of the following model:

$$Race_{i\,2002} = Race^{\dagger}_{i\,2002} * \beta + Status_{ki} * \gamma_k + Controls_{ki} * \delta_k + \varepsilon_i, \tag{1}$$

where $Race_{i\ 2002}$ is the respondent's race in 2002, $Race^{\dagger}_{i\,2002}$ is the respondent's race in 2002 with 1.6% of cases switched from black to nonblack (or vice versa), and $Status_k$ and $Controls_k$ are the respective status variables and controls included in the model. We

TABLE 1
PREDICTING RACIAL CLASSIFICATION WITH AND WITHOUT CLASSIFICATION "BLIPS"

	BLACK		WHITE	
	Includes Blips	Excludes Blips	Includes Blips	Excludes Blips
Unemployed......	.34***	.41***	−.29***	−.30***
Impoverished.....	.35***	.25*	−.36***	−.39***
Incarcerated......	.29+	.27	−.25**	−.31**
Received welfare...	.15*	.20	−.13***	−.15**

NOTE.—Data are from the 1979 NLSY. Each column represents a different model with specifications following from Saperstein and Penner (2012, table 4), except that these models are limited to the years in which it is possible to identify blips in both current and lagged classification (1982–94).

+ $P < .10$.

* $P < .05$.

** $P < .01$.

*** $P < .001$.

that all measures of race are measured with error, we introduce changes to both the independent and dependent variables.[9] We estimate 1,000 replications of both KDH's original procedure and our modified version of their procedure and report the average coefficients from these models in table 2.[10] Table 2 presents estimates from baseline models, models using KDH's approach ("IV only"), and models with our modification ("IV&DV"). Finally, to provide a sense of where the baseline model results fall in the distribution of the IV&DV results, we report the proportion of the IV&DV coefficients with a magnitude larger than the baseline coefficient. We take a similar ap-

requested KDH's data for our reply, but KDH did the randomization in 100 separate Excel spreadsheets and were unable to share these files; thus we cannot exactly reproduce their results. Our understanding is that they sorted the data according to a random number, and changed the race measure for the independent variable for the first 1.6% of respondents. So, to replicate KDH in our eq. (1) we also sort the data randomly and change $Race^{\dagger}_{i\,2002}$ for the first 1.6%. These results are reported in table 2 under "IV only."

[9] We also estimate the model:

$$Race^{\dagger\dagger}_{i\,2002} = Race^{\dagger}_{i\,2002} * \beta + Status_{ki} * \gamma_k + Controls_{ki} * \delta_k + \varepsilon_i, \qquad (2)$$

where $Race^{\dagger\dagger}_{i\,2002}$ and $Race^{\dagger}_{i\,2002}$ are the respondent's race, each perturbed equally. For eq. (2), we change $Race^{\dagger}_{i\,2002}$ for the first 0.8% and change $Race^{\dagger\dagger}_{i\,2002}$ for the second 0.8%, resulting in the same number of changes, and the same overall number of discrepancies as found between the original 1979 and 2002 self-identifications (see table 2, IV&DV, for results). We also estimated a version of eq. (2) that allows the measurement error in the dependent and independent variables to be independent, which produced similar results (not shown).

[10] Unlike KDH, we do not report the average of our standard errors from these models, because the average standard error does not reflect the variation of the average of the coefficients (cf. Rubin 1987).

TABLE 2
Reexamining KDH's Approach to Measurement Error

	Black				White			
	Baseline	IV Only	IV & DV	P(B<I)	Baseline	IV Only	IV & DV	P(B>I)
Self-identification model:								
Unemployed	.76	.85	.37	.02	−.41	−.59	−.28	.04
Impoverished	.83	.86	.42	.03	−.59	−.75	−.35	.00
Incarcerated	.27	.42	.22	.44	−.55	−.61	−.30	.05
Received welfare	.25	.53	.24	.48	−.07	−.41	−.20	.94
Interviewer classification model:								
Unemployed	.32		.33	.54	−.29		−.26	.02
Impoverished	.35		.36	.61	−.37		−.32	.00
Incarcerated	.32		.30	.43	−.33		−.28	.07
Received welfare	.18		.11	.14	−.12		−.11	.26

Note.—Models include other status variables and controls from Saperstein and Penner (2012, table 4). Cols. P(B<I) and P(B>I) report the proportion of the 1,000 replicates of the IV & DV models with coefficients of greater magnitude than the baseline results (i.e., how often does eq. [2] produce coefficients as large as the baseline model). In the interviewer classification models, we do not replicate the "IV only" (KDH) approach because the racial classification from a given survey year can be both an independent variable and a dependent variable.

proach to extend the analysis to self-identification as white, and interviewer classification as both black and white.[11]

Our small change to their procedure has large consequences. Like KDH, when we introduce changes only to the independent race variable, we find results that are similar in magnitude or larger than the baseline coefficients. However, when we perturb both the independent and dependent race variables, this is no longer the case. The pattern of results in the baseline models is infrequently produced by the IV&DV model; for example, the IV&DV coefficient for unemployment is as large as the baseline coefficient just 4% of the time when predicting self-identification as white, and 2% of the time when predicting self-identification as black.[12] Put simply, we show that their results can only be obtained if one makes the unfounded assumption that measurement error only affects race when it is an independent variable.

[11] When predicting self-identification as white, we replicate the procedure described above and match the level of fluidity observed in our white-nonwhite comparisons. To apply the same general approach to the 17 years of classification data, we choose a classification at random from among the racial classifications observed for a given individual, set this classification as their "true race," and then introduce random changes to recreate the observed levels of fluidity. In the classification models, we do not replicate the IV only approach because the classification from a given survey year can be both an independent and dependent variable.

[12] When we estimate models with status factors individually (following AIL's approach), we find that the IV&DV approach produces coefficients as large as the baseline models less than 5% of the time for 10 of the 16 contrasts.

Thus, although we cannot condone KDH's approach to evaluating measurement error, we conclude that there is nothing in KDH's critique to suggest measurement error is the only reason the relationship between social status and racial categorization appears to run in both directions.

WHO IS REALLY "AT RISK" OF RACIAL FLUIDITY?

The critique that we overstated the case for racial fluidity raises questions about (1) how common such fluidity is likely to be and (2) how general is the relationship between status and racial categorization. We address each of these issues in turn. Both AIL and KDH assert that meaningful changes are less common than we suggested and, when they do occur, such changes are limited to people who are racially "ambiguous." We do not disagree that categorical fluidity is observed more frequently in particular subpopulations, and we stated as much in "Racial Fluidity and Inequality" (p. 707). However, it is problematic to assume that either racial ambiguity or categorization as "Hispanic" (or "Latino") is a static characteristic. Further, it is important to clarify it is not the *level* of categorical racial fluidity but rather the *process* through which social status influences racial perceptions that we propose to be more general.

Significance of Fluidity not Primarily about Frequency

In "Racial Fluidity and Inequality" we highlighted that more than 20% of the 12,686 NLSY respondents experienced at least one change in their racial classification. AIL get a lower estimate because they apply survey weights. Although we were careful to discuss the level of fluidity relative only to the sample, we acknowledge that we did not explicitly state that we were not using the NLSY to establish the level of racial fluidity as a population parameter. From our perspective, a "true" estimate of racial fluidity is about as meaningful a concept as a "true" measure of race. Even gauging whether there is a lot of racial fluidity, or only a little, depends on how stable one expects race to be.

To us, the significance of racial fluidity stems not from its common-ness but from its utility for understanding processes of racial categorization. Fluid cases provide leverage that studying racially stable people cannot: they allow us to ask, what predicts being assigned to, or being removed from, a particular racial category? For this purpose, what matters most is whether a given sample provides enough variation to study the correlates of category assignment. That said, we recognize that establishing the magnitude and scope of racial fluidity is likely to interest many researchers and address these issues below.

To weight or not to weight.—As our aim was to exploit the repeated racial classifications to better understand predictors of categorization, the issue of weighting is most relevant for our multivariate analyses. We estimate coefficients for our key status variables using both the 1979 sample weight used by AIL and a customized panel weight. If anything, the evidence for our claims is stronger when we use the survey weights (see table 3).[13]

However, in considering whether or not our estimates should have been weighted, it is important to recognize that the survey's weighting schemes treat racial categories as fixed strata. Respondents are weighted differently depending on how they were classified by NLSY in 1978—which is not necessarily consistent with how respondents later identified themselves, were perceived by interviewers, or would have been recorded in the 1970 census (to which the weighted population distributions were pegged). So, if one qualified as "Hispanic" based on the survey screener, one would be assigned separately calculated "Hispanic" weights throughout.[14] Yet, respondents also could self-report Hispanic origins in 1979, or answer "yes" that they were "Hispanic, Latino, or of Spanish origin" in 2002. The complexity of who counts more or less when the weights are employed is highlighted by the fact that just 467 (32%) of the 1,437 respondents who have "Hispanic" weights are consistently "Hispanic" across all three measures. Given this, and the fact that weighting did not affect our multivariate analyses, we present unweighted frequencies and model estimates throughout our reply.

Fluidity and ambiguity.—AIL and KDH emphasize that the vast majority of Hispanics (as defined by the 1979 measure) have fluid racial classifications, and that these "ambiguous" cases account for a sizeable proportion of the overall fluidity in the sample. They also imply that much of the racial fluidity would be eliminated with a more appropriate set of ethnoracial categories. We disagree that a better measure of race would eliminate either fluidity or ambiguity; some cases will fit better in a given classification scheme than others, and this is true for all types of classification (Zerubavel 1991). Further, rather than thinking of racial ambiguity as a relatively fixed

[13] See https://www.nlsinfo.org/weights/nlsy79 for information on customized weights. The similarity between our results with and without the survey weights is not surprising, given that our models controlled for many respondent characteristics used to create the weights.

[14] To be classified as "Hispanic" based on the 1978 screener, respondents had to self-identify as "Mexican American," "Cuban," "Puerto Rican," or "Other Spanish," or live in a household where one of their parents identified as such. Respondents also were assigned a Hispanic categorization if their family surname was listed on the U.S. Census Bureau list of Spanish surnames and the householder or householder's spouse reported speaking Spanish at home as a child (see NLSY memo dated 10/4/1978). This contrasts with the 1979 self-identification measure, for which respondents selected "origin or descent" categories that included "Cuban," "Chicano," "Mexican," "Mexican-American," "Puerto Rican," "other Hispanic," and "other Spanish."

TABLE 3
PREDICTING RACIAL CATEGORIZATION WITH AND WITHOUT WEIGHTS

	BLACK			WHITE		
	Unweighted	1979 Weights	Panel Weights	Unweighted	1979 Weights	Panel Weights
Self-identification model:						
Unemployed	.76***	.79***	.79***	−.41***	−.42***	−.42***
Impoverished	.83***	.93***	.93***	−.59***	−.65***	−.65***
Incarcerated	.27	.37	.33	−.55*	−.69**	−.69**
Received welfare	.25	.35	.35	−.07	$-.23^{+}$	$-.23^{+}$
Interviewer classification model:						
Unemployed	.32***	.31***	.31***	−.29***	−.27***	−.27***
Impoverished	.35***	.51***	.51***	−.37***	−.49***	−.49***
Incarcerated	.32*	.39*	.39*	−.33***	−.42***	−.42***
Received welfare	.18**	.28***	.28***	−.12***	−.21***	−.21***

NOTE.—Data are from the 1979 NLSY. Model specifications follow Saperstein and Penner (2012, table 4). Models predicting self-identification and classification use different panel weights, corresponding to the different survey years on which they draw.
$^{+}P < .10$.
* $P < .05$.
** $P < .01$.
*** $P < .001$.

characteristic of the person being categorized, as AIL and KDH seem to, we think of racial ambiguity as the result of a confluence of factors from the individual's own characteristics to the circumstances in which the categorization takes place. Neither people nor populations are always racially ambiguous; ambiguity (or lack thereof) is socially constructed and entwined with the classification scheme, such that different people will be "ambiguous" in different schemes.

We illustrate that fluidity and the expectation of ambiguity do not always go hand-in-hand by comparing levels of racial fluidity in the NLSY across a range of individual characteristics likely to be associated with categorical ambiguity (or difficulty fitting particular individuals into U.S. racial classification schemes). This point can be seen most clearly in table 4 among NLSY respondents who selected "origin or descent" categories in 1979 that might indicate multiracial heritage (e.g., black and Asian Indian); if anything, as a group, they are less likely to have fluid racial classifications over the course of the survey than those who report a single origin (or whose multiple origins do not cross racial boundaries).[15] Racial fluidity is more

[15] Unlike AIL, we do not assume reporting either a Hispanic origin or "mainly" European origins trumps reporting multiple origins that cross conventionally defined racial lines. For example, we consider people who identified as French, Indian-American or Native American, and Irish as "multiracial," and people who identified as black, German, and Puerto Rican as both "Hispanic" and as "multiracial."

TABLE 4
Individual Ambiguity and Racial Fluidity

	% With at Least One Discrepant Racial Classification	*N*
Reported multiracial origins in 1979	19	1,380
Did not report multiracial origins in 1979. . .	21	11,116
Reported a Hispanic origin in 1979	87	1,976
Did not report a Hispanic origin in 1979 . . .	8	10,520
Not born in the United States	70	868
Born in the United States	17	11,736
Self-id changed between 1979 and 2002	45	1,544
Self-id did not change	18	6,090

Note.—Data are from the 1979 NLSY. All indicators are coded following Saperstein and Penner (2012).

common not only among people who report Hispanic origin in 1979 but also among people who were not born in the United States,[16] and people whom we classified as changing their self-identification between 1979 and 2002. Yet some fluidity is present even among respondents who are non-Hispanic, U.S. born, and stably self-identified. Thus, racial fluidity and racial ambiguity should be treated as distinct analytical concepts, regardless of how related they might seem.[17]

To address AIL and KDH's claims that the residual "other" category drives the observed fluidity, we compare the NLSY to Add Health. We find similar levels of change in racial classification between waves 3 and 4 in Add Health as we do year-to-year changes in NLSY (4.5% vs. 6.0%, respectively), despite Add Health's more specific classification scheme, which included "American Indian or Alaska Native" and "Asian or Pacific Islander" but did not include "other." Although neither survey provided "Hispanic or Latino" as a category, unlike AIL and KDH, we are skeptical that adding another option for the interviewers' classifications would yield racial stability. A "Hispanic" category might make some people's classifications more stable, but could make other people's classifications less stable than they were when there were fewer options available; the boundaries between white and Hispanic, black and Hispanic, or Asian and Hispanic are far from clear.

We agree with AIL and KDH that examining how the level of fluidity differs across subpopulations is necessary to fully understand processes of racial

[16] In the NLSY, 60% of foreign-born respondents reported Hispanic origin in 1979, but 74% of people who reported Hispanic origin were born in the United States.

[17] We prefer to think of observing racial fluidity as a sufficient condition for the presence of racial ambiguity, and racial ambiguity as a necessary (but not sufficient) condition for observing racial fluidity.

categorization, and the relationship between racial fluidity and ambiguity is important to consider. Nevertheless we chose to focus our efforts instead on the predictors of categorization because the question of whether or not status characteristics change along with changing racial categorizations is what connects the social construction of race to broader issues of inequality.

The Role of Status in Racial and Ethnic Categorization

It might seem counterintuitive that observed categorical racial fluidity could be concentrated in particular subpopulations, while the relationship between social status and racial categorization could be a more general one. This apparent paradox is partially resolved by recognizing that the status factors associated with a change in categorization from nonwhite to white in one year (e.g., graduating from college) can also help to explain why someone might continue being categorized as white in the future. Further, even if status shapes racial perceptions for everyone, some people's observed racial categorizations can still be more sensitive to changes in social position.

To demonstrate this, we briefly discuss the merits of thinking of racial categorization in terms of continuous probabilities. We then respond directly to AIL's critique by presenting fixed effects models estimating the association between social status and racial categorization across a series of subpopulations. We conclude this section by examining whether status factors also shape who identifies as Hispanic.

Observed categorical change vs. continuous probabilities of categorization.—Instead of each person in the United States being assigned a race in a static, categorical sense, we find it helpful to think of everyone as having a probability of identifying or being classified in each category that represents the chance that they will identify or be seen as a particular category at a particular point in time (see "Racial Fluidity and Inequality," pp. 706–8). We argue that status considerations—and changes in social position, in particular—have the potential to nudge these probabilities by a few percentage points, in one direction or the other. Whether a change in a person's probability of categorization results in observed fluidity depends on a host of other factors including the categories offered, who is doing the categorizing, and where the person being categorized was located in the various probability distributions to begin with. It is in this sense of continuous probabilities of categorization, and how those probabilities are shaped by status cues, that we intended our results to be seen as applicable to Americans in general.

The best evidence of this comes from our collaborative work on the psychological process of racial categorization (Freeman et al. 2011). We replicated our basic finding from the NLSY in an experimental setting by showing that the same faces are racially categorized differently depending on whether they are portrayed in high-status (business suit) or low-status (janitorial coveralls) clothing. A novel mousetracking tool allowed us to examine whether or not the status cues had an effect on racial perception when the categorization itself appeared unaffected. In the low-status condition, even when individuals ultimately classified faces as white, their mouse moved significantly closer to the box for making a "black" classification than it did when the same face was presented in the high-status condition. Conversely, in the high-status condition, when a face was classified as "black," average mouse trajectories veered significantly closer to the "white" category. These results suggest that even when status cues do not change how people are racially classified in categorical terms, they can play an important role in shaping racial perceptions more broadly.

General vs. subpopulation specific results.—AIL, in particular, question the claim that the bidirectional relationship between social status and racial categorization is a more general phenomenon. AIL interpreted their evidence of a relationship between social status and classification as "weak" based on models restricted to people who self-identified in 1979 using either Hispanic origin categories or origin categories that crossed racial lines (see AIL, table 2).[18] We revisit their conclusion across a broader range of subpopulations, including only the NLSY cross-sectional sample and only non-Hispanic or non-multiracial respondents. AIL note issues of statistical power in their models, and we too consider most of our models to be underpowered. Nevertheless, the exercise is useful to demonstrate that, overall, estimates for the associations of racial categorization with status factors are consistent in direction and magnitude regardless of subpopulation (see table 5).

Like AIL, we estimate a series of models with a common set of controls (for considerations such as age) in which our key status factors are entered individually.[19] Unlike AIL, our models include respondent fixed effects.[20] We also show results for the full sample with and without respondent fixed effects so the direction, magnitude, and statistical significance of the

[18] AIL come to this conclusion, in part, because they do not treat either educational attainment or marital status as racialized indicators of social position, as we did.

[19] AIL misunderstood how the models in "Racial Fluidity and Inequality" were estimated. Our dependent variable was not the presence or absence of a change in classification. Like AIL, our dependent variable was coded to predict a particular racial category (e.g., white vs. all else).

[20] Random effects models like those used by AIL can be useful, in that they allow researchers to examine covariates that do not vary within individuals. However, in this case, we prefer fixed effect models as they provide a more conservative test of the effects of time-varying status factors.

TABLE 5

Comparison of Model Estimates Predicting Racial Categorization With Different Sample Restrictions

	Whole Sample (no FE)	Fixed Effect Models: Whole Sample	Cross-sectional Sample	Non-Hispanic (1979 and 2002)	Hispanic (1979 or 2002)	Non-multiracial (1979 and 2002)	Multiracial (1979 or 2002)	Non-AIAN, Hispanic or Multiracial	AIAN, Hispanic or Multiracial
Self-identified as white:									
Unemployed	−.057***	−.012	−.008	−.016**	−.013	−.015*	−.001	−.007	−.028
Impoverished	−.069***	−.015+	−.014+	−.005	−.029	−.013+	.005	−.003	−.048*
Incarcerated	−.087***	−.047**	−.058*	−.034**	−.101	−.038*	−.182*	−.007	−.152**
Received welfare	−.043***	.004	−.009	−.013*	−.006	.012	−.054+	−.008+	−.019
Self-identified as black:									
Unemployed	.019***	.008**	.006*	.008**	.010	.009**	−.004	.008*	.009
Impoverished	.021***	.003	.003	.003	.002	.001	.015	.002	.006
Incarcerated	.020*	.008	.016	.009	.002	.009	−.003	.007	.011
Received welfare	.016***	.005+	.008*	.006+	.001	.004	.017	.007+	.003
Classified as white:									
Unemployed	−.027***	−.002	.002	.000	−.006	−.002	−.001	.000	−.003
Impoverished	−.032***	.003	.002	−.001	.015	.003	.002	.000	.014
Incarcerated	−.043***	−.008	−.013	−.007+	−.022	−.005	−.032*	−.005	−.018
Received welfare	−.029***	−.003	.002	−.001	−.015	−.001	−.016*	−.001	−.008
Classified as black:									
Unemployed	.005***	.002*	.001	.001	.005+	.001	.006	.001	.005*
Impoverished	.006***	.000	.001	.000	−.003	.000	.001	.000	.000
Incarcerated	.008***	.003	.001	.005+	−.004	.003	.007	.004	.003
Received welfare	.006***	.001	.000	.001	.002	.001	.001	.001	.002

Note.—Data are from the 1979 NLSY. Each coefficient represents a different model predicting racial categorization using the relevant status variable and controls for age, living in the South, year fixed effects, and interviewer characteristics (age, gender, education, and race). The first model also includes a control for prior racial classification/identification (as relevant), and additional controls for whether the respondent reported a Hispanic origin in 1979, multiple races in 1979, or was born outside the United States. Following Saperstein and Penner (2012), SEs in the non-FE models account for the clustering of observations within interviewers; fixed effect models account for the clustering of observations within respondents. AIAN—American Indian or Alaska Native.

+ $P < .10$.

* $P < .05$.

** $P < .01$.

*** $P < .001$.

estimates can be compared across the full range of model specifications.[21] We do not expect all of the coefficients in table 5 to be statistically significant, particularly given our original fixed effect model results (see "Racial Fluidity and Inequality," app. table A3, where status factors were entered simultaneously rather than separately). When we split the sample into smaller subpopulations, we have even less power to distinguish statistically significant differences than in our original analyses. However, there is still information to be gained by comparing estimates from these fixed effects models relative to estimates from models with less conservative controls but more statistical power.

The strongest evidence supporting our perspective in table 5 can be found in the relationship between unemployment and racial categorization—for both self-identification and interviewer classification. Across all four panels of fixed effects models, 28 of 32 estimates for the association between long-term unemployment and racial categorization are in the expected direction (decreasing the odds of white categorization and increasing the odds of black categorization), and 10 of those estimates are at least marginally statistically significant. Other associations show more mixed patterns, as in the case of poverty, which significantly lowers the odds of self-identifying as white in four of the eight comparisons but, if anything, seems to increase the odds of being classified as white by interviewers when respondent fixed effects are included.[22] Although individual status factors might not exhibit the expected associations across each type of racial categorization and for every subpopulation in table 5, we believe the results support our core claim that social status can influence racial categorization. Determining which specific factors matter, under what circumstances, for which populations, remains an important task for future research.

Evidence for differential ethnic attrition.—However, the decision to limit one's study to particular ethnoracial subpopulations, such as "Hispanics," is not as straightforward as AIL's analysis implies. Researchers have demonstrated that people with parents or grandparents born in Spanish-speaking countries who are doing well socioeconomically are less likely to identify as Hispanic in the census or other surveys (see, e.g., Duncan and Trejo 2011). Alba himself highlighted the case of "disappearing Mex-

[21] The baseline results include a control for race in the prior survey year. Subsequent models introduce dummy variables (fixed effects) for each respondent. If individuals had a "true" race that was associated with their likelihood of experiencing changes in social status, our fixed effects models would account for this because they control for all time invariant characteristics. They also address concerns about biased correlations with status that can affect between-person comparisons when there is random noise in the race measures.

[22] None of the estimates that run counter to our hypotheses are statistically significant.

icans," concluding that "selective departures from the Mexican-American group are on a scale that could easily impact on the measurement of group characteristics over time" (Alba and Islam 2009, p. 120). The challenge for understanding immigrant incorporation—and inequality—is that if people who are doing well are less likely to explicitly identify as Hispanic then their success in the United States is effectively removed from the population average for "Hispanics" and counted instead in the population average for "non-Hispanics." Thus, the category "Hispanic" should not be treated as exempt from the processes that affect racial categorization more broadly.

We find evidence that this process occurs in the NLSY. When we estimate models predicting who identifies as Hispanic in 2002, among respondents who reported a Hispanic origin in 1979, we find that people are significantly more likely to identify as Hispanic if they reported experiencing long-term unemployment or receiving welfare (see table 6). These results hold regardless of whether we limit our analyses to respondents who reported Hispanic origin in 1979 or to respondents who reported their mother, father, or paternal grandfather was born in a Spanish-speaking country. Associations with having ever been impoverished or incarcerated also are in the expected direction and of similar magnitude but are not statistically sig-

TABLE 6
Selective Ethnic Attrition in Hispanic Self-Identification (2002)

	Restricted to Respondents Reporting 1979 Hispanic Origin	Restricted to Respondents Reporting Hispanic Family Birthplaces
Unemployed	.31*	.36*
	(.14)	(.18)
Impoverished	.26	.42^{+}
	(.17)	(.24)
Incarcerated	.16	.33
	(.25)	(.33)
Received welfare	.38**	.57**
	(.13)	(.18)

Note.—Data are from the 1979 NLSY (includes restricted use data). $N = 1{,}398$ and 782 persons, respectively. Each coefficient represents a different model predicting Hispanic self-id in 2002 using the relevant status variable and controls for whether the respondent reported multiple races in 1979, was born outside the United States, age, living in the South, year fixed effects, and interviewer characteristics (age, gender, education, and race). Following Saperstein and Penner (2012), SEs in these models account for the clustering of observations within interviewers; models that account for clustering within respondents yield similar results.

$^{+}P < .10$.
* $P < .05$.
** $P < .01$.
*** $P < .001$.

nificant. In addition to bolstering our claim that "Hispanic" should not be treated as a static characteristic, these results further demonstrate the role that social status can play in influencing ethnoracial categorization.

WHAT IS THE CAUSAL MECHANISM?

We turn finally to concerns about our explanation for the relationship between racial fluidity and racial inequality. KDH assert that we did not adequately support our contention that racial stereotypes were shaping racial categorization. This is because our goal was not to isolate racial stereotypes as a specific cause of fluidity. Although it is difficult to think of a context in which racial stereotypes are not somewhat implicated, we view our research as merely consistent with this explanation, not as establishing the mechanism. In that sense, we agree with KDH that future research should examine specifically how social status, and other factors, come to affect people's racial classifications. Possible mechanisms include observable characteristics or behaviors of the targets of classification, attitudes or beliefs held by the person doing the classifying, and the context of the interaction itself.[23] To facilitate this effort, we assess the suggestions made by KDH and offer some additional insight.

We do want to be clear, though, about the limits of what can be established using observational data.[24] We interpret our results as suggesting that not only does one's racial categorization shape one's social status, one's social status can also shape perceptions of one's race. Readers looking for causal evidence that status shapes perceptions of race should consider our experimental manipulation of status cues (Freeman et al. 2011). However, even that study—which shows results are consistent with simulations from a neural model of stereotype-related processes—does not conclusively prove that stereotypes are the causal mechanism that link perceptions of status and perceptions of race.

[23] AIL also suggest that characteristics of the classifiers play an important role. Unlike AIL, we believe interviewers' racial conceptualizations, their exposure to racial diversity, and their feelings about their own place in the racial/status hierarchy, rather than their demographic characteristics per se, are likely to explain differences in classification.

[24] This is why, for example, we avoided using the word "cause" to describe our models with multiple lagged values. As KDH note, these are often referred to as "Granger cause" models. We interpreted those models as telling us whether a given status factor "provides useful information in predicting a person's current racial classification above and beyond the information provided by knowing how they were racially classified in previous years" (p. 705). KDH assert that our interpretation is not consistent with how such models are described in econometrics textbooks, but they omitted a sentence in the middle of their quote from Stock and Watson that reads, "Granger causality means that if X Granger-causes Y, then X is a useful predictor of Y, given the other variables in the regression."

Hearing Status, Seeing Race

KDH assert that they conducted better tests of our causal mechanism by exploiting variation in the interviewer's level of knowledge about the respondent's answers. We disagree, in part because explicit survey responses were only one of several factors we noted that might affect racial categorization (see, e.g., pp. 698, 688 n. 9). Other information—above and beyond physical appearance—might come from visual or contextual cues, such as the respondents' dress or demeanor, the way their house was furnished, who else was present during the interview, or the characteristics of the respondents' neighborhood. We have speculated elsewhere that changes in social status experienced in the past might continue to exert influence on subsequent racial classification as a result of these factors—as well as accumulation of advantage or disadvantage—regardless of whether the interviewer explicitly heard the information (see, e.g., Saperstein, Penner, and Kizer 2014).

Although many of these mechanisms are difficult to study using survey data, the NLSY does provide leverage on whether or not the interviewers' hearing specific information might shape their racial classifications of respondents.[25] It is in this spirit that KDH suggested contrasting the racial classification from the 1978 screener with later racial classifications. Interviewers were instructed to record race by observation in the screener before they asked a battery of questions about family income (NORC 1978, p. 56)—rather than making their racial classification after the entire interview was completed, as in subsequent years. However, we cannot recommend using the screener classification for this purpose because the informant for the household screener was unlikely to be the person who became the respondent in 1979. Screener information was obtained from the householder, householder's spouse, or other knowledgeable adult resident in 1978. Given that the cohort was restricted to people ages 14–21 in 1979, there is little reason to believe the person whose race the interviewer observed for the household screener was the same person whose racial classification was recorded in 1979.[26]

Racial stereotypes and reported drug use.—KDH's empirical strategy for the 1980 drug use questions also hinges on the interviewers' hearing about respondents' drug use (or lack thereof) in one year (1980), but not the next (1981). However, respondents recorded their answers to the 1980 drug use items confidentially (CHRR 2001, p. 51), using a pencil and paper form that was sealed in an envelope before it was returned to the interviewer.

[25] KDH's table 4 finding was addressed in our original n. 16.

[26] Further, for the screener, interviewers were instructed to assign all family members the same race as the householder or, in cases of apparent mixed ancestry, the race of the father (see NORC 1978, p. 105).

Therefore, the only way interviewers would know about respondents' drug use in 1980 was if respondents asked for assistance when filling out the form, or if they reported having been arrested for possession or use of various drugs as part of a separate interviewer-administered section on contact with the criminal justice system.

There are additional drug use supplements in the 1988, 1992, 1994, and 1998 surveys that include similar questions, and also have variation on whether or not the interviewer heard the respondent's answers. In some cases, interviewers administered the drug use supplement out loud and recorded the respondents' answers, while in other cases the respondents completed the questions confidentially. Thus, we were able to examine whether the association between respondents' recorded level of drug use and how they were racially classified varies depending on whether or not interviewers heard about the respondents' drug use.

We present estimates from logistic regressions predicting racial classification in the current year with information about the respondents' drug use, whether or not they filled out the drug use questions confidentially, and the interactions between the two (see table 7). Following KDH's example we use reported marijuana use to predict racial classification as white. We also examine how cocaine use is related to racial classification, but we do this separately for cocaine and "crack" cocaine, with the latter posited to increase the odds that the respondent will be seen as black. We consider the results in table 7 somewhat preliminary and interpret them cautiously. In placebo regressions predicting (1) interviewer classification in the subsequent year and (2) self-identification in 2002, the interaction effects were not significant (as expected), but we did not have time to fully explore variation by survey year or exploit the 1988 random assignment of respondents to confidential or nonconfidential response modes.[27] Nevertheless, we provide these analyses as an example highlighting how future research can explore such issues.

We find some evidence that when respondents report marijuana use the interviewer is more likely to later classify them as white. We are less troubled by the direction of this relationship than KDH seem to be, as it was probably more common, particularly during the 1970s and '80s, for Americans to think of people who smoke marijuana as disaffected white youths (see, e.g., Peterson 1985). However, based on the nonsignificant interaction effects, we conclude there is little evidence that hearing about marijuana use, per se,

[27] Of course, mode of data collection also affects who reports drug use (Hoyt and Chaloupka 1994). So, even with the experiment, differences in the relationship between status and racial classification cannot be definitively attributed to whether or not interviewers heard the information (i.e., we cannot rule out that any mode effects were driven by differential reporting).

TABLE 7
PREDICTING RACIAL CLASSIFICATION USING VARIATION IN INTERVIEWERS' KNOWLEDGE ABOUT RESPONDENTS' DRUG USE

	Marijuana Predicting Classification as White	Cocaine Predicting Classification as White	Crack Predicting Classification as Black
Reported drug use:			
1–2 times	.02	.04	1.94**
	(.12)	(.14)	(.69)
3–10 times	.23*	.07	.86***
	(.12)	(.15)	(.24)
>10 times	.17*	.05	.14
	(.09)	(.12)	(.50)
Confidential form	.01	.08	.23
	(.10)	(.07)	(.20)
Drug use X confidential:			
1–2 times	.14	.14	−2.00**
	(.15)	(.19)	(.71)
3–10 times	.10	.08	−1.26*
	(.15)	(.21)	(.53)
>10 times	.17	.15	.80
	(.11)	(.15)	(.63)
N (person years)	33,303	33,272	16,335

NOTE.—Data are from the 1979 NLSY. Models include controls for prior racial classification, whether the respondent reported a Hispanic origin in 1979, reported multiple races in 1979, was born outside the United States, age, living in the South, year fixed effects, and interviewer characteristics (age, gender, education, and race). SEs account for clustering within interviewers. Marijuana and cocaine use models use data from 1988, 1992, 1994, and 1998; crack usage was only reported separately in 1994 and 1998. Omitted category for drug use is "never."

$^{+}P < .10$.
* $P < .05$.
** $P < .01$.
*** $P < .001$.

is related to being classified as white (as opposed to other characteristics that are associated with reporting marijuana use or other ways in which marijuana use might be cued). We also find little evidence that reporting cocaine use influences racial classification or that interviewers are differentially likely to classify respondents as white based on whether or not they hear this information.

By contrast, not only is reporting crack use associated with how interviewers racially classify respondents, but the association also varies based on whether or not interviewers hear this information.[28] These results suggest that in some cases learning about characteristics, behaviors, and ex-

[28] The different pattern for respondents reporting higher levels of crack usage implies that the relationship between reported drug use and racial classification might vary not only according to the drug involved, but also according to level of use. However, the

periences that are highly racialized can prime Americans to see and think of one another in ways that reinforce preexisting stereotypes.

SUMMARY AND CONCLUSION

Given the breadth and depth of the issues raised by the AIL and KDH comments, our reply has covered considerable ground. We thus briefly summarize our most important empirical clarifications.

1. *We find no evidence that the reciprocal relationship between social status and racial categorization is simply an artifact of measurement error*. This is true both when we use our preferred approach, which removes classification blips from our models, and when we use KDH's approach to introducing random error.
2. *Models predicting racial categorization using survey weights produce similar results as models estimated without weights*. However, racial fluidity poses a challenge to standard weighting schemes, and the NLSY weights are based on a fixed definition of the Hispanic population that differs from the fixed definition of "Hispanic" used by AIL.
3. *Racial fluidity is typically higher in subpopulations thought of as racially ambiguous, but is not zero for any population*. Racial ambiguity itself is socially constructed, and membership in these subpopulations also is not necessarily fixed. Thus, we focused less on variation in the level of fluidity and more on the predictors of racial categorization at a given point in time.
4. *Evidence congruent with a reciprocal relationship between social status and racial categorization can be found across a wide range of subpopulations, cohorts, and data sets*. These observational results are buttressed by experimental evidence suggesting that status cues shape racial classifications, and affect racial perceptions more broadly, even in cases where the category selection does not change.
5. *The mechanisms that link social status to racial categorization could be visual or contextual and are not necessarily related to responses relayed directly during a survey interview*. Yet, in some cases, as with the association we find between reported crack use and being classified as black, explicitly hearing certain information might prime widespread racial stereotypes. Understanding when and how different types of information shape racial perceptions is an important frontier for research.

relatively small number of frequent crack users renders this conclusion highly speculative.

Is It Race That's Rigid?

The distance between our perspective and those of AIL and KDH cannot be measured by the empirical critiques alone; the difference is also marked by a vast chasm in terms of racial conceptualization. Despite KDH's claim to the contrary, social science research has never "documented" that race is rigid (p. 233); social scientists—like most people—simply assumed individuals' categorizations were fixed and, unsurprisingly, kept analyzing their data in ways that reinforced that belief. The comments by AIL and KDH follow in this same vein; underlying both is an assumption that there is a "true" and presumably stable measure of race (or ethnoracial background) out there to be found.

In contrast, we do not think people are "white," or "black," or "Hispanic" as an inherent part of their being. We do think that Americans are regularly put into these categories and that identifying and/or being seen as white or black or Hispanic (or Asian, or mixed race, etc.) has consequences for their opportunities, outcomes, and position in society. We suspect that it matters when, or by whom, one is likely to be racially categorized in ways associated with disadvantage and when, or by whom, one is likely to be advantaged. Thus, the repeated measures of racial classification in the NLSY interest us as proxies for how people are seen across contexts in their everyday lives.

We understand that our colleagues are unsettled by the complexity this perspective introduces into the business as usual of studying racial inequality. However, advocating that social scientists give up the true race assumption is not the same as suggesting that the research community should abandon its effort to understand the role of race in society. Racial categorization continues to be real in its consequences whether or not self-identification and classification are fixed characteristics. If anything, our findings suggest that the meanings of whiteness and blackness are remarkably durable and that the relationship between race and inequality, in the aggregate, is even stronger—not weaker—than is assumed in standard stratification research. This will be true no matter how much individual fluidity there is, as long as that fluidity is selective and the categories themselves maintain their meaning and overall position in the social hierarchy. Fluidity and stability in racial categorization work in tandem to maintain the status quo. Indeed, if race is not necessarily a fixed characteristic, then explaining whose categorization remains stable and why becomes just as important as explaining whose does not.

Thus, although we are flattered that AIL believe our framework for thinking about race and inequality will be more useful in the future—as immigration and interracial marriages continue to diversify the population—we hope researchers will not wait to reexamine the conventional wisdom about race and racial categorization in the United States. Revisiting rigid assumptions, and treating them instead as empirical questions, has the potential to change what we think we know about a lot of things—including the relationship between race and inequality.

Aliya Saperstein
Stanford University

Andrew M. Penner
University of California, Irvine

REFERENCES

Alba, Richard, Noura Insolera, and Scarlett Lindeman. 2016. "Comment: Is Race Really So Fluid? Revisiting Saperstein and Penner's Empirical Claims." *American Journal of Sociology* 122, in this issue.

Alba, Richard, and Tariqul Islam. 2009. "The Case of the Disappearing Mexican Americans: An Ethnic-Identity Mystery." *Population Research and Policy Review* 28 (2): 109–21.

Bound, John, Charles Brown and Nancy Mathiowetz. 2001. "Measurment Error in Survey Data." Pp. 3705–3843 in *Handbook of Econometrics,* vol. 5. Edited by E. E. Leamer and J. J. Heckman.

CHRR (Center for Human Resource Research). 2001. *NLSY79 User's Guide*. Ohio State University, CHRR NLS User Services.

Duncan, Brian, and Stephen J. Trejo. 2011. "Who Remains Mexican? Selective Ethnic Attrition and the Intergenerational Progress of Mexican Americans." Pp. 285–320 in *Latinos and the Economy*, edited by David L. Leal and Stephen J. Trejo. New York: Springer.

Freeman, Jonathan B., Andrew M. Penner, Aliya Saperstein, Matthias Scheutz, and Nalini Ambady. 2011. "Looking the Part: Social Status Cues Shape Race Perception." *PLoS One* 6 (9): e25107.

Hoyt, Gail Mitchell, and Frank J. Chaloupka. 1994. "Effect of Survey Conditions on Self-reported Substance Use." *Contemporary Economic Policy* 12 (3): 109–21.

Kramer, Rory, Robert DeFina, and Lance Hannon. 2016. "Racial Rigidity in the United States: Comment on Saperstein and Penner." *American Journal of Sociology* 122, in this issue.

NORC (National Opinion Research Center). 1978. *National Longitudinal Survey of Labor Force Behavior (NLSY) Interviewer's Manual: Household Screening.* Chicago: NORC.

Peterson, Ruth D. 1985. "Discriminatory Decision Making at the Legislative Level. An Analysis of the Comprehensive Drug Abuse Prevention and Control Act of 1970." *Law and Human Behavior* 9 (3): 243–69.

Rubin, Donald B. 1987. *Multiple Imputation for Nonresponse in Surveys*. Hoboken, N.J.: John Wiley & Sons.

Saperstein, Aliya, and Aaron Gullickson. 2013. "A Mulatto Escape Hatch? Examining Evidence of U.S. Racial and Social Mobility in the Jim Crow Era." *Demography* 50 (5): 1921–42.

Saperstein, Aliya, and Andrew M. Penner. 2012. "Racial Fluidity and Inequality in the United States." *American Journal of Sociology* 113 (3): 676–727.

Saperstein, Aliya, Andrew M. Penner, and Jessica M. Kizer. 2014. "The Criminal Justice System and the Racialization of Perceptions." *Annals of the American Academy of Political and Social Science* 651:104–21.

Simonovits, Gabor, and Gabor Kezdi. 2016. "Economic Hardship Triggers Identification with Disadvantaged Minorities" *Journal of Politics.* DOI: 10.1086/685586.

Young, D. M., Diana T. Sanchez, and L. S. Wilton. 2015. "Too Rich for Diversity: Socioeconomic Status Influences Multifaceted Person Perception of Latino Targets." *Analyses of Social Issues and Public Policy.* doi: 10.1111/asap.12104.

Zerubavel, Eviatar. 1991. *The Fine Line: Making Distinctions in Everyday Life*. New York: Free Press.