Measuring Race/Ethnicity in a "Post-Obama" Era

Keywords: race, ethnicity, identity, racism, group consciousness, mixed race, multi-racial

Introduction and Motivation

Our proposal centers around a new item to measure respondents' racial/ethnic self-identification. This item enables a test of a lingering gap between theory and practice. Theoretically, social science accepts a constructivist view of group identities like race and ethnicity – that they are socially defined and vary across contexts of structural constraint, social interaction, and individual choice. Empirically, while most social surveys now allow respondents to identify with more than one race/ethnicity, the underlying concept in extant measurement remains fixed and categorical. If a more contingent, continuous reality were measured categorically, this gap between theory and practice would be potentially consequential – both to our study of racial/ethnic politics and to any aspects of electoral politics more generally where a person's race/ethnicity is a contributing factor.

We propose here an "Identity Points Allocation" approach that would allow the ANES community to estimate this gap and to study the empirical implications of our racial and ethnic self-identification as an explanatory variable or as an outcome interest in itself. These are not mere matters of passing curiosity. The target population for the ANES continues to grow in its racial/ethnic diversity just as the electoral politics of nation continues to increase in its racialized and polarized tenor. We are in a potentially transformative moment in the salience of racial/ethnic identity. Innovation in our approach to measurement is warranted if we stand a chance of capturing that transformation as it happens *in vivo*.

Background and Previous Surveys

Most surveys today (including the ANES, General Social Survey, and Panel Study of income Dynamics) follow the lead of the 2000 decennial census. Where respondents had previously been forced to choose one among a list of racial/ethnic groups, respondents are now allowed to represent a

mixed race heritage and choose more than one from that list. In measurement terms, these question formats presume that the given response categories validly and reliably capture the extent of how a person thinks of herself in racial/ethnic terms. That is, they presume a question design that is easily interpreted, the information relevant to answering it easily retrieved, judgments based on that information capably made, and those judgments squarely mapped onto given response categories. A problem persists with this presumption. In the pre-2000 Census world, survey respondents with mixed race heritage faced two difficulties: choice is constrained to be *singular* and *equal.* The singularity constraint requires respondents to choose just one racial/ethnic category regardless of whether individuals might in fact be partial to multiple categories. The equality constraint requires, computationally, that two persons who choose the same category are identical with respect to the fit of that category to their life circumstances. The equality constraint conversely also requires that two persons who each choose a different category share nothing in common with respect to their racial attachments or ancestry.

In sum, the verdict on our existing measurement practice hinges on getting an estimate of whether meaningful information is lost in this gap between theory and practice. Current measurement approaches do not allow for this. Surveys like the American Community Surveys, the ANES, GSS, and PSID address the singularity constraint but not the equivalence constraint. There have been, to be fair, a few previous efforts to adopt a more flexible, graded approach to measuring racial/ethnic self-identification. One such study (Brady and Kaplan 2000) uses multiple indicators to develop factor analytic scales and a more graded measure of the ethnic identity of Estonians and Slavs. [1] However, this approach cannot generate estimates of how much variation in racial/ethnic identity is lost by prevailing direct approaches racial/ethnic self-identification used in social surveys.

---

[1] Henry Brady and Cynthia Kaplan, "Categorically Wrong? Nominal Versus Graded Measures of Ethnic Identity," *Studies in Comparative International Development* 35, 3(2000): 56-91.

Another noteworthy effort directly compares four different indicators of race in the National Longitudinal Study of Adolescent Health (NLSAH) – self-identification by teenagers at school, at home, by their parent(s), and by interviewer observation (Harris and Sims 2002; see also Smith 2001).[2] Here the researchers demonstrate variability and social construction by employing an approach that shows how racial/ ethnic composition varies sharply across different contexts. The comparison, however, remains between discrete, categorical indicators of race and retains the equivalence constraint in a measurement approach. Both studies reveal some of the limitations of our existing measures, but we are still left wanting a more direct test of the fluidity or fixity of racial/ethnic self-identification.

We propose a different approach for the 2016 ANES Pilot Study. The intuition is simple: we can make significant headway not only by allowing for multiple selections out of a given set of racial/ethnic response categories but also by allowing for varying intensities of attachment to those selections. Specifically, we propose an Identity Point Allocation (IPA) approach would give respondents a stock of ten "identity points" to allocate as they see fit across a defined set of racial/ethnic response categories. This approach has an analog to electoral choice. The pre-2000 Census approach to measuring race/ethnicity shares a resemblance to the one-person, one-vote system in facing singularity and equivalence constraints. The IPA approach, by contrast, shares a resemblance to cumulative voting and addresses both constraints.

The proposed question format has been tested in several previous surveys: a 2003 telephone survey of adult Californians conducted by the erstwhile Survey Research Center at the University of California, Berkeley; a 2006 Polimetrix web-based module in the Time Sharing Experiments in the Social Sciences series; a 2007 telephone survey of adults in Los Angeles County conducted by

---

[2] David Harris and J.J. Sims, "Who is Multi-racial? Assessing the Complexity of Lived Race. *American Sociological Review* 67, 4 (2002):614-627. Tom Smith takes a similar approach by comparing three approaches – interviewer observation, a census-based "first mention" format, and the self-identification format used in the GSS – and finds "minimal impact on the GSS time-series (2001, 7)."

Interviewing Services of America.[3] More recently, it has also been picked up by the Pew Research

Center in a 2015 survey and it is currently in the final stages of approval and has undergone cognitive

interviewing for inclusion in the 2016 General Social Survey. In each survey, respondents were

asked some variant of the following:

> "In identifying a person's race and ethnicity, we often use just one racial or ethnic category. Sometimes, however, more than one racial or ethnic category is applicable. Imagine if our race and ethnicity could be describe by using a 10 point system to allocate as we see fit to whichever racial and ethnic categories that we think accurately describes a person. For example, if you think of someone as half-white and half-Asian, you might allocate 5 points to each. Or if you think of someone as mostly black but with some Hispanic heritage, you might allocate 9 points for African American and 1 point for Latino. Suppose you are asked to describe your own background this way. How would you describe your race and ethnicity using this 10-point system?"

Specific Items

   We propose the inclusion of a reworded version of the IPA item and, as available survey time

allows, up to eight additional items that we believe would be valuable to us and to the broader ANES

user community in conducting analysis with the IPA item. Our proposed item is as follows:

   – "In describing our background, we often ask someone which racial or ethnic groups best describes them. Imagine if we used instead a 10 point system where points are allocated to whichever racial or ethnic groups the respondent thinks accurately describes themselves.

[Randomize VERSION A for one-half of the sample; and VERSION B for the other half of the sample.]

[VERSION A] For example, if you think of someone as half-white and half-Asian, you might allocate 5 points to each. Or if you think of someone as mostly black but with some Hispanic heritage, you might allocate 8 or 9 points for African American and 1 or 2 points for Latino.

[VERSION B] For example, if you think of someone as just Latino, you would allocate all 10 points to Latino. Or if you think of someone as mixed race, but more African American than white, you might allocate 6 or 7 points to black and 3 or 4 points to white.

---

[3] Please see the Appendix for more details on how responses can differ by question wording, and for pre-test results showing how identity point allocation varies by question format and racial group.

Now think of your own background in racial and ethnic terms. How would you describe your race and ethnicity using this 10-point system?"

- ☐ African American
- ☐ White
- ☐ Hispanic or Latino
- ☐ Asian American
- ☐ Native American or American Indian
- ☐ Other [SPECIFY]

There are two key modifications to the revised IPA item. First, the question wording is significantly trimmed and simplified at the beginning and end of the item. This would ease the cognitive work required to answer this item and reduce the item non-response (which was an issue with the 2003 survey, but not the 2006 and 2009 surveys). In the cognitive interviewing and other pre-testing for the 2016 GSS, this simplified question wording appears to do what is expected of it. Second, we are proposing to randomize between two item versions that vary the examples of point allocations given to the respondent. In version A, one example is an equal 5 points bi-racial allocation and the other example is an unequal bi-racial allocation. In version B, one example is a mono-racial 10 points allocation and the other example is an unequal bi-racial allocation.

This single item has the highest priority for this module. In addition to this item, we also propose to ask respondents to define their parents' racial/ethnic background using the IPA format. These two questions would come after the IPA self-identification item:

-   "How would you describe your (father's / mother's) [ROTATE] race and ethnicity using this 10-point system?"

These items are a high second priority and would allow us to more fully examine the extent to which respondents' family backgrounds define their identity point allocations. These items above on parental racial/ethnic background are also an additional way to pinpoint individuals who identify with a single race (to the currently used ANES race variable) but are of mixed racial backgrounds.

In addition, we would propose three additional items enable a study of the empirical relationship between multiracial identity and one's political orientation. Scholars as far back as Robert Park and

Everett Stonequist have speculated on this relationship and the increasing multiple-race population raises important questions about the future of minority group solidarity in US politics. Yet, we know little about the opinions of multiple-race identifiers and from where those opinions emerge. Thus one very helpful measure to add to this module would be the following companion to the current ANES party identification measure, but asked of respondents' parents. This item should follow the existing party ID variable and fall under the "voting patterns" set of respondent background variables.

- "Generally speaking, does your (mother / father) think of (herself / himself) as a Republican, Democrat, Independent, or what?"

The other political item we propose is to ask respondents to define President Barack Obama using the IPA scale. Measuring respondents' racial definition of Obama using this IPA approach should be of interest to the ANES user community for several reasons. First, it allows some comparison between self-definition and external ascription using the IPA item. Second, it also enables us to examine whether the subjective racial definitions of President Obama are predictive of respondents racial and political attitudes. This item should come after the items asking respondents to describe their parents' race/ethnicity using the IPA approach. The specific item wording we propose is:

- "Now think about Barack Obama. How would you describe President Obama's race and ethnicity using this 10-point system?"

Taken together, the items in this proposed module should give the ANES user community a potentially very useful set of questions with which to better understand the contingent and varied nature of racial/ethnic self-identification, more inclusively estimate the US multiracial population, and delve more deeply into the social, economic, and political consequences of racial/ethnic change *in vivo.*

Appendix One: Four Hypothetical IP Distributions*
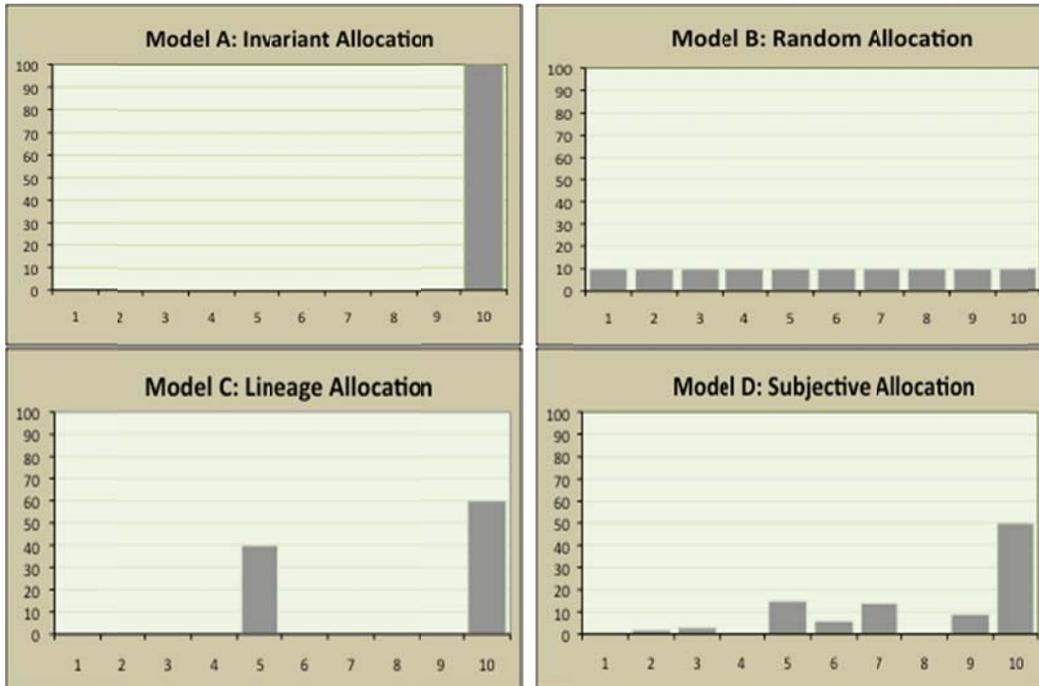
* Appendix One presents four possible distributions of identity points for a hypothetical sample population of respondents:

- "Invariant allocation" (Model A) is one distribution of identity points that is implicit in the current ANES measurement approach. Namely, in this model, all respondents think of themselves mono-racially and allocate all 10 identity points to the category they identify with.
- "Random allocation" (Model B) is a fictionalized realm in which where racial/ethnic boundaries are so arbitrary or contingent that, for a given sample population, every point allocation is equally likely.
- "Lineage allocation" (Model C) is the other possible distribution that is implicit in the current ANES approach. Here, 60% of respondents think of themselves mono-racially and 40% think of themselves in bi-racial terms. When respondents think of themselves in bi-racially, the default is "invariant allocation" of their parental lineage; that is, they evenly divide their 10 identity points between the two racial/ethnic groups.
- "Subjective allocation" (Model D) is a hypothetical distribution in which the allocation of identity points varies across multiple contexts of choice and frames of reference. Here ascribed parental lineage is one among many possible allocations and individual recall of one's parental lineage may be faulty or selective and they may favor one parent's background for a range of subjectively defined reasons. Allocation is also open to other subjective considerations like one's phenotype, geographic origin, racial ascription by others, and other behavioral factors like one's friendship ties, cultural consumption, and racially coded mien.

Appendix Two: IP Allocation Across Racial/Ethnic Groups*
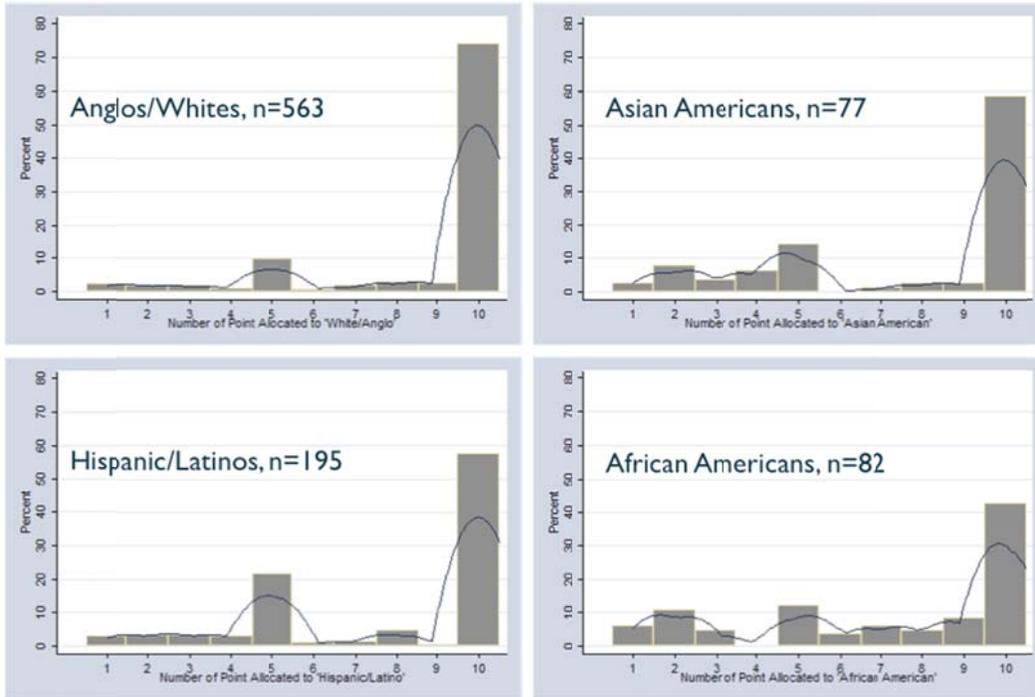
* Data are from the 2003 Golden Bear Omnibus Survey; field dates were 9/17 to 11/17/2003; n=1,050. Respondents in Figure Three below are coded as belonging to a racial/ethnic group if they allocate at least one identity point to that group. The distribution of identity points is shown for the four most common racial/ethnic groups – Whites, Blacks, Latinos, and Asians; the distribution is shown in two non-parametric forms: as a histogram and as a smoothed kernel density.

Appendix Three: Identity Point Allocation by Types of Identification and Group*
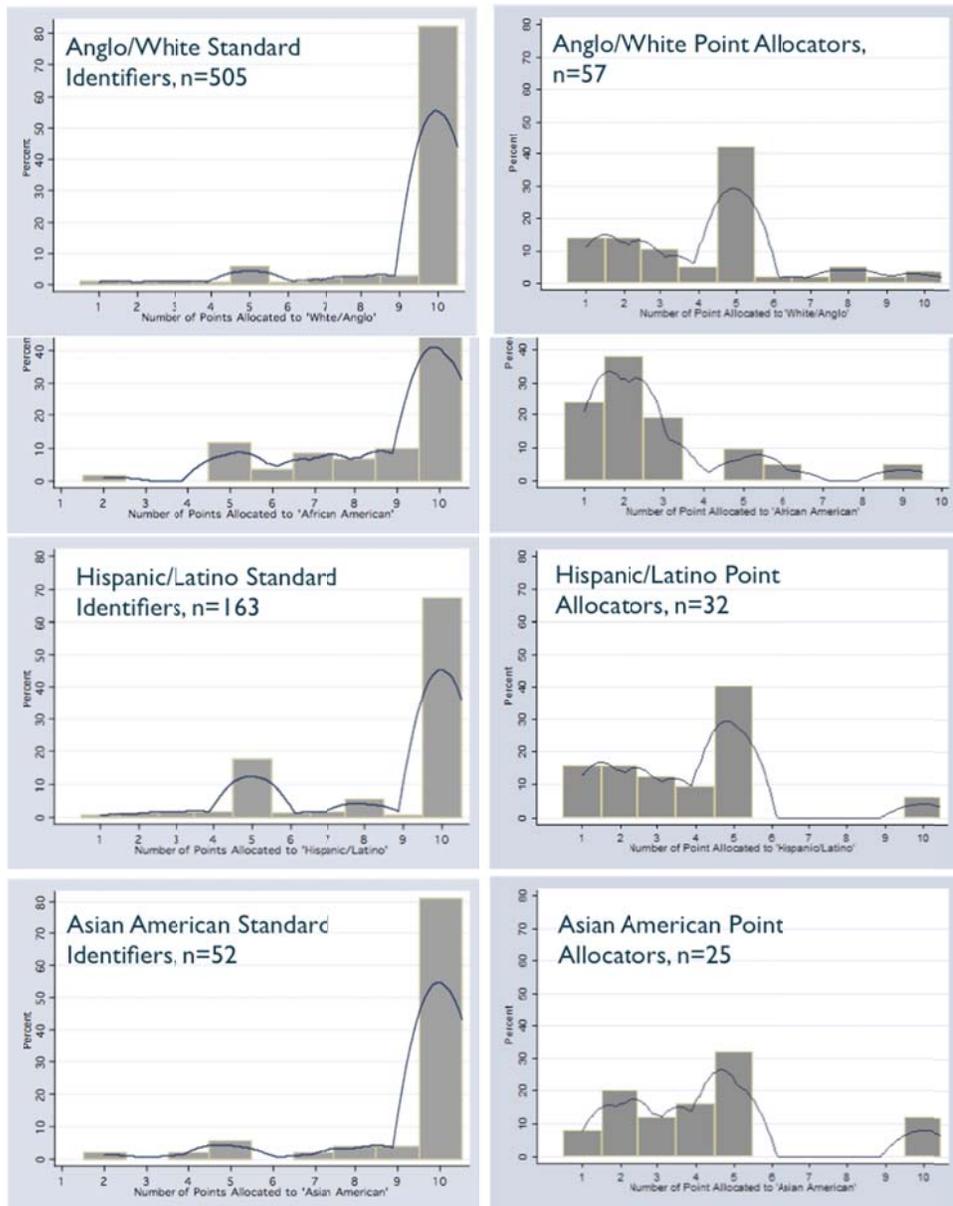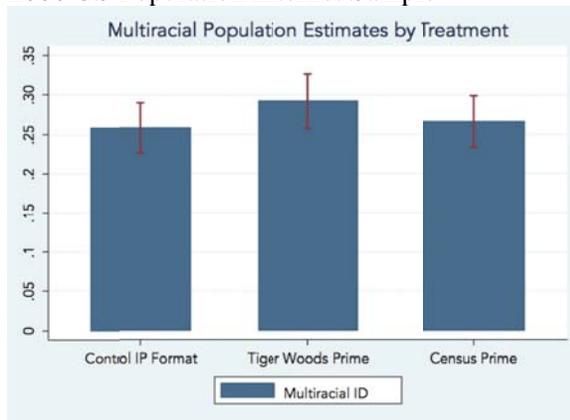

* Data are from the 2003 Golden Bear Omnibus Survey. Figure Three compares two distinct types of identification: (a) those individuals who identify with a given racial/ethnic category in the current ANES question format ("standard identifiers"); (b) those individuals who do **not** identify with that category in the standard format but who nonetheless allocate one or more identity points to that category in the IP format ("point allocators").

Appendix Four. Comparing Estimates of the Mixed Race Population

2003 California Sample*
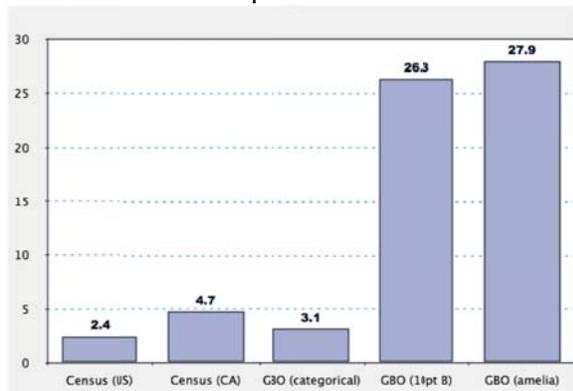
2006 US Population Internet Sample**

* Survey data are from the 2003 Golden Bear Omnibus Survey. Estimates of the US and California multiracial populations are from the 2000 decennial census. The third column is the within-survey estimate of the California multiracial population (using item wording similar to the ANES). The fourth column is the estimate using the IPA item. The fifth column attempts to weight for item non-response (roughly 28% to the IPA format) by using King's approach to multiple imputation (see Gary King, *et al*. 2001. "Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation." *Am. Pol. Sci. Review* 95 (1): 49-69.)

** Data are from a 2006 module in the Time Sharing Experiments in the Social Sciences online surveys. Field dates were 3/13-3/20/2006; n=2,076. Respondents were randomly assigned to one of three versions of the IP allocation question: (1) a "control" version in which respondents are asked the identity point allocation format without any prior question; (2) a "cultural prime" version in which respondents are first asked to define Tiger Woods using the IP format; (3) a "political context" version in which respondents are told, "Now suppose you are asked to describe your own racial background in this way for *governmental purposes,* like the Census questionnaire. How would you describe your race or ethnicity using this 10-point system?" The within-survey estimate of the US multiracial population based on the current categorical measurement approach was 1.6%; the population parameter estimate in the 2006 American Community Survey was 1.9%. [N.B.: all respondents were required to answer this item, so the item non-response was 0%.]