

# Meaning and Measurement

## Reorienting the Race Politics Debate

Michael A. Neblo

*Ohio State University, Columbus*

Is race politics about racism, ideology, or group conflict? After decades, this debate seems scarcely closer to resolution, despite the enormous theoretical, empirical, and normative issues at stake. I argue that a misguided approach to interpreting public opinion has stymied the debate. All three theories implicitly try to read a person's motives for supporting or opposing proposals off of their placement in the so-called complex space of contemporary opinion about race. However, I show that because the supposed complexity of the issue space is based on a methodological artifact, any attempt to read qualitative differences in motives from it must fail.

**Keywords:** *race; symbolic racism; ideology; group conflict*

Why do some Americans favor affirmative action and other policies designed to help blacks, while others oppose them? A simple enough question, yet perhaps no other debate has vexed political science for so long, or so vigorously. Symbolic racism (SR) theorists claim that a fusion of antiblack affect and a kind of punitive version of the Protestant ethic is the central factor determining whites' opinions about race policy. Principled ideology (PI) theorists argue that since the civil rights movement, antiblack affect has receded as the primary determinant of whites' attitudes about racial policy and that political ideology is now the most important factor. Group conflict (GC) theorists agree that antiblack affect is not the main issue, but they suggest that race politics *never was* about affect. Race policy is, and always has been, about whites' attempts to maintain their privileged place in the American social hierarchy.<sup>1</sup>

For PI theorists, white opposition to policies designed to help blacks is typically a matter of reasonable disagreement with left-liberal politics, and as such must be respected. For SR and GC theorists, this constitutes a whitewash of racism and greed and, as such, should be opposed. The empirical and normative issues at stake seem relatively straightforward. Thus the puzzling question is why a debate among such respected scholars, with mountains of evidence, and almost three decades of engagement, appears, to many observers, scarcely closer to resolution.

I argue that the debate has been stymied by a misguided approach to interpreting the meaning of public opinion about race. And this approach has been abetted,

in turn, by a subtle, yet highly consequential, mistake in the *measurement* of public opinion about race. All three contending theories implicitly try to read motives off of a person's placement in the apparently complex space of post-civil rights era opinion about race.<sup>2</sup> However, the purported complexity of the issue space is illusory, and thus any attempt to read qualitative differences in motive attribution off of it must fail. Therefore orienting the debate around the dimensionality of the issue space has led to intractable conceptual problems.

### Complexity and the Issue Pluralism Thesis

Scholars supporting the SR, PI, and GC theses disagree about a lot when it comes to the meaning of public opinion about race politics. However, they all agree that race politics in the United States has become "complex."<sup>3</sup> For example, PI theorists claim

---

**Author's Note:** I would like to thank Paul Sniderman, Don Kinder, Lynn Sanders, David Sears, Jim Sidanius, Larry Bobo, George Marcus, Skip Lupia, Don Green, Michael Dawson, Mark Hansen, Charles Larmore, Cass Sunstein, Nancy Burns, Jake Bowers, Cara Wong, Adam Berinsky, Ted Brader, Craig Volden, Eric MacGilvray, Chad Flanders, Andrew Rehfeld, Chad Cyrenne, Kirk Greer, Scott Blinder, Adam Maze, Tom Nelson, Nick Nelson, Kathleen McGraw, Harwood McClerking, Ismail White, Corrine McConaughy, Eileen McMahon, and Daniel Davis for their generous discussion and comments over many years on various incarnations of this article. In addition, many thanks to the editors and the anonymous reviewers for their patience and persistence in helping me to make this a better paper.

that before the civil rights movement, “the issue of race was, through and through, a matter of right and wrong” (Sniderman and Piazza 1993, 1). It was unidimensional. A person’s attitude regarding one question on the racial agenda correlated closely with the others. PI theorists argue that race politics today exhibits “issue pluralism”—that is, it has broken into three distinct agendas, which they label “social welfare” (SW), “equal treatment” (ET), and “race conscious” (RC). They offer evidence that the correlation among items on different agendas has become modest in the post–civil rights era and conclude that the structure of race politics has shifted.<sup>4</sup>

PI theorists put a lot of effort into establishing the issue pluralism thesis. The crucial question, though, is why distinguishing these agendas is important for motive attribution. Either it is a non sequitur, or they are suggesting that distinct agendas imply that the reasons why one would support or oppose items on these agendas differ systematically. If so, we could begin to infer different motives for opposing or supporting them. This seems to be the tacit argument in *The Scar of Race* (Sniderman and Piazza 1993, 20, emphasis added): “This diversity of responses to racial issues is, we want to suggest, *the single most important feature of contemporary racial politics*. . . . The more consistent [whites’] reactions to racial issues, the more those issues boil down to the same single question; the less consistent their responses, the more they tend to react to particular policies in their own terms.” PI theorists see the issue pluralism thesis as an essential step toward the goal of showing that race politics today is primarily about political ideology (PI), rather than a new kind of racism (SR) or an age-old struggle for power (GC).

However, PI theorists face both an empirical and a theoretical problem with this approach. The empirical problem is that, despite appearances, the *structure* of white opinion about race politics has not changed. “The single most important feature of contemporary racial politics” turns out not to be a real feature of racial politics at all. This finding cuts at the heart of the PI argument, even on its own terms. The theoretical problem is that one can imagine, for example, a staunch libertarian reasonably and sincerely opposing the ET agenda in the same way a traditional conservative might oppose the RC and SW agendas.<sup>5</sup> Thus, if PI theorists want to read motives off of agendas at all, they are doing something that is different from the SR theorists in degree, rather than in kind.

Like PI theorists, SR theorists agree that in the old days, race politics was simpler. Today, in addition to

those who are basically hostile or basically sympathetic to blacks, there are also those who merely appear sympathetic. They go on to explain that many of the people who now say they are sympathetic look and act a lot like those who used to be hostile—only now, they are less overtly so. SR theory suggests that rather than undergoing a thorough change in their feelings about blacks, many whites have learned how to appear respectable. But underlying this veneer, whether they realize it or not, is something very similar to the racial animosity that motivated overt racist behavior in days past. Because of social norms, the framing of this animosity has had to shift from some genetic deficiency to a failing at the level of black culture and individual effort. SR is theorized to consist of a fusion of latent antiblack affect combined with a kind of harsh individualism characteristic of the Protestant ethic (Kinder and Sears 1981).

Note, however, that this formulation makes the racial landscape multidimensional by grafting it onto the legitimating ideology of a broader cultural politics. If symbolic racism is truly a blend of antiblack affect and the Protestant ethic, then it should have a markedly weaker effect on the ET agenda. The whole idea behind the Protestant ethic is that we are warranted in being harsh in denying SW (and a fortiori group based benefits, e.g., the RC agenda) precisely because we presume an antecedent equality of opportunity. Indeed, five of the six items on Kinder and Sanders’s (1996) racial resentment scale implicitly presume equality of opportunity if one answers them in the way indicating racial resentment. So people high on the scale should, at least, not actively oppose the ET agenda as they would the other two. There might be some attenuated correlation in that people high on symbolic racism might believe that society already approximates equal treatment, and thus they might not go out of their way to affirmatively endorse measures to secure it. However, they should not actively oppose them either, as they would the other two. Thus we should still be able to observe a difference. If not, the legitimating ideology is doing no work, and we are back in the realm of plain, if covert, negative affect.

In the other direction, we would expect a somewhat *greater* effect for symbolic racism on the RC agenda. Without endorsing their reasoning, we may note that such people would think that, in addition to offending against responsibility, as does SW, RC policies also offend against the individualism so characteristic of the Protestant ethic. Handouts are bad enough, but handouts based on group membership, rather than individual

criteria, would offend doubly. These theoretically implied differences in the effect of symbolic racism constitute the sense in which SR theorists are *implicitly* committed to a multidimensional conception, along with their *explicit* empirical finding of multidimensionality (Kinder and Sanders 1996).<sup>6</sup>

I want to suggest, however, that there is an important sense in which race politics has remained every bit as simple as before. While the conceptual distinctions that PI theorists propose are surely quite plausible, the question remains as to whether they are an adequate representation of mass opinion, and if so, what meaning this should have for how we conceive race politics. On the basis of the evidence so far, PI theorists have been hasty in concluding that the structure of racial politics has fundamentally shifted. They point out that since the 1980s, some of the intercorrelations between major survey items have become quite modest, inferring from this that race politics cannot be unidimensional anymore (Sniderman and Piazza 1993, 19-27). However, a decrease in the intercorrelations between items is necessary but *not* sufficient for a change in the dimensionality of an issue space.

### Issue Pluralism as a Measurement Artifact

An analogy might help make my point more clearly. Imagine a test designed to measure mathematical ability in a wide range of populations. Since mathematical ability varies tremendously, one would need a very wide range of questions to distinguish all ability levels. Now, the correlation between a problem that asks a participant to solve  $5 + 3 = X$  and a problem that requires the participant to solve a differential equation will be very small. However, it does not follow that the two questions are not measuring the same construct, in this case, mathematical ability. If one cannot correctly solve the simple arithmetic, one will surely not be able to solve the differential equation. As a corollary, if one *can* solve the differential equation, one can certainly do addition. Call this phenomenon (e.g., the relationship between simple arithmetic, high school mathematics, and higher mathematics) *lexical ordering*.

Returning to race politics, let us consider two items drawn from the 1991 National Race and Politics Survey:

1. [There are some large companies where blacks are underrepresented.] Do you think that [these] large companies should be

required to give a certain number of jobs to blacks, or should the government stay out of this?

2. How about laws protecting people—many of whom are [black]—from discrimination in hiring and promotion? Are you [in favor of/opposed to] such laws?

Question 2 belongs to what PI theorists call the ET agenda, and question 1 they categorize as RC. Indeed, the Pearson correlation between these two items is quite modest ( $r = .130$ ). However, if they and other interagenda questions are lexically ordered, then we cannot infer from modest correlations that they belong to a higher dimensional space; that is, the issue dimension may be wide, but the space need not be complex or pluralistic.

In logical terms, the condition for lexicality is that answering the so-called hard question correctly is sufficient for inferring that the so-called easy question is answered correctly:  $H \rightarrow E$ . From this we can derive its contrapositive,  $\sim E \rightarrow \sim H$ , as the second necessary (and jointly sufficient) condition for lexicality. In the terms of two-valued logic, these conditions are equivalent (and thus redundant). However, since we are dealing with empirical data, we cannot expect exceptionless rules, and when we move to probabilities, there are two distinct ways in which data can fail to approach lexicality. We will say that  $X$  is “typically sufficient” for  $Y$  if the conditional probability of  $Y$  given  $X$  approaches 1 ( $P[Y|X] \approx 1$ ) and if the ratio between the conditional and raw probability of  $Y$  is significantly greater than 1 ( $P[Y|X]/P[Y] > 1$ ).<sup>7</sup> Similarly, we will say that  $X$  is “typically necessary” for  $Y$  if the conditional probability of not- $X$  given not- $Y$  approaches 1 ( $P[\sim X|\sim Y] \approx 1$ ) and if the ratio between the conditional and raw probability of not- $X$  is significantly greater than 1 ( $P[\sim X|\sim Y]/P[\sim X] > 1$ ). Translating the conditions for the equal protection and job quota questions given previously, we get  $P(\text{EqProt}_{\text{Support}} | \text{Quota}_{\text{Support}}) \approx 1$  as the first condition and  $P(\text{Quota}_{\text{Oppose}} | \text{EqProt}_{\text{Oppose}}) \approx 1$  as the second. Together, they imply that virtually no one will simultaneously oppose equal protection and support quotas:  $P(\text{EqProt}_{\text{Oppose}} \& \text{Quota}_{\text{Support}}) \approx 0$ .

As we can see from Table 1, the two questions are indeed lexically ordered. The upper right-hand cell of the table has only forty-one out of fourteen hundred participants ( $P[\text{EqProt}_{\text{Oppose}} \& \text{Quota}_{\text{Support}}] = .029$ ). Furthermore, of these forty-one participants, only two graduated from college, suggesting that participants with this answer pattern might not have had as clear an

understanding of the issues at stake. If someone opposes equal protection, then they will almost certainly oppose quotas ( $P[\text{Quota}_{\text{Oppose}}|\text{EqProt}_{\text{Oppose}}] = .85$ ), and if they support quotas, they will almost certainly support equal protection ( $P[\text{EqProt}_{\text{Support}}|\text{Quota}_{\text{Support}}] = .89$ ). However, the obverse of each of these statements is not true; that is, knowing someone opposes quotas tells us nothing about his or her stand on equal protection, and knowing that a person supports equal protection tells us nothing about his or her stand on quotas. The information conveyed via conditional probabilities is not symmetric. Finally, these results are not a degenerate case caused merely by the skew of each variable: multiplying the raw probabilities yields a predicted number of participants simultaneously supporting quotas and opposing equal treatment almost twice as large as what we observe. As in the math test analogy, the questions are lexically ordered. Thus they are part of the same dimension, but discriminate in the tails of that dimension, rather than contributing to the body of the underlying factor.

This finding should not be particularly surprising. Just as math items are ordered because various operations within mathematics presuppose each other, support for various race policies will be ordered because the normative commitments underwriting them are nested. For example, supporting materially enhanced equality of opportunity (SW) presupposes a commitment to formal equality of opportunity (ET). Similarly, both of these are entailed by support for equality of outcomes (RC). They form a logical hierarchy of responses to inequality, and thus we would expect differing rates of support as we move up the chain.<sup>8</sup>

If I am right that the three apparently distinct issue agendas are lexically ordered (approximately), then it may be that the structure of race politics has not changed: there is still one dimension, but people are distributed over a wider range on that dimension. On this conjecture, the well-replicated finding that opinion about race is three-dimensional would be an artifact of the methods used to arrive at it, that is, standard confirmatory factor analysis.<sup>9</sup> PI theorists argue for issue pluralism directly from item intercorrelations. SR theorists go on to estimate a confirmatory factor analytic model on American National Election Survey data and find “three correlated, but distinct packages of opinion on race policy. Federal assistance, equal opportunity, and affirmative action present three different faces of the contemporary political debate over race” (Kinder and Sanders 1996, 297). Notice that these categories are nearly identical to the PI theorists’ SW, ET, and RC agendas.

**Table 1**  
**Quotas by Equal Protection Laws**  
**(Binary Recode)**

	Equal Protection Laws		
	Support	Oppose	Total
Quotas			
Support	322	41	363
Oppose	797	240	1,037
Total	1,119	281	1,400

However, factor analysis assumes that variation in responses to an item is exclusively a function of a person’s position on the underlying trait. The relative difficulty of the item plays no role whatsoever. Thus standard factor analytic scaling is ill suited to deal with a trait that varies over a wide range and is measured with items that have substantially different distributions. The problem is especially acute when the pool of items being scaled is relatively small, the variables are not continuous, or when the jumps in the relative skew between lexically ordered items are abrupt. All of these conditions obtain in the present data, and all of them can lead to artifactual inflation of the dimensionality of standard factor analytic models. Thus the “distinct packages of opinion on race policy” may be only artifactually distinctive.

A different scaling technique, the logistic latent trait model, is designed to handle just such problems. The latent trait model estimates both participant differences on the trait *and* the scale location for each item. An alternating, iterative procedure is used to estimate the parameters. Given starting estimates for each participant’s racial ideology, item locations are estimated and are then used to refine the estimates of the participants’ ideology.<sup>10</sup> Since the range of support for various policies is quite wide, it is very plausible that standard factor analysis is leading us to a false inference about the fundamental structure of white opinion about race politics.

To test the issue pluralism versus center/tail hypotheses more systematically, I administered a two-part survey to a sample of adults (age eighteen plus and a U.S. citizen) drawn primarily from north-eastern Illinois. Though this is a sample of convenience, I took pains to sample across social groups. Demographically, it is a reasonable approximation of adults in the Chicago metropolitan area. It is evenly split on gender (51 percent female, 49 percent male), reasonably spread on age (minimum eighteen years/maximum eighty years,  $\mu = 39.8$ ,  $\sigma = 15.1$ ), and

somewhat more educated (27 percent high school or less, 51 percent college or more). Unsurprisingly for Chicago, ideological self-placement is less skewed (53 percent liberal/47 percent conservative) than party identification (44 percent Democrat/25 percent Republican). Though the sample was racially diverse, the analyses reported here are based on 237 complete white protocols. The survey asked 106 questions about race, including almost all of those used on major national surveys, plus dozens that I adapted or wrote myself.<sup>11</sup>

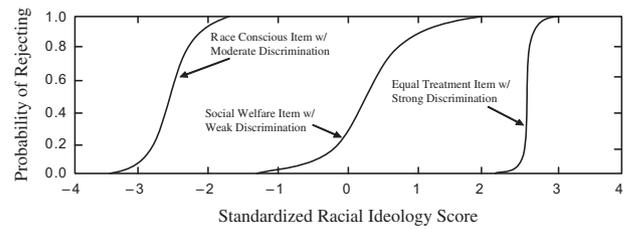
To test the dimensionality and lexicality hypotheses on these data, I estimated a two-parameter logistic latent trait model. Latent trait theory hypothesizes that an individual's survey answers are explained by his or her level on the posited latent trait. Each item's characteristic curve traces the probability of answering a given item in the direction associated with higher levels of the latent trait as a monotonically increasing function of the participant's level on the trait. Thus the model can evaluate how well, and over what portions of a scale, items discriminate between participants. It is given as follows:

$$P_g(\Theta) = \frac{e^{Da_g(\Theta-b_g)}}{1 + e^{Da_g(\Theta-b_g)}}$$

$P_g(\Theta)$  is the proportion of participants with latent trait level  $\Theta$  who answer item  $g$  in the direction positively associated with the trait.  $D$  is a constant,  $a_g$  is the discriminating power of the item for the trait, and  $b_g$  is the difficulty of the item (i.e., extremity). One can think of the model as estimating a logistic regression for each item on the entire scale, where  $b_g$  is the intercept and  $a_g$  is the coefficient. Thus  $a_g$  tells us how well the item discriminates between participants on the trait. Larger coefficients indicate steeper slopes, which indicate a narrower (i.e., more efficient) range of the trait over which we can accurately predict the answer to the item. Similarly,  $b_g$  tells us where on the scale the item is discriminating between participants.

To give a visual sense of how this model and its parameters work, Figure 1 plots the item characteristic curves representing three hypothetical race policy questions. The first item is from the RC agenda and has a large negative difficulty parameter, suggesting that one must be very racially liberal to avoid rejecting the policy. The curve is moderately steep, suggesting that its discrimination parameter is reasonable, that is, that it loads well onto the latent trait. The second item, from the SW agenda, is located near zero on the standardized racial ideology scale,

**Figure 1**  
**Item Characteristic Curves**



meaning that supporting or rejecting the policy discriminates among racial moderates leaning left or right. The curve is not very steep (i.e., it has a smaller discrimination parameter), indicating that the item does not fit squarely onto the main racial ideology dimension. The third item, from the ET agenda, has a large difficulty parameter, suggesting that one must be quite racially conservative<sup>12</sup> to reject the policy. Moreover, the curve's steep discrimination parameter suggests that it loads powerfully onto the latent trait.

When we use the latent trait model to test for lexicality, the results are quite striking. For the purpose of getting clear on how the results bear on our hypothesis, I will start by focusing on six items—a pair of questions from each of the three purportedly separate agendas:

- Q78 If they are not violent, whites have the right to try to exclude blacks from their neighborhoods. (ET)
- Q47 If black people are not getting fair treatment in jobs, the government should see to it that they do. (ET)
- Q100 I'd be willing to have my taxes raised a little to improve education for poor blacks. (SW)
- Q22 Most welfare programs are a waste of taxpayers' dollars. (SW)
- Q99 Preferential hiring of blacks is wrong because it gives blacks advantages they haven't earned. (RC)
- Q12 Because of past discrimination, preference in hiring and promotion should be given to blacks. (RC)

Table 2 shows, just as the lexicality thesis predicts, that the ET items have difficulty parameters on one extreme of the distribution (a given participant must be very racially conservative to oppose them), the RC items are located on the other end, and the SW items are in between. What is more, all of these items have

**Table 2**  
**Difficulty and Discrimination Parameters for Six Items**

Question	Agenda	Difficulty	Discrimination
Q78 (exclude neighbor)	Equal treatment	1.414	1.753
Q47 (fair in jobs)	Equal treatment	1.121	0.779
Q100 (education)	Social welfare	0.558	1.610
Q22 (welfare waste)	Social welfare	-0.307	1.261
Q99 (unearned advantages)	Race conscious	-0.333	1.519
Q12 (past discrimination)	Race conscious	-0.805	1.009
Mean/ <i>SD</i> for all items <sup>a</sup>	All	0.254/1.215	1.045/0.469

a. Nine of the 106 questions were excluded from this model because they exhibit unfolding properties that make it impossible to recover reasonable estimates under the assumptions of the logistic latent trait model. For example, someone could disagree with Q83, "Affirmative action is legitimate only if limited to soliciting a diverse *applicant* pool," either because he or she is too racially conservative to countenance even very limited affirmative action or because he or she is too racially liberal to limit affirmative action to the application stage.

large discrimination parameters *on the same underlying trait*, which is indicative of a single dimension. The ET items discriminate between hard and soft racial conservatives, the RC items discriminate between hard and soft racial liberals, and the SW items discriminate between centrists leaning left and right.

Depending on how one maps each item onto the PI theorists' three agendas, not all questions cleanly conform to the lexicality thesis (though the majority do). Furthermore, in most cases where they do not, there is a good explanation for why. Take, for example, the following National Election Survey (NES) item, which has been widely used in the race politics debate:

Q28 While equal opportunity for blacks to succeed is important, it's not really the government's job to guarantee it. (ET)

This is an archetypal ET agenda item. However, its difficulty parameter did not cluster closely with the other ET questions. On consideration, though, it was not difficult to see why. The question is blatantly double-barreled, and what is more, the barrels are pointed in opposite directions with respect to the latent trait. If one wanted to deny the first clause but assent to the second, what is the appropriate response? Most people would probably respond "agree," but it is remarkably unclear for a question that made it onto national surveys of the very highest quality. It turns out that the second clause was necessary because, without it, the skew in support for "equal opportunity to succeed" was so high that the item was not useful. Put another way, without making the question double-barreled, the item would surely have had the difficulty parameter of an archetypal ET agenda item.

Table 3 presents a formal test of lexicality, looking at all of the relevant items comprehensively.<sup>13</sup> The

logic of the test is simple: if the agendas are approximately lexical, then their items' difficulty parameters should be significantly different from each other since difficulty locates them along a single dimension of racial ideology (analogous to the difficulty of math questions or left-right in politics).

All three tests are both statistically and substantively significant, and in the correct order, providing strong evidence that the three agendas are lexically ordered along a single dimension. Thus, despite some exceptions, the overall results from the larger policy item pool in my data offer strong support for the lexically thesis.<sup>14</sup>

However, the larger item pool in my data might be part of the problem; that is, such a heavy racial focus in the survey might induce a kind of artifactual consistency across items that would cause us to falsely infer unidimensionality. To test for this possibility, I randomly assigned half of the participants to get the race questions in reverse order. The rationale behind the tests is fairly straightforward. If the large number of racially oriented items is inducing consistency, then items asked later in the survey should load more powerfully on the underlying factor. Thus, for each ordering condition, I compare the first ten questions asked on the survey to the last ten questions. Similarly, I compare the same blocks of ten questions *across* the two ordering conditions, that is, when they appear at the beginning of the survey and when they appear at the end. In the context of the latent trait model, the discrimination parameter is analogous to the absolute value of a factor loading, so it is the relevant measure of consistency.

The first row of Table 4 compares the first and last ten questions' discrimination parameters for participants who got the first question order. The second row compares the first and last ten questions

**Table 3**  
**ANOVA on Difficulty Parameter by Agenda**

Tukey's HSD Multiple Comparison	Mean Difference	SE	<i>p</i> Value
Equal treatment vs. social welfare	0.70812	0.24505	<0.024
Equal treatment vs. race conscious	2.16394	0.23215	<0.000
Social welfare vs. race conscious	1.45583	0.22197	<0.000

for the reverse-question-order participants. The third row compares the parameters for Q1–10 when they appear at the beginning of the survey versus the end of the survey (in the reverse order condition),<sup>15</sup> and the fourth row compares Q97–106 when they appear at the beginning of the survey versus the end of the survey. None of the four tests are even remotely significant,<sup>16</sup> suggesting that reactivity is not driving consistency across items.

However, since my data are not drawn from a national random sample, I want to bolster my argument by reanalyzing NES data on the seven items used to derive the three-factor model. While the latent trait model's parameter estimates are not reliable with so small of an item pool, factor analysis using polychoric correlations can correctly scale items that are skewed and noncontinuous, and thus we can get a test of unidimensionality, if not of lexicality.<sup>17</sup> Polychoric correlations posit continuous latent response variables underlying the ordinal scale of typical Likert items and correct the Pearson correlation for the known deviations from its assumptions. Otherwise, the factor analysis proceeds as normal; that is, we use the same procedures but with a more appropriate correlation matrix. The results from the reestimation using polychoric correlations differ importantly from Kinder and Sanders's (1996) original findings.<sup>18</sup>

They estimate two models, starting with the simplest possible model, that is, a single latent factor. Obviously, this constitutes a test of the unidimensionality thesis. They reject this model decisively (Kinder and Sanders 1996, 295): "This specification is clearly wrong. . . . The single factor model fits the variance-covariance matrix poorly. . . . Chi-square with 14 degrees of freedom = 163.31 ( $p < .01$ ), adjusted goodness of fit = .888. . . . Opinion on race policy is evidently not single-minded." It would appear that the issue pluralism thesis, in some form, must be true.

The model that they settle on postulates three correlated dimensions: federal programs, equal opportunity (EO), and affirmative action (AA; with one item, school desegregation, loading on both the EO and AA factors). Kinder and Sanders (1996, 295) refer to this specification as "a slightly more elaborate model."

I do not think that this description is apt. If one needs three oblique dimensions and an item with multiple factor loadings to explain only seven questions, the model is hardly parsimonious. This is not a criticism in that if the data are complexly structured, then there is nothing to be done about it. Reality sometimes resists our attempts at simplification. However, for purposes of comparing the results of my reanalysis, I want to highlight the baroque structure invoked to secure an adequate fit for these data. Nonetheless, the fit is indeed very good: chi-square = 13.20 (prob. = .213), adjusted goodness of fit index = .988. Note that Kinder and Sanders's three factor labels are very close to the PI theorists' SW, ET, and RC agendas, respectively. Their results lend strong support, not just to the issue pluralism thesis, but to the PI theorists' precise formulation of it.

When I reestimated their models using a matrix of polychoric correlations, the picture changed dramatically. The chi-square statistic on the one-factor model came out to 41.34 ( $p < .01$ ), a rather impressive 75 percent reduction from the original model, though still fairly large. However, the adjusted fit index jumped all the way to .981, a huge improvement and a very solid fit even in absolute terms (i.e., it only differs from the fit of Kinder and Sanders's [1996] preferred model in the third decimal place). What are we to make of these somewhat contradictory results? We can get some leverage on this question by looking at the diagnostics for the reanalysis of the more complex model. Here, the chi-square came in at 12.11, yielding a borderline rejection ( $p = .09$ ) but a near-perfect fit index of .996.

Clearly one of the statistics is misleading in that the chi-square points toward further complicating a model that is almost perfect (and probably already overfitting the data). Bollen (1989, 267) lists "several reasons for exercising caution in the use of the chi-square estimates," two of which are particularly germane to the present case. First, the chi-square approximation assumes that the observed variables have no kurtosis, and simulation studies indicate that skewness "leads to excessively large chi-square estimates." The data in this model are both highly skewed and leptokurtic

**Table 4**  
***t*-Tests on Discrimination Parameters for Order Effects**

Order	Questions	<i>M</i>	<i>SD</i>	<i>t</i>	<i>df</i>	<i>p</i> Value
Standard	1–10 vs. 97–106	–0.215	0.8054	–0.844	9	<0.420
Reverse	1–10 vs. 97–106	0.2131	0.67105	1.004	9	0.342
Between	1–10	–0.0413	0.3962	–0.33	9	0.749
Between	97–106	0.0394	0.93082	0.134	9	0.896

(overpeaked).<sup>19</sup> More important, the chi-square assumes that the null hypothesis  $\Sigma = \Sigma(\Theta)$  holds exactly. Put differently, we can see by rearranging terms that in this context, the chi-square constitutes a joint test that every element of the residual matrix  $\Sigma - \Sigma(\Theta)$  equals zero. This is a severe test, and with a sufficiently large sample, only a degenerate model could avoid rejection. As Bollen (1989, 268) puts it, “the situation is analogous to rejecting the null hypothesis of a zero correlation coefficient of, say, 0.03 . . . [a] large sample may give us confidence that the population correlation is unlikely to be zero, but the substantive significance of a 0.03 correlation in most situations is nil. So it is with the chi-square test.” It would seem that a model with a fit index over .98 constitutes a case in point. One or more elements of the residual matrix are probably not strictly equal to zero, but the substantive significance of this fact is nil.

So, should we prefer the maximally simple model with an adjusted fit index of .98? Or a model that uses three oblique dimensions and an item with multiple factor loadings to explain only seven questions, and gains us a trivial amount of fit? Given that Kinder and Sanders (1996, 295) describe one of their other models with a *lower* fit index (Adjusted Goodness of Fit Index = .97) as fitting “splendidly,” and that the parsimony fit index for the three-factor model actually goes *down* significantly, the answer is clear.<sup>20</sup> Opinion about race policy is unidimensional, and thus the issue pluralism thesis is false.

### What Does the Single Dimension Measure?

The first thing to note is that whatever the new scale measures, it does so reliably. The unidimensional factor solution for the seven-item NES battery yields an internal reliability coefficient  $\rho$  of .895. However, we know that the reliability of a scale tells us almost nothing about its validity. And questions about what a scale measures are really questions about validity.

The NES battery represents a good sample of important race policy questions, so it has a certain amount of *content* validity, if we restrict ourselves to the policy domain. As a result, the unidimensional scale is also likely to have *criterion*-related validity in predicting a participant’s position on new policy questions. However, the race politics debate hinges on explaining the reasons underlying support and opposition to those policies, so the scale’s manifest content and ability to predict new policy positions beg all of the crucial questions.

Earlier research pursued the *construct* validity of the three agendas by showing somewhat different associations with key variables such as education and ideology as well as affect toward blacks in general and black political figures. However, there is very little evidence of such differential relationships in the NES data. Table 5 compares the Pearson correlations of the old three-dimensional solution’s factor scores (i.e., for the ET, SW, and RC agendas) with ideology, education, the NES feeling thermometers for blacks (affect) and Jesse Jackson (candidate). In every case, the three correlations with a given construct validity variable are of the same sign, and the differences in magnitude are quite modest. Indeed, *none* of the twelve pairwise tests for differences in *r* (i.e., ET-SW, ET-RC, and SW-RC each for ideology, education, affect, and Jackson) reaches statistical significance at the .05 level, even without correcting for multiple comparisons. Thus the unidimensional analysis is not obscuring significant construct differences.

In the math test analogy, we had a straightforward interpretation of an item’s difficulty. Items with higher difficulty parameters were more difficult to answer correctly, and those who could answer them had more mathematical ability. But what are the analogues to *correct* and *ability* in the context of opinion about race? One possible interpretation comes from PI theorists themselves. As I discussed previously, PI theorists link their overall argument closely to the issue pluralism thesis. They concede that if issue pluralism is false, “racial policies tend ‘to boil down to

**Table 5**  
**Pearson's *r*: Factor Scores × Construct**  
**Validity Variables**

	Ideology	Education	FT: Blacks	FT: Jackson
Equal opportunity	-.22	.131	.388	.432
Social welfare	-.31	.104	.419	.432
Race conscious	-.268	.031	.417	.539

Note: FT, feeling thermometer.

Source: Data from 1986 National Election Study.

the same single question: Are you sympathetic to [blacks] as a group, are you indifferent to them, or do you dislike them?" (Sniderman and Piazza 1993, 18). Given this concession, the evidence for unidimensionality presented previously would be devastating to the PI position. On PI theory's own terms, we would be compelled to interpret those low on ability as sympathetic toward blacks, and those high on it as disliking them. Indeed, ability here would be a propensity to feel racial antipathy. But PI theorists would surely not want to grant this interpretation. Nor should they.

PI theorists should never have conceded that scaled policy questions necessarily boil down to *any* single thing relevant to motive attribution, much less sympathy–antipathy. The whole premise that motive attribution can be read off of dimensionality has confounded this debate. Without reformulation, the unidimensionality results discussed previously would also confound SR theory because their construct could not have differential effects on the various agendas, as a commitment to the Protestant ethic would imply.

Nonetheless, for some people, intuitively, it seems almost certain that the unidimensional scale *would* be a valid proxy for racism. However, just as surely, for other people, the answers reflect their conservative (or liberal) principles. And for yet others, their scores are driven by a perception of group conflict, or some mix of these three. However, if we think that there might be such heterogeneity in motives, then standard discriminant validity tests that do not account for it will only further confuse the issue. The key problem for the race politics debate is to sort *which* participants are motivated by what. Correlates of the single dimension on an undifferentiated sample will not help us here if unit homogeneity does not hold. Policy factor scores themselves will not help either. Nor will the content validity of the questions since their usefulness for inferring motives is precisely what is at stake. So, unless one wishes to beg the

question mightily, *ability* and *correct* lose much of their normative connotations here since there is no obvious mapping between scores and motives. This line of approach to resolving the race politics debate has exhausted itself.

None of this is to say that there is no hope of making progress in adjudicating between the contending theories. Quite to the contrary, these results set the stage for such progress. However, disaggregating the relevant heterogeneity in motives is a complex task that will require separate, extended consideration using a new approach. In concurrent research (Neblo 2008), I show how accounting for such heterogeneity can move us toward resolving this thirty-year debate by demonstrating that each of the major theories has part of the story right, but for independently identifiable subsets of participants.<sup>21</sup> This article clears the way for this move by showing how an exclusive strategy of parsing concepts has run out of steam, setting up the move in concurrent work of showing how these concepts take on differential significance in a taxonomy of *people*.

## Conclusion

The issue pluralism thesis in race politics is based on a methodological artifact. The various policy questions that make up what were thought to be three separate issue agendas are lexically ordered in a way that confounded standard factor analytic attempts at scaling them. More appropriate methods confirm that while the scope of race politics may have expanded over time, the policy space remains structurally simple. Yet this subtle statistical mistake has led the race politics debate down a very long blind alley. Insisting that the single race dimension means the same thing for everyone beyond its manifest behavioral content—a general tendency to express positive opinions about blacks, consider their socioeconomic situation a collective problem, and support policies intended to help them—would just reproduce the mistake that has vexed us for so long. Thus breaking the stalemate will require a new strategy for getting leverage on motive attribution.

## Notes

1. These are highly compacted versions of more subtle theories, the details of which I discuss as they become necessary. Many excellent scholars have participated in this debate, so it is with some hesitation that I single out Sears and Kinder as the leading figures in the SR camp (see Kinder and Sears 1981; Kinder and Sanders 1996; Kinder and Mendelberg 2000; Sears 1994;

Sears et al. 1997; Sears and Henry 2005; Tarman and Sears 2005). SR has also been known as the *New Racism*, and *racial resentment*. While there are important differences in formulation, I make a general argument, and so for simplicity, I have chosen to stick with the original label. *Principled ideology* is my term. Sniderman has been the most vigorous proponent of the PI thesis (Sniderman et al. 2000; Sniderman and Carmines 1997; Sniderman and Piazza 1993; Sniderman and Tetlock 1986; Kuklinski et al. 1997). Bobo is probably the most prominent GC scholar (Bobo 1983, 1998, 2000). Sidanius's social dominance theory goes well beyond the context of American race politics, so, though theoretically important, it exceeds the scope of this analysis (Sidanius and Pratto 1999). The volume edited by Sears, Sidanius, and Bobo (2000) deserves special mention as the most direct engagement of these competing theories to date. Amazingly, this is a highly selective list of contributions to this debate.

2. Throughout, when I refer to "opinion about race," I mean the opinions of whites. The number of black, Asian, and Latino respondents in my sample does not allow for adequate separate analyses.

3. The nature and significance of that complexity, however, is a matter of some dispute. For example, while SR theorists acknowledge three agendas, they sometimes argue for their practical equivalence on the grounds that the various agendas have similar determinants (see, e.g., Tarman and Sears 2005).

4. To my knowledge, no one actually has pre-civil rights era public opinion data comparable to the post-civil rights era data. Thus one cannot prove a structural shift directly, nor am I aware of anyone who tries to do so by other means. However, neither am I aware of anyone who has been particularly keen to challenge the intuitive claim that the issue was unidimensional before the civil rights movement.

5. In "Motive Matters: Liberalism and Insincerity" (Neblo, n.d.), I develop the normative implications of conjoining the empirical and theoretical problems posed here.

6. To be clear, these are my arguments. No SR theorists argue for the differential effects that I do, and some argue, to the contrary, that the uniform effects that we actually observe bolster the theory because they imply internal consistency for the construct (e.g., Tarman and Sears 2005).

7. Determining what constitutes a satisfactory "approximation to 1" depends on how much information is added by a given condition (in our case,  $X$  or  $\sim Y$ , respectively). This is where the second criterion comes in: the ratio between the conditional and raw probabilities is a measure of the informational value-added of knowing the condition.

8. This schema is obviously general and approximate. For example, some forms of affirmative action (e.g., recruiting a diverse applicant pool, as opposed to quotas) follow the logic of SW since they have the properties of enhanced equality of opportunity vs. rigid equality of outcome. However, such items exhibit unfolding properties consistent with the story that I am telling about a logical hierarchy.

9. Guttman (1950) was the first to correctly scale lexically ordered data. However, Guttman scaling has been replaced by these more powerful procedures. Polychoric factor analysis is designed to handle skewed ordinal items within a general factor analytic framework. It is closely related to the logistic latent trait model that I present later. Bock, Gibbons, and Muraki (1988) lay out the advantages of the item response theory (IRT) approach. However, Knol and Berger (1991) argue that polychoric factor analysis performs almost as well on all but unusual data sets. IRT

has one advantage not discussed in the literature, though, namely, that one can formally test for lexicality with it.

10. This model is very similar to models that have become fairly prominent in political science via the spatial analysis of roll call votes and ideal point estimation. For example, Poole and Rosenthal (1985) use a special case of the general latent trait model to estimate NOMINATE scores for Congressional role call votes. Clinton, Jackman, and Rivers (2004) generalize this approach, and Martin and Quinn (2002) use it to estimate ideal points for Supreme Court justices. Delli Carpini and Keeter (1996) also use IRT to build their knowledge index.

11. For the complete list of race questions, see <http://polisci.osu.edu/faculty/mneblo/papers.htm>. The race items are not balanced in terms of the direction for scoring. However, item analysis suggests that this is not what is driving the unidimensional solution.

12. I use the phrase *racially conservative* here to bracket the "is it really racism" question. (Wood 1994; Tetlock 1994; Sears 1994). I am open to the idea that some, perhaps many, racially conservative whites are also racist. In the present context, I merely want to avoid prejudicing the matter, so to speak. In "Three-fifths a Racist: A Typology for Analyzing Public Opinion about Race" (Neblo, 2008), I try to begin disentangling these phenomena.

13. Twenty-two of the items sort cleanly onto one of the three policy agendas. ET consists in Q28, Q38, Q47, Q59, Q78, and Q87. SW consists in Q8, Q13, Q20, Q22, Q29, Q48, and Q100. RC consists in Q1, Q12, Q25, Q53, Q65, Q70, Q77, Q89, and Q99.

14. The latent trait model does not allow for a direct, formal test of unidimensionality of the kind that I present for the NES data. However, all three agendas have items with very large discrimination parameters, indicating that they load quite well onto the single latent trait. Moreover, I did test for significant differences on discrimination parameters by agenda. None of the post hoc comparisons were statistically significant at .05, suggesting that all three agendas' items load onto the latent trait equally well. These results constitute an indirect test of unidimensionality on these data. However, I rely on the NES data for the stronger test of dimensionality, and these data for the stronger test of lexicality.

15. I ran the same analyses using five items and twenty items as well, and the results were the same. Obviously, when Q1–10 appeared at the end of the survey, they had different numbers for the participants who got that ordering condition. However, I had to fix one designation for purposes of analysis and discussion.

16. Clearly none of the four tests are remotely significant statistically. However, the mean difference in the first and second rows might seem marginally significant on substantive grounds. For present purposes, though, the difference is misleading. Note that the two means are of almost exactly the same magnitude, though of opposite sign. This means that the two sets of ten items naturally differ in how much agreement they elicit, but without reference to where they appear on the survey. The sign gets reversed because I reversed which set was getting subtracted from the other for purposes of the calculation; that is, the difference between them attributable to position is actually substantively miniscule:  $.21500 - .21310 = .0190$ .

17. Though I have not been able to find a discussion in the literature, I believe that it may be possible to test for lexicality using polychoric factor analysis. The technique uses so-called cut points, which I conjecture can be used to derive estimates analogous to the difficulty parameter in the latent trait model. Polychorics presume that the underlying variable is normal, so it

can only partially correct for high degrees of skew in the observed ordinal scale. However, partial correction is conservative since full correction would only strengthen my argument.

18. For a prominent application of polychoric factor analysis in political science, see Niemi, Craig, and Mattei (1991). I have not been able to replicate Kinder and Sanders's (1996) results precisely (though the differences are small), perhaps because of the management of missing data and *don't know* (DK) responses. However, I ran the analyses several ways, and the basic pattern holds up. Optimal imputation and pairwise procedures (as implemented in the major software packages) cannot calculate polychoric correlations. Thus I ran my analyses first using listwise deletion, and second, by recoding missing and DKs to the median. EQS offers two ways to handle ordinal data. I ran both an asymptotically weighted generalized least squares procedure (AWGLS) as well as a maximum likelihood estimation procedure that produces corrected tests and *SEs*. The results are based on the AWGLS estimates using the list-wise deleted data with robust tests and indices.

19. The chi-square's problem with skew and kurtosis is not perfectly corrected by using polychoric correlations, and the robust statistics are limited in their ability to correct for severe skew and kurtosis, except with extremely large data sets.

20. For the complex model, PGFI = .365, and for the simple model, PGFI = .495. Since EQS does not provide a PGFI, these two analyses were run in LISERL using a slightly different procedure from the EQS models.

21. In Neblo (2008), I develop and deploy this strategy in much greater detail.

## References

- Bobo, L. 1983. Whites' opposition to busing: Symbolic racism or realistic group conflict? *Journal of Personality and Social Psychology* 45: 1195-210.
- . 1998. Race, interests, and beliefs about affirmative action: Unanswered questions and new direction. *American Behavioral Scientist* 41: 985-1003.
- . 2000. Race and beliefs about affirmative action: Assessing the effects of interests, group threat, ideology, and racism. In *Racialized politics: The debate about racism in America*, edited by D. O. Sears, J. Sidanius, and L. Bobo, 137-64. Chicago: University of Chicago Press.
- Bock, R., D. R. Gibbons, and E. Muraki. 1988. Full-information item factor analysis. *Applied Psychological Measurement* 12: 261-80.
- Bollen, K. 1989. *Structural equations with latent variables*. New York: John Wiley.
- Clinton, J., S. Jackman, and D. Rivers. 2004. The statistical analysis of roll call data. *American Political Science Review* 98: 355-70.
- Delli Carpini, M., and S. Keeter. 1996. *What Americans know about politics and why it matters*. New Haven, CT: Yale University Press.
- Guttman, L. 1950. The basis for scalogram analysis. In *Measurement and prediction*, edited by S. A. Stouffer, L. Guttman, E. Suchman, P. F. Lazarsfeld, S. Star, and J. A. Clausen, 78-96. Princeton, NJ: Princeton University Press.
- Kinder, D. R., and T. Mendelberg. 2000. Individualism reconsidered: Principles and prejudice in contemporary American opinion. In *Racialized politics: The debate about racism in America*, edited by D. O. Sears, J. Sidanius, and L. Bobo, 44-74. Chicago: University of Chicago Press.
- Kinder, D. R., and L. M. Sanders. 1996. *Divided by color: Racial politics and democratic ideals*. Chicago: University of Chicago Press.
- Kinder, D. R., and D. O. Sears. 1981. Prejudice and politics: Symbolic racism versus racial threats to the good life. *Journal of Personality and Social Psychology* 40: 414-31.
- Knol, D. L., and M. P. Berger. 1991. Empirical comparison between factor analysis and multidimensional item response models. *Multivariate Behavioral Research* 46: 457-77.
- Kuklinski, J. H., Paul M. Sniderman, Kathleen Knight, Thomas Piazza, Philip E. Tetlock, Gordon R. Lawrence, and Barbara Mellers. 1997. Racial prejudice and attitudes toward affirmative action. *American Journal of Political Science* 41: 402-19.
- Martin, A., and K. Quinn. 2002. Dynamic ideal point estimation via Markov chain Monte Carlo for the U.S. Supreme Court, 1953-1999. *Political Analysis* 10: 134-53.
- Neblo, Michael A. n.d. Motive matters: Liberalism and insincerity. <http://polisci.osu.edu/faculty/mneblo/papers.htm>.
- . 2008. Three-fifths a racist: A typology for analyzing public opinion about race. *Political Behavior* 30, no. 3 (April 10), doi:10.1007/s11109-008-9060-2. <http://www.springerlink.com/content/y70pk8p76837pu2p/fulltext.pdf>.
- Niemi, R., S. Craig, and F. Mattei. 1991. Measuring internal political efficacy in the 1988 National Election Study. *American Political Science Review* 85: 1407-13.
- Poole, K., and H. Rosenthal. 1985. A spatial model for legislative roll call analysis. *American Journal of Political Science* 29: 357-84.
- Sears, D. O. 1994. Ideological bias in political psychology: The view from scientific hell. *Political Psychology* 15: 547-56.
- Sears, D. O., and P. J. Henry. 2005. Over thirty years later: A contemporary look at symbolic racism. *Advances in Experimental Social Psychology* 37: 95-150.
- Sears, D. O., J. Sidanius, and L. Bobo, eds. 2000. *Racialized politics: The debate about racism in America*. Chicago: University of Chicago Press.
- Sears, D. O., C. VanLaar, M. Carrillo, and R. Kosterman. 1997. Is it really racism: The origins of white Americans' opposition to race-targeted policies. *Public Opinion Quarterly* 61: 16-53.
- Sidanius, J., and F. Pratto. 1999. *Social dominance: An intergroup theory of social hierarchy and oppression*. New York: Cambridge University Press.
- Sniderman, P. M., and E. G. Carmines. 1997. *Reaching beyond race*. Cambridge: Harvard University Press.
- Sniderman, P. M., Gretchen C. Crosby, and William G. Howell. 2000. The politics of race. In *Racialized politics: The debate about racism in America*, edited by D. O. Sears, J. Sidanius, and L. Bobo, 236-79. Chicago: University of Chicago Press.
- Sniderman, P. M., and T. Piazza. 1993. *The scar of race*. Cambridge, MA: Belknap Press/Harvard University Press.
- Sniderman, P. M., and P. E. Tetlock. 1986. Symbolic racism: Problems of motive attribution in political analysis. *Journal of Social Issues* 42: 129-50.
- Tarman, C., and D. O. Sears. 2005. The conceptualization and measurement of symbolic racism. *Journal of Politics* 67: 731-61.
- Tetlock, P. E. 1994. Political psychology or politicized psychology: Is the road to scientific hell paved with good moral intentions? *Political Psychology* 15: 509-29.
- Wood, J. 1994. Is symbolic racism racism? A review informed by intergroup behavior. *Political Psychology* 15: 673-86.