Common Attitudinal and Social-Psychological Scales May Exaggerate the Dimensionality of Human Differences

Alan S. Gerber*[1], Gregory A. Huber[2], Mackenzie Lockhart[3], Jack D. Walker II[4]

This Draft: 16 July 2026

Psychometric scales have proliferated in political science research to explain important outcomes, treating each new scale (e.g., racial resentment, social dominance orientation, authoritarianism, etc.) as measuring a distinct underlying construct, trait, dimension, or factor. We analyze approximately 180 items from 37 commonly used scales to show that many of these scales are highly correlated. We demonstrate that correlated but non-causal scales can appear statistically significant by inheriting variance from the true causal construct, inflating associations and false positives when tested in isolation. At the item level, we find that constituent items of scales are often correlated strongly with items from other scales, suggesting limited independence. Exploratory factor analysis reveals six latent factors that cut across established scale boundaries. The first explains a large share of the common variance and powerfully predicts partisanship and vote choice. These results underscore the need for greater attention to redundancy in political psychology measurement.

Word count: 12,363

---

*Authors are listed in alphabetical order.

[1] Sterling Professor of Political Science, Institution for Social and Policy Studies, Yale University. Email: alan.gerber@yale.edu.

[2] Corresponding Author: Forst Family Professor of Political Science, Institution for Social and Policy Studies, Yale University. Email: gregory.huber@yale.edu.

[3] Independent Researcher. Email: mackenzie.lockhart@gmail.com.

[4] Pre-doctoral Fellow, Institution for Social and Policy Studies, Yale University. Email: jack.walker@yale.edu.

Transparency statement: The authors declare the use of generative AI in the research and writing process. According to the GAIDeT taxonomy (2025), the following tasks were delegated to GAI tools under full human supervision: Data analysis. The GAI tool used was: Claude Sonnet 4.6. Responsibility for the final manuscript lies entirely with the authors. GAI tools are not listed as authors and do not bear responsibility for the final outcomes. Declaration submitted by: Collective responsibility.

Acknowledgments: We thank Kevin Arceneaux, Andrew M. Engelhardt, and Jennifer Jerit for helpful comments. We thank Matt Blyth, John J. Cho, Philip Moniz, and Itamar Yakir for assistance.

Psychometric scales are widely used in social scientific research generally and political science research specifically. One of the striking features of contemporary scholarship is that the number of scales that researchers have identified and use to predict important political outcomes continues to grow over time. Given that researchers have identified so many different psychological characteristics that predict important outcomes, an important question is whether this growth of scales reflects the discovery of new underlying human differences.

Answering this question requires assessing scale validity, including how independent these scales are from one another. If conceptually independent scales are nonetheless correlated as measured, then a researcher who uses only a single psychological scale to predict an outcome may produce biased estimates. This is because, in a model that excludes other scales, the estimated effect of the single scale will proxy the effect of other correlated scales that also affect the outcome. There is also the possibility that the underlying traits these scales measure are not independent, in which case different scales could reflect different ways of measuring a smaller set of characteristics of individuals. If so, the proliferation of scales may not reflect the discovery of new human differences, but instead the relabeling of already identified differences.

Prevailing researcher conventions for assessing the validity of psychometric scales in political science often focus on demonstrating internal consistency (e.g., a high Cronbach's alpha, a measure of the correlation among scale items) and, sometimes, discriminant validity (e.g., using confirmatory factor analysis to show that survey items from different scales load onto distinct dimensions). Most social science work also uses demonstrations of "positive"

predictive validity (i.e., whether a scale predicts an outcome the construct should affect according to a specific theory) to illustrate the validity (and importance) of a scale.

But high internal consistency does not address the question of whether scales are distinct from one another. And, in practice, tests of discriminant validity are usually applied to pairwise comparisons of a scale to only a small number of other scales. Moreover, even if two scales (and their component items) are only modestly correlated, they might still measure, in part, the same underlying factor(s), such that a novel scale might offer very little predictive power once the shared factor(s) measured across multiple scales is (are) accounted for. Finally, many tests of predictive validity are limited in that they do not simultaneously account for correlated scales, which can generate spurious associations in both scale validation and theory testing.[5]

To clarify these concerns, we identify two related yet distinct problems. The first we label the *estimation problem*, which is about model specification. When scholars estimate regressions using a single scale, they assume that other related scales are either (1) uncorrelated with the scale in question or (2) are irrelevant to the outcome. If any omitted scale is correlated with the one being studied, the first assumption is not satisfied. If this is the case, and the second assumption does not also hold, then the estimated relationship may reflect omitted variable bias instead of true causality. Crucially, this is not a problem of mere correlation—it arises only once a theoretical model specifies which (omitted) constructs matter.

[5] One limitation of confirmatory factor analysis (CFA) is that it starts from an assumption that a scale measures a unique construct, rather than letting the correlation of items across scales be a measure of any underlying shared constructs. In short, CFA seeks to validate the assumption that a scale captures a shared trait, which we argue ignores the possibility that assumption is false, particularly if individual items across scales measure shared constructs.

For example, consider the following scenario: one researcher believes scale A affects an important outcome through a specific theoretical mechanism and another researcher believes that scale B does so through a distinct mechanism, but both researchers estimate their models in isolation. In this scenario, if A and B are correlated, each model would be mis-specified if both scholars are correct.

The estimation problem may also be a statistical consequence of a deeper one, which we label the *dimensionality problem*. Social scientists have come to treat a growing number of scales as distinct constructs (for variety, we use "construct" here interchangeably with "trait," "factor," and "dimension"). But different scales may measure the same underlying construct, or mixes of constructs.[6] As previously noted, scholars generally regard high internal consistency and some test of predictive validity as sufficient for demonstrating the merit of a new scale. However, high internal consistency within a scale and predictive power are not evidence of discrimination across scales. Tests of discriminant validity show that scale items are more correlated with one another than with items from another other scale but do not address the question of whether multiple scales are measuring some underlying common construct(s). And, as we note above, when tests of discriminant validity are conducted, they are rarely applied to a large number of scales (and their items), limiting the scope of these tests.

It may be that if one were to examine a broad range of scales and their component items, many of them measure a much smaller number of latent constructs. Thus, despite

[6] Note that different scales might measure the same underlying factor either because the same factor is captured by the questions in both scale or due to measurement issues (in which how respondents interpret and answer survey questions is affected by some common underlying factor, independent of the trait the scales seek to measure).

apparent discriminant validity, different scales may (in part) measure common dimensions or different mixes of multiple such dimensions. This conceptual redundancy in which multiple scales measure a smaller number of common factors would also produce an estimation problem: If multiple scales measure overlapping dimensions, then single-scale regression models will be interpreted as demonstrating a scale's predictive validity, even if scales are redundant.

In this paper, we investigate these issues empirically in four ways. First, we document that although scales are widely employed in leading political science journals, researchers usually do not account for multiple scales simultaneously or test assumptions about their independence, meaning existing work seldom addresses either the estimation or dimensionality problems introduced above. Second, we administer 191 unique survey items that are used to construct 38 commonly used scales to 2,001 respondents. Analyzing these data, we document substantial correlations across many nominally distinct scales and demonstrate using simulations how such correlations can generate false positive results (i.e., scales can appear statistically significant even when effects originate in correlated scales).[7] This highlights the estimation problem in that single-scale regressions can produce statistically significant but substantively spurious findings (which would also pose a problem for tests of predictive validity).

That finding that many scales are correlated raises the concern that they in fact measure common underlying concepts (i.e., the dimensionality problem). For this reason, our third

[7] As we show below, this concern about bias is not purely hypothetical. If we use individual scales to predict 2024 vote choice, we find that estimates in models without other scales produce estimates that are dramatically larger than in models that account for other scales.

analysis examines, at the item-by-item level, whether high inter-scale correlations reflect a smaller latent structure than current practice suggests. We find that items from different scales frequently load onto a much smaller number of dimensions, with the first dimension broadly corresponding to opposition to improving the status of non-majority groups, suggesting conceptual redundancy across the measurement landscape. It appears that across scales the underlying latent construct space is smaller than hypothesized, or at the very least, that novel scales often measure both existing factors and novel traits.

Fourth, to illustrate the implications of this redundancy, from our 191 survey-item pool,[8] we construct six atheoretical scales using only internal correlation as a criterion (i.e., we identify the six items that load most highly onto each of six novel dimensions identified using exploratory factor analysis). Only one such scale largely recovers an existing scale (need for cognition), while the others draw from multiple existing scales. Accounting for just the first two of these novel scales (each constructed as a mean scale from only six items each) dramatically reduces the apparent association between extant scales and a key political outcome similar to those they are often used to predict in published work, 2024 presidential vote choice. Additionally, our factor analysis reveals that several commonly used scales either load heavily onto this common first factor, a second factor, or a mix of these factors. Overall, these results confirm both the estimation and dimensionality concerns, demonstrating that what researchers tend to treat as distinct constructs often represent different combinations of shared underlying factors.

[8] For these analyses, we collapse the men and women stereotype endorsement scales into net anti-women stereotype endorsement relative to men.

In the conclusion we discuss why these findings have important implications for how scholars engage in theory testing and suggested changes in research practices. Additionally, we also highlight the broader implications of these findings for how scholars engage in theory development.

Demonstrating Construct Validity

Our analysis builds on prior work across disciplines that proposes standards for assessing whether a scale is valid. We briefly review this work, focusing on whether these tests would (1) properly distinguish empirical effects of valid scales from effects of correlated but theoretically distinct scales and (2) detect that different scales measure common constructs. If scales do measure distinct constructs, we argue that extant tests of predictive validity do not generally consider the possibility that effects could originate in other correlated scales, which would render those tests less informative because of the estimation problem discussed previously. Tests of discriminant validity, which seek to measure whether scales measure unique or common constructs, are direct tests of scale independence, addressing the dimensionality problem noted above. But, in practice, these tests may provide limited information given how they are implemented in prior political science work for three reasons. First, these tests are binary tests of two scales, so they do not assess whether multiple scales that might appear distinct in binary tests nonetheless measure, at least in part, one or more shared traits. Second, these tests are generally applied only to comparisons of a new scale to a few other scales (and their component survey items), so we lack information about scales that are excluded from these tests. Third, rarely do scholars assess whether specific scales add predictive power compared to multiple other scales or to common constructs measured across multiple scales, so

we lack informative evidence about the implications of either the estimation or dimensionality problem.

The psychometric literature primarily addresses construct validity in three ways: internal consistency, discriminant validity, and predictive validity. Scale construction often begins with a researcher selecting items they believe measure a common trait, either through direct inspection of the items as constructed or by reference to another external source of validity (e.g., in the case of the Big 5 personality traits, by the groupings of words used to describe specific personality types). This intuition is generally affirmed by assessing the correlation among these items. Cronbach (1951) introduced alpha, a measure of internal consistency of scales. Cronbach's alpha compares the average covariance among scale items to the total variance of the scale; a higher alpha indicates greater internal cohesion. This measure has enjoyed widespread use for decades and remains among the key practices in political science and psychometrics more broadly for establishing construct validity (e.g., Barker et al. 2021). In practice, almost all work in political science introducing a new scale provides evidence about internal consistency.

Tests of discriminant validity were developed soon after. Campbell and Fiske (1959) established the concept of discriminant validity by outlining a set of patterns that should appear in correlations across traits and measures. Their key argument for our purposes is that correlations between the same trait across testing methods should be high ("convergent validity") whereas correlations between different traits should be lower ("discriminant validity"). In the survey context, the logic can be expressed as the argument that discriminant validity is demonstrated when (a) the correlations between items used to construct a single

scale are larger than the correlations between items in different scales and (b) the correlations between different scales are small.

Different techniques have been proposed to formalize tests of discriminant validity, but they share a common intuition in comparing the relative correlations of items used to construct a single scale to the correlations of items across scales. Jöreskog (1969) developed computational methods for confirmatory factor analysis, which allows researchers to test whether observed variables load onto hypothesized latent factors as theorized (testing whether items within a scale load primarily onto a single latent factor while items from different scales load onto distinct factors), providing a statistical framework for evaluating the validity patterns outlined by Campbell and Fiske. Building on earlier work by Fornell and Larcker (1981), which proposes the average variance extracted (AVE) criterion, Henseler et al. (2015) introduce the heterotrait-monotrait (HTMT) ratio as a criterion for assessing discriminant validity between individual pairs of scales. The HTMT ratio compares the average correlation between items across two different scales (heterotrait) to the geometric mean of the average of item correlations within each scale (monotrait). A low HTMT ratio indicates discriminant validity. Applied to the survey context, discriminant validity exists when the items used to construct a scale are (relatively) more correlated with one another than with items in other scales.

In the political science context, we are not aware of any work using the scales we study that conduct confirmatory factor analysis or HTMT tests to assess discriminant validity compared to other commonly used scales. Some work has examined whether attitudes about some policy issues are multidimensional (see, for example, the interchange between Goren 2008 and Jacoby 2008 on the dimensionality of spending attitudes), and other work has

examined whether different traits explain attitudes (e.g., Goren et al. 2016), but systematic analysis of the discriminant validity of most scales appears rare.

Additionally, a key limitation of any binary comparison of two scales (e.g., an HTMT test) to assess their independence is that survey items in multiple scales could measure one or more common traits, so that even if within-scale item correlations are larger than across-scale item correlations, there could still be a (large) common component to multiple measured scales. This implies it would be useful to assess for a larger set of the survey items used to construct different scales if there are one or more shared dimensions that are being measured across multiple scales. It may be that different scales measure common traits in addition to unique dimensions, but without a systematic analysis of multiple scales simultaneously, it is not possible to discern if multiple scales are measuring some common trait.

Finally, predictive validity concerns whether a scale successfully correlates with outcomes that theory links to the underlying construct. In the classical formulation by Cronbach and Meehl (1955), a scale exhibits predictive validity if it correlates with theoretically relevant outcomes alongside null effects for unrelated outcomes. In most political science work, predictive validity is much less systematically addressed than discriminant validity, and it is usually tested in a single-scale model and by showing only that a scale predicts relevant outcomes (i.e., "positive" predictive validity),[9] but not that it does not predict outcomes it should not (i.e., "negative" predictive validity, which for space reasons we do not take up in greater detail). In practice, rarely do researchers test whether a construct predicts an outcome

[9] Such models often control for other demographic (e.g., age, education, etc.) characteristics and some other political predispositions (e.g., partisanship) but usually do not also include multiple other scales.

uniquely relative to other scales, nor are we aware of studies that test whether a new scale adds predictive value relative to underlying constructs that may be measured across multiple scales.

Overall, prior work on scale validity is only partially informative of the two problems we identify—estimation and dimensionality. In terms of estimation, a scale with high internal consistency and clear discriminant validity may still produce a biased estimate of the effect of the scale on an outcome if a test of predictive validity is done in isolation, without considering the independent effects of other correlated but omitted (and theoretically motivated) scales. From a social science theory-testing perspective, this bias also means that empirical estimates of predictive validity may not be informative tests of theoretical claims because they may be biased and therefore sensitive to false positives.

In terms of dimensionality, the extant literature largely concludes that establishing discriminant validity by showing scale items are more correlated with one another than with items from another scale items (or simply that the items in one scale load onto a common factor) is sufficient to demonstrate that a new scale captures a distinct construct. As implemented, however, this approach rarely considers a large set of potentially related scales (and their component items), and so it can only speak to distinctiveness among the scales to which it is applied. We are also unaware of work that has considered the possibility that different scales measure different "mixes" of a smaller number of underlying dimensions. In terms of statistical practice, confirmatory factor analysis is not compared to a more naïve approach in which the items used to construct multiple scales are analyzed in an exploratory manner to understand which items correlate with one another, either within or across scales. Such exploratory factor analysis would provide informative evidence to validate key researcher

assumptions about scale construction, which is that scale items measure the trait they are intended to capture and not other theoretically relevant constructs. Additionally, such a factor analysis would allow a researcher to assess empirically what is distinctive in each extant scale and set up an informative test of predictive validity: Does a scale offer any additional explanatory power beyond the effects of some mix of latent traits common across multiple scales? In the most general terms, along the lines of Campbell and Fiske (1959), we argue that if a scale does not offer what we call *discriminant predictive validity*, meaning predictive power beyond the effect of other scales and factors common to multiple scales, a researcher has not demonstrated a proposed new trait is novel rather than a different measure of some previously identified trait(s).

Scale Selection

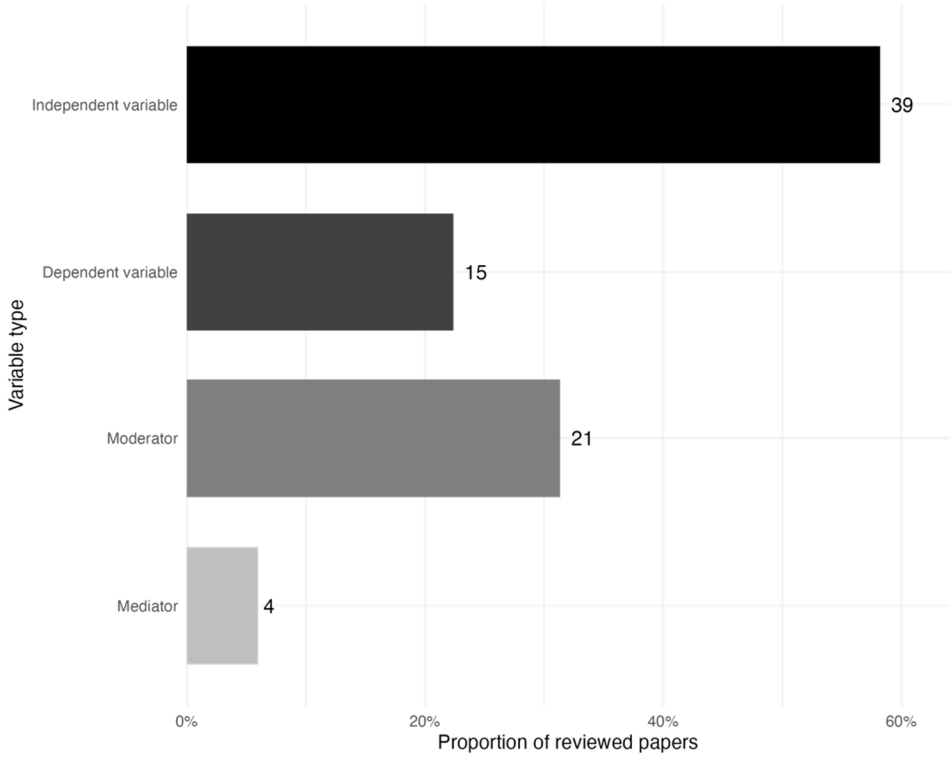
To focus our analysis, we first identified 38 of the most frequently used social-psychological and attitudinal scales in recent political science research by reviewing issues of the *American Political Science Review*, the *American Journal of Political Science*, the *Journal of Politics*, and *Political Psychology* for the five-year period between 2020 and 2024. In addition to this list, we included common scales either asked on the American National Election Studies or otherwise well-known to researchers. The list in our study is not exhaustive and does not fully represent the proliferation of psychological scale usage in political science. Given survey fielding constraints, we omitted scales that require large amounts of survey time to field, and we used shortened or adapted (e.g., made applicable to a U.S. context) batteries where available.[10] We

[10] See Ansolabehere et al. (2008) and Bakker and Lelkes (2018) for a discussion about the effects of using shortened survey scales.

present the full list of scales, with a brief description of each and their original sources, in Appendix Table A2.

In our literature sweep we identified 67 papers using at least one of the 38 psychological scales we specified. For each paper, we coded whether a scale was used as an independent, dependent, moderating, or mediating variable. Figure 1 summarizes this usage. Note that some papers use scales as more than one type of variable (e.g., a study may use a given scale as an explanatory variable and as a moderating variable in an interaction term), so the proportions sum to more than one.

Figure 1. Use of psychometric scales in identified papers.

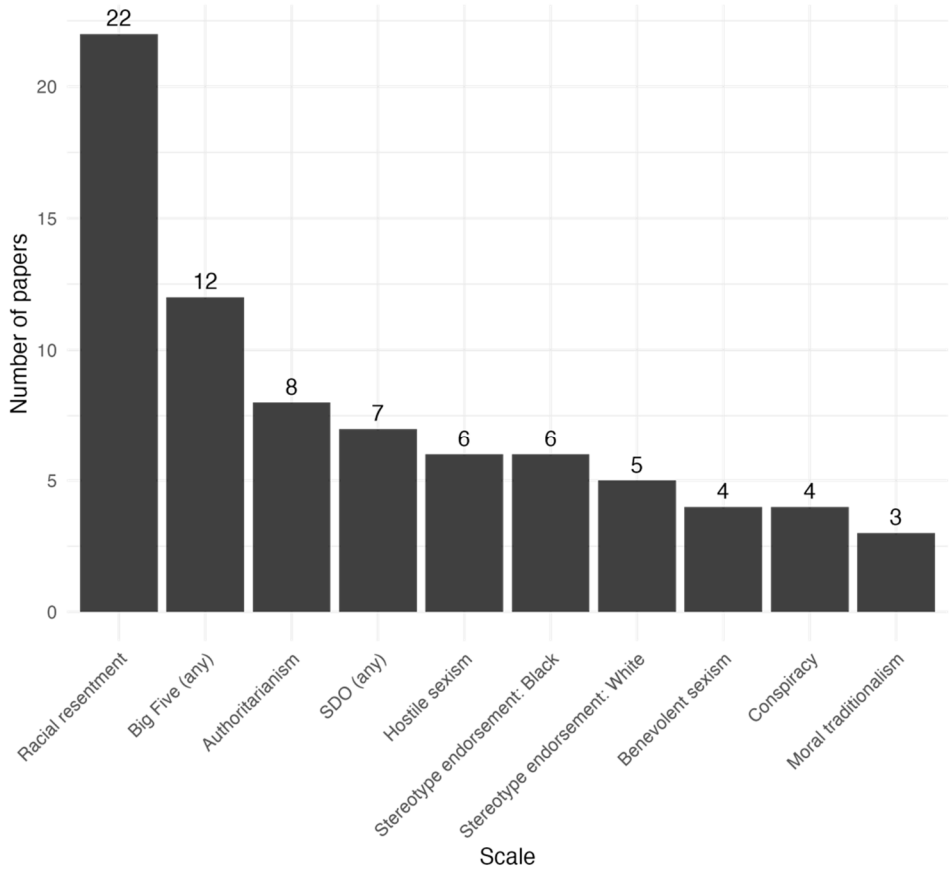


Notes: Sample is all papers identified in literature sweep that use one of the 38 scales we study (see text for details).

Our review finds that psychometric scales are used as independent and moderating variables most frequently (roughly 90% of the time). Our analysis will therefore focus on these

uses. Approximately 60% of all reviewed papers use a scale as an independent variable, and 30% use a scale as a moderating variable. Scales are less frequently used as dependent variables (approximately 20% of papers) and rarely used for mediation analysis (under 10% of papers). In Figure 2, we plot the frequency of use of the most commonly used scales from our review, again restricted to the papers we identified that use a scale in the four ways specified in Figure 1.

Figure 2. Most commonly used scales in identified papers.



Notes: Plot shows all scales used more than twice as one of the four key variable types in papers identified in our literature sweep.

The racial resentment scale was used most often, appearing in 22 papers from our review. One or more components of the Big Five was used in 12 papers, authoritarianism was used in eight papers, social dominance orientation (SDO) was used in seven papers, hostile

sexism scale was used in six papers, and stereotype endorsements for Black and White individuals were used in six and five papers, respectively. The remaining scales in our list were used in no more than five papers across our review.

We next turn to whether analysis using these scales also accounts for other scales. As we note above, if scales are correlated, omitting a correlated scale risks a false positive if other scales also affect the relevant outcome. In Appendix Table A3, we list the papers from our systematic review that use a psychometric scale either as an independent or moderating variable. We report the specific scale(s), its (their) specific use(s), and the number of other psychometric scales that are controlled for in at least some analysis in the paper (i.e., if a scale is used as an independent variable, is there at least one model estimated in the paper, including robustness models and results reported only in an appendix, that controls for one or more other scales?). We include in this count scales beyond our list of 38, as well batteries of the same name that may differ in item content from our list. We tried to be inclusive in our assessment of whether a paper accounted for other scales to bias our estimates in favor of more conservative researcher practices.

We estimate that approximately 41% of papers using a psychometric scale as an independent or moderating variable do not control for at least one other scale in at least some analysis. In the remaining 59% of papers that control for at least one other scale, the average number of other factors accounted for is 3.9, a figure which is influenced upward by a single paper that accounts for 29 other scales. The median is 2 and the mode is 2.

Overall, this review reveals the breadth of the usage of psychological scales in political science. Scales are used mostly to predict outcomes and somewhat frequently as moderating

variables (as sources of heterogeneity in effects). Importantly, when scales are used in these ways, analysis only accounts for the effects of any other scales about 60% of the time, and when it does so, the median practice is to account for only 2 other scales. If scales were independent of one another—either in terms of the constructs they measure or their underlying measurement properties—or if other scales have no effect on the outcome of interest, this pattern might be inconsequential. But are scales correlated or distinct from one another? After introducing our dataset, we take up this question first.

Data

We collected these data as part of the [removed for peer review], which was approved by the [removed for peer review] Institutional Review Board. Respondents provided informed consent.[11] This study, administered by YouGov, included a group of 2,001 respondents who were interviewed specifically to measure the psychometric scales used in this paper. The survey was fielded February 2–15, 2024. Given the length of the survey, it was split into two parts that each took approximately 20 minutes to complete. To protect against acquiescence bias, many scale items were reverse-coded. Further details appear in the Appendix A1.

This sample was drawn from a broader election survey that included a baseline sample of approximately 130,000 respondents in the United States recruited from YouGov's online opt-in panel. When using weights provided by YouGov, the sample is reflective of the American

[11] Consistent with the *Principles and Guidance for Human Subject Research*, respondents were informed of the purpose of the study and provided information regarding data privacy and protection (i.e., anonymity). The study did not engage in deception or intervene in political processes.

public on observable demographics. Our primary analysis is unweighted, but weighted results are available upon request.

For each scale (listed in Appendix Table A2), we used the coding rule described in the source article to create an additive scale from the component items, reverse coding them as appropriate. We normalize each scale to have a mean of 0 and a standard deviation of 1, which facilitates comparisons of magnitudes across scales.

Our survey included 191 unique items used to construct the 38 scales we analyze here. Missing data is therefore a concern because if we employ listwise deletion we are left with 1,520 of our original 2,001 respondents. Fortunately, missing data are relatively rare in our sample: 76% of respondents skipped no items, 16% skipped one, 99% skipped 4 or fewer, and the maximum number of items skipped is 7 (3.6% of items asked). Overall, we are missing 800 item responses across 2,001 respondents, a missing data rate of 0.2% (800/(191*2001)). To retain the largest possible sample for analysis, we impute missing values for all items using mean imputation. This is a conservative approach because unlike using other scale questions or other (non-scale) survey responses to predict missing answers, it should not artificially increase the correlation among items or with the outcomes we use the scales to predict.[12] Our core analysis sample is therefore composed of 2,001 respondents.

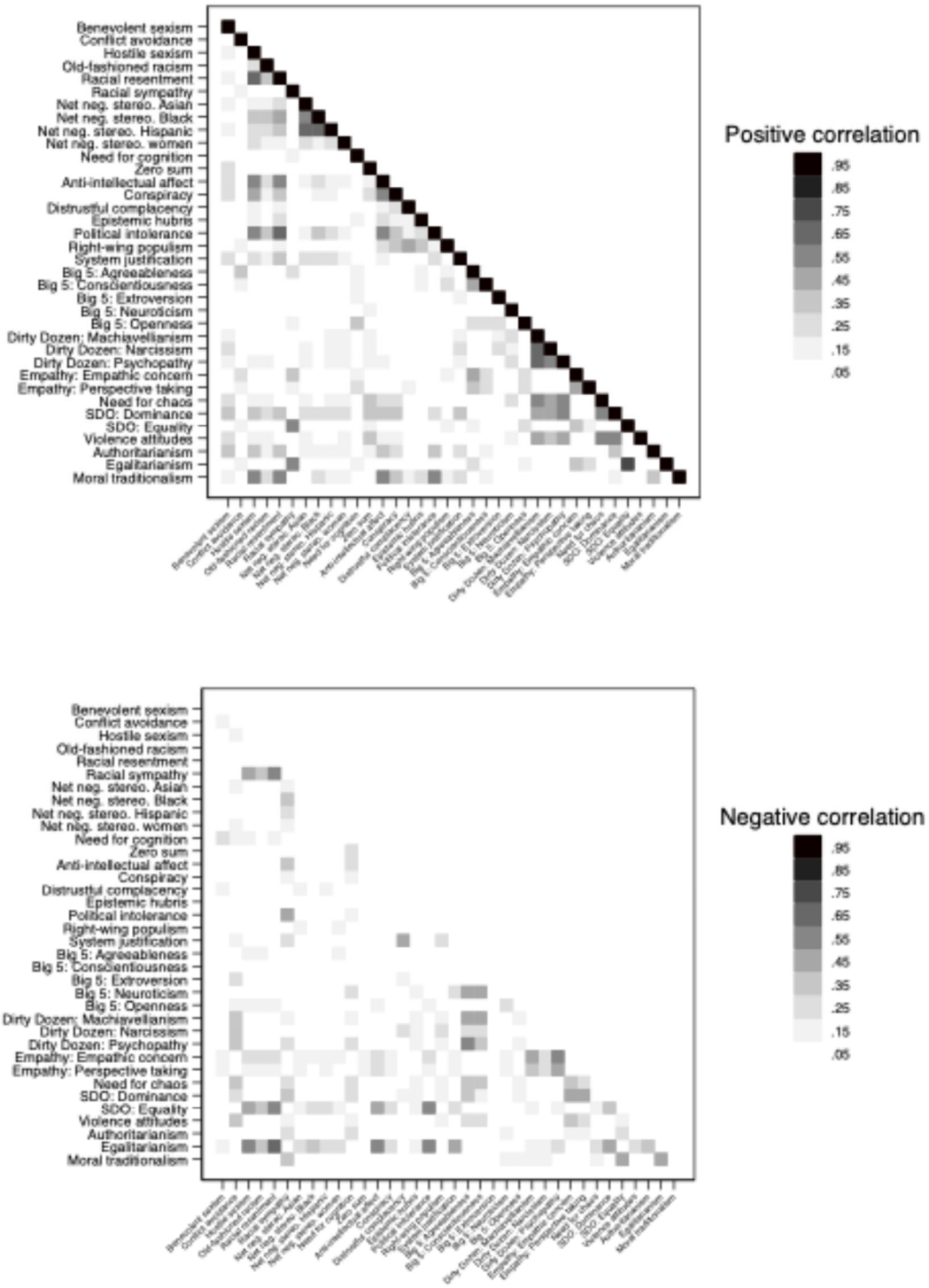
Commonly used scales are correlated

We present the correlations among the scales we study graphically in Figure 3. In this section we use net stereotype endorsements because this is how these scales are generally

[12] We also conducted parallel analysis (available upon request) restricted to the 1520 respondents who answered all items and find similar results.

used in the political science literature, shrinking our scale count to 36. For women, this is net stereotypes relative to men. For Black, Hispanic, and Asian people, this is net negative stereotypes relative to White people. The top facet shows positive correlations and the bottom facet shows negative correlations, with darker shadings indicate larger correlations among two scales. The figure is organized so that scales that are broadly similar thematically are grouped together into four bundles: (1) intergroup attitudes, (2) beliefs about elites and political trust, (3) Big Five personality traits, and (4) traditional values, negative personality traits, and interpersonal behavior.

Figure 3. Positive (top) and negative (bottom) correlations among scales for 36 commonly used scales.



Notes: Darker shadings indicate larger correlations (see text).

There are some scales that are relatively uncorrelated with other scales. For example, the extraversion dimension of the Big Five and net anti-women stereotype endorsement are not highly correlated, positively or negatively, with any other scale (absolute value of all correlations <0.3). But many commonly used scales are correlated in recognizable "families." At the extreme, consider the dominance dimension of social dominance orientation (SDO), shown in the sixth-from-last row and sixth-from-last column. It is positively correlated at greater than 0.30 with 13 other scales: benevolent sexism, hostile sexism, racial resentment, zero-sum mentality, anti-intellectual affect, conspiracy, system justification, all three of the "Dirty Dozen" measures, need for chaos, violence attitudes, and authoritarianism. It is negatively correlated at less than -0.30 with 6 scales: racial sympathy, the agreeableness dimension of Big Five, the empathic concern dimension of empathy, the perspective taking dimension of empathy, the equality dimension of social dominance orientation, and egalitarianism. Overall, the social dominance dimension is highly correlated (absolute value of correlation >0.30) with 19 of the 35 other scales (54%).

Between these extremes, there is substantial variability, but almost all scales are at least somewhat correlated with other scales.[13] Only two scales have no correlations (absolute value) greater than 0.3, and the median is 6.5 with an average of 6.7. In other words, most scales are

[13] Some work argues that constructs such as right-wing authoritarianism and social dominance orientation are foundational dimensions influencing sociopolitical attitudes, implicitly suggesting that other constructs are downstream from them, and thus high correlation is expected (e.g., Duckitt and Sibley 2009). However, in practice, researchers introducing new scales rarely theorize their constructs as downstream from higher-order ones. We note that causal hierarchies still produce the same estimation and dimensionality problems we identify, but our critique concerns redundancy in scales at the same theoretical level. For example, if one partitions a Big 5 trait (e.g., openness) into its facets, but the facets are in practice highly correlated, estimating the effect of a facet and theorizing about its effect in isolation risks both the estimation and dimensionality problems.

meaningfully correlated with at least one other scale, and on average scales are correlated in non-trivial ways with about 20% of the other scales in our sample.

To underscore this intuition, we also conducted an exploratory factor analysis for the 36 scales listed in Figure 3.[14] The first dimension of this factor analysis, which has an eigenvalue of 7.2 and explains 43% of the variance, generates factor loadings whose absolute value is greater than 0.3 for 26 of the 36 scales (72%). The second dimension has an eigenvalue of 4.2 and explains an additional 25% of the variance. It has factor loadings whose absolute value exceeds 0.3 for 15 of the scales (42%). The unrotated factor loadings for the first five dimensions, which cumulatively explain 97% of the observed variance, are reported in Appendix Table A4 and the scree plot for the factor analysis is shown in Appendix Figure A1 This means that although researchers have constructed 36 unique scales, 2 latent dimensions appear to account for more than two-thirds—67%—of the measured variance across these scales.

High Correlations Between Scales Increase Risk of Estimation Problem

The observed high correlations between scales suggests the estimation problem we identified earlier may be more than a theoretical problem. In particular, as we describe in the prior section, published papers that use these scales as predictive or moderating variables do not frequently account for other correlated scales that might also affect outcomes. To assess whether this omission might lead to the concerns raised by the estimation problem, we conduct a thought exercise: Suppose that the effect of a researcher's chosen scale of interest on an outcome was 0, but the effect of another scale on the same outcome was meaningful. How

[14] We conducted principal factor analysis in Stata using the factor command.

frequently would a researcher who only examined the correlation between their scale of interest and the same outcome incorrectly conclude that the scale they measured predicted the outcome, even if it had no effect on the outcome? While such a thought exercise is stark, it is broadly applicable to other scenarios as well. For example, suppose two positively correlated scales each had a positive effect on an outcome, but a researcher only estimated the effect of their scale of interest. One would be similarly concerned that their estimate could be biased upward by proxying the additional effect of the other (omitted scale). (And, of course, the direction of bias could run in the opposite direction if the omitted correlated scale had a negative effect on the outcome or the scales were negatively correlated.)

To answer this question empirically we conducted a simulation exercise. For each of these 36 scales, we simulated in our dataset of *N*=2,001 a scenario in which the chosen scale had a linear effect of 0.15 on a generated outcome Y_i, where the outcome was determined stochastically as $0.15 \cdot scale_i + \varepsilon_i$, where $\varepsilon_i \sim \mathcal{N}(0,1)$ (i.e., drawn from a normal distribution with mean 0 and standard deviation 1). Note these simulations assume a particular error variance ($\sigma^2 = 1$). If the error variance were larger, false positive rates would be lower, whereas if it were smaller, false positive rates would be higher. We chose 0.15 and 1 because with these parameters, in the true regression (simulated outcome predicted by the scale that caused the outcome), the estimated scale effect is always significant. For each scale, we then ran 35 regressions, one for every other scale, where we estimated the effect of that scale on the outcome and recorded both the estimated regression coefficient and its standard error.

We plot the results of this simulation exercise in Figure 4. In each facet, we plot the estimated regression estimate for 35 distinct regressions. Each point is the estimated coefficient

(horizontal axis) and absolute value of the t-statistic (vertical axis) for the effect of the chosen scale on the outcome when the outcome is simulated to in fact be caused by another scale. Points in black are statistically significant (|t| > 1.96), while those in gray are not. In the label for each facet, we also list how many of the 35 estimates are statistically significant. Of the 1,260 points plotted in the figure, 329 (26%) are statistically significant.

The figure shows that false positives would be likely even in the absence of true effects. At one extreme, the upper rightmost facet displays results for the dominance dimension of SDO. When the true effect on the outcome is caused by 19 of the other 35 scales the simulation nonetheless returns a statistically significant bivariate estimate, a false positive rate of 54%. At the other extreme, for the extroversion dimension of the Big Five (third from bottom row, third from right), only 2 of 35 estimates are significant, a false positive rate of only 6%. Across all scales, 9.1 estimates (26%) are significant on average (the median is 9).

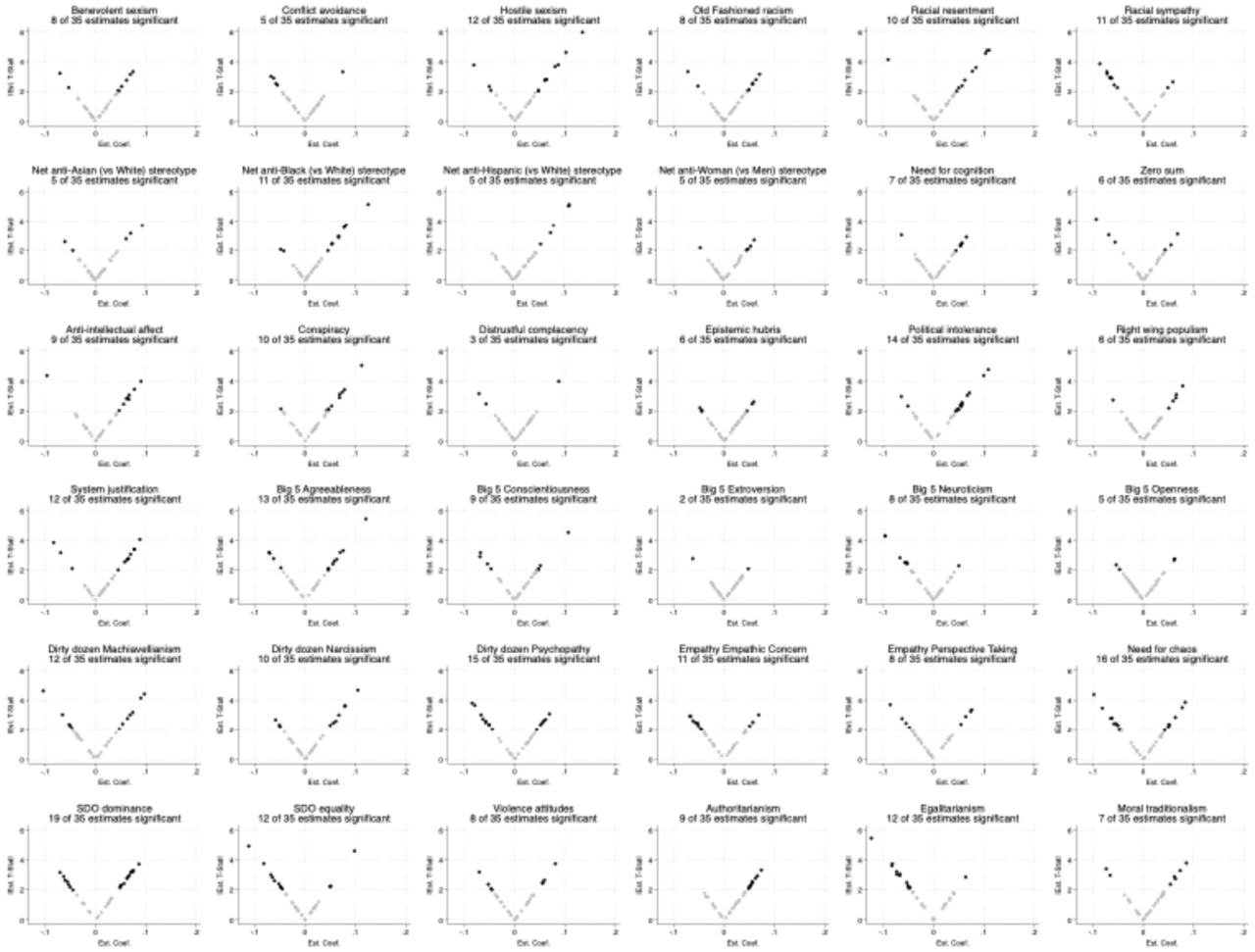
Across scales, the magnitude of this false positive effect is directly proportional to the correlation between the scales,[15] so it is follows that the false positive rate is highest for those scales that are highly correlated (negative or positively) with other scales (See Figure 3). Additionally, while our chosen parameters (true β=0.15; standard deviation of error=1) affect the absolute frequency of the false positive rate, the relative frequency of false positives across

[15] When two scales (or more generally, two variables) are correlated, a scale that does not cause the outcome can still appear statistically significant because it inherits variance from the true causal scale. If X_1 causes Y with coefficient β, and another scale (we will call X_2) is correlated with X_1 at ρ, then the estimated coefficient from regressing Y on X_2 alone is approximately $\hat{\beta}_2 = \beta\rho$, where the corresponding t-statistic is approximately $\frac{\beta\rho\sqrt{n}}{\sigma}$. Thus the false positive rate increases monotonically with the absolute correlation $|\rho|$, the true effect size $|\beta|$, and the sample size n, and decreases with higher error variance σ^2.

scales is unaffected by these choices (those scales that are more correlated with other scales are more likely to yield false positives).

These results are particularly consequential for the most widely used scales identified in our literature review. Returning to Figure 2, note that of the most frequently used scales (used at least five times), many also have the highest false positive rates. Racial resentment (used in 22 articles) has a false positive rate of 29%, the two dimensions of social dominance orientation (9 articles) have an average false positive rate of 44%, for authoritarianism (8 articles) it is 26%, for hostile sexism (6 articles) it is 34%, and for net anti-Black stereotypes (6 articles) it is 31%. The average false positives for the Big Five (14 articles) is a substantively lower 21%, although it is notable that one component of the Big Five, agreeableness, has a false positive rate of 37% (in part because it is highly negatively correlated with need for chaos, social dominance orientation, violence attitudes, and all three components of the Dirty Dozen, among other scales). Comparing Figure 3 with Appendix Table A3 also makes clear that empirical models that test for the effects of these scales rarely account for the other scales that they are most correlated with, meaning that they do not robustly address the estimation problem.

Figure 4. Estimated effects of scale on outcome when true causal effect is from each of 35 other scales.



Notes: Each simulated scale effect on outcome is 0.15 with an error SD of 1.00. Each point is estimated coefficient and |t-statistic| for binary regression with robust standard errors of each scale (facet) on outcome caused by each other scale. All scales standardized to mean 0 and standard deviation 1. Points in black are statistically significant (p<0.05, n=2,001) and points in gray are not.

The above analysis uses simulations, but we have also conducted analyses using observed outcomes to give a sense of the potential magnitude of the estimation problem in actual data. Specifically, we compare the estimated effect of each scale in predicting a salient political outcome, vote choice in the 2024 presidential election, to the estimate recovered from a model that simultaneously includes all 36 of these scales at once. This finding is reported in Appendix Figure A2 and shows that the magnitude of estimated effects, and the prevalence of statistically significant estimates, is greatly reduced in analysis that accounts for the effects of other scales. In a binary regression framework, 30 of 36 scales (83%) are estimated to have a statistically significant effect on vote choice, but in the analysis controlling for all scales, only 7 of these estimates remain significant (23%) and these estimates are attenuated on average by 69% in the controlled models. (By contrast, three estimates are significant only when controlling for all other scales.) While this test is conservative, it also suggests that prior empirical analysis is likely substantially underpowered to distinguish effects of a scale from the effects of correlated scales because one would generally need larger sample sizes than are present in many survey datasets to distinguish effects of correlated factors.

Item-level Analysis Supports Dimensionality Concern

Our preceding analysis examined correlations across scales at the scale level, finding that many scales are correlated with other scales. This raises the concern that they in fact measure common underlying concepts (or at least that they have correlated measurement properties), the core of the dimensionality problem raised above. But these scales are created using multiple items, and so a natural follow-on question is to assess why it is that different scales are correlated. Is it that individual items in different scales are highly correlated, or alternatively

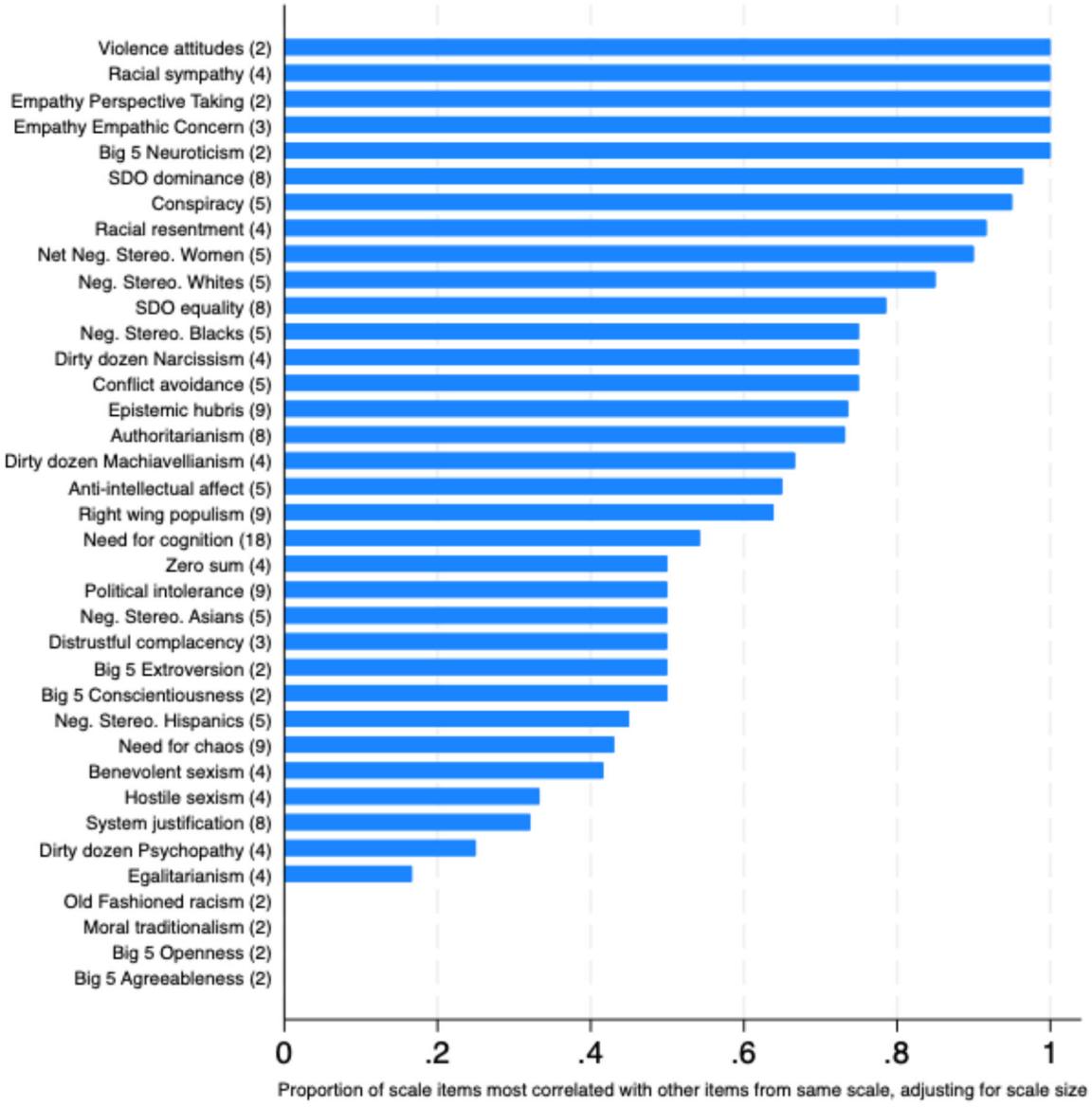
that scale items are generally most correlated with other items in the same scale, but the cumulative scales are nonetheless correlated? A finding that scale items are correlated with items from other scales would suggest that, as constructed, different scales capture one or more common factors, despite using distinct questions in their creation. By contrast, if different scales are highly correlated despite their component items being most correlated with items from the same scale, it suggests individual scales measure both unique traits and some shared traits.

To answer this question empirically, we calculated the correlation between each pair of individual survey items across all the 191 survey items we fielded to construct the scales we analyze. (In this analysis we move from using net racial stereotypes [e.g., Black minus White] to raw racial stereotypes, increasing our scale count to 37. Results using net racial stereotypes are similar but require reusing scale items for beliefs about White people to construct net differences on each measure for non-White groups, mechanically increasing the correlations between those scales.) Each scale is composed of multiple items, ranging from 2 to 18, with a median of 4 and an average of 5.03. For a given scale created from N items, we calculated how frequently the N-1 most correlated items (absolute value of correlations) are from the same scale. For example, suppose a scale had 5 items. For each of the 5 items in the scale, we calculate the proportion of the time the 4 other most correlated items are from the same scale (since there are at most N-1 other items it can correlate most highly within the same scale). If scales are entirely "self-contained," they should correlate most highly only with items from the same scale. We then calculate the average, by scale, of this number across all scale items. This

analysis normalizes the numerator by the maximum possible number of items from the same scale (*N*-1), so results are not mechanically affected by the number of items in a scale.

This is likely a conservative test of dimensionality for two reasons. First, there are differences in survey item design across scales, so that if common modes of asking questions (e.g., agree-disagree scales with a fixed number of response options) tend to produce common answers, these design features will inflate within-scale correlations. Second, scales are often initially constructed by fielding multiple items and choosing those that are most correlated, so researcher design choices will tend to favor items with high internal correlations (this is of course usually done on a theoretically motivated basis to pick the best items to measure a latent trait).

Figure 5. Frequency with which scale items are most correlated with other items from the same scale.



Notes: Vertical axis lists scales with their number of component items in parentheses. The figure shows the average proportion of time the N-1 most correlated items for each scale item are from the same scale, where N is the number of survey items in the scale. We calculate the average of this quantity for all N items in each scale. If all scale items correlated most highly with items in the same scale, all proportions would be 1. Scales sorted by decreasing proportions.

Figure 5 summarizes the results of this exercise at the scale level, sorted in decreasing order of the average proportion of scale items most correlated with other scale items. The pattern of results varies substantially across scales. For five scales all the most correlated items are from the same scale (violence attitudes, racial sympathy, empathy [both perspective taking and concern], and the neuroticism dimension of the Big Five), while for four scales none of them are (old-fashioned racism, moral traditionalism, and both the openness and agreeableness dimensions of the Big Five). Averaging the results across all scales (i.e., taking the average of the bars shown in the figure), 59% of scale items are most correlated with other scale items. For some of the most widely used scales (e.g., racial resentment, social dominance orientation) these figures are relatively high (>70%), while for others they are much lower (less than 50% on average for the Big Five, both benevolent and hostile sexism, and moral traditionalism).

At the same time, that scale items are most correlated with other scale items does not fully address the concern that they might nonetheless be highly correlated with other items from different scales because they measure a similar construct, even if the magnitude of the correlation is smaller. To understand whether this concern is warranted, we conducted an exploratory factor analysis where we jointly scaled all 191 of these individual survey items. The resulting scree plot is displayed in Appendix Figure A3 and displays the characteristic “elbow” after six factors, meaning there are six factors that explain larger proportions of the common variance (from 25% for factor 1 and 16% for factor 2 to 5% for factor 6; total variance explained by the first 6 factors is 68%), while each additional factor after that explains a much smaller portion of the cumulative variance (less than 3% for factors 7-9, less than 2% each for factors 10-20, and less than 1% each for all other factors).

Which items load most heavily on to each factor? This approach is motivated by another thought exercise: Suppose a naïve researcher was building scales from these items and decided that they were going to construct six unique scales (shown in the scree plot). Which items would they end up choosing for each scale? We pick an arbitrary number of six items as a reasonable size for a survey scale and use these data to identify the items that load most heavily onto each of the six dimensions. These items, along with their factor score and the original scale from which they are drawn, are shown in Table 1.[16] We report unrotated factor loadings below but include orthogonally rotated scores in the appendix (see Table A5).[17] (Complete scale loadings for the first 6 dimensions, with scale loadings below 0.3 [absolute value] suppressed, are displayed in Appendix Table A6.) For each factor, we also report the Cronbach's alpha, which we note is high in all cases and above common thresholds for deeming a scale reliable. (This is not surprising because high factor scores correlate with high inter-item variance.) These six new scales are only modestly correlated with one another (absolute correlations <0.30) with one exception: Factors 3 and 6 are negatively correlated at -0.78.

[16] Alternative factoring methods yield virtually identical results. Maximum likelihood factor scores correlate at 0.98 with our reported scores. Iterated principal factor scores correlate at 0.99. And principal component factor scores, which make more restrictive assumptions than principal factor analysis, correlate at 0.99.

[17] As we show below, in a model including both factors 1 and 2, each predicts key political outcomes despite their correlation. Rotation limits this predictive power by separating genuinely shared variance. Rotation reduces the correlation between extracted factors, but it cannot remove the shared variance from the underlying data. Indeed shared variance across scales is this paper's central point: Researchers using individual scales in practice have no ability to "rotate away" common structure, so reporting unrotated solutions better reflects the way scales are used in practice.

Table 1. Highest loading items for six factors (scales) from item-level factor analysis.

Question wording	Source scale	Factor score
Factor 1 ($\alpha = 0.88$)		
Inferior groups should stay in their place.	Social dominance orientation: Dominance	0.691
If certain groups of people stayed in their place, we would have fewer problems.	Social dominance orientation: Dominance	0.691
It's probably a good thing that certain groups are at the top and other groups are at the bottom.	Social dominance orientation: Dominance	0.684
Sometimes other groups must be kept in their place.	Social dominance orientation: Dominance	0.673
It's really a matter of some people not trying hard enough; if Blacks would only try harder they could be just as well off as Whites.	Racial resentment	0.646
This country would be better off if we worried less about how equal people are. [Reverse-coded]	Egalitarianism	-0.645
Factor 2 ($\alpha = 0.83$)		
Foreigners in the United States should have the same access to social welfare benefits.	Political intolerance	0.649
Foreigners in the United States should be allowed to vote at the local level.	Political intolerance	0.644
Foreigners in the United States should be allowed to vote at the federal level.	Political intolerance	0.626
Generations of slavery and discrimination have created conditions that make it difficult for Blacks to work their way out of the lower class.	Racial resentment	0.584
Over the past few years, Blacks have gotten less than they deserve.	Racial resentment	0.567
I tend to expect special favors from others.	Dirty Dozen: Narcissism	-0.548
Factor 3 ($\alpha = 0.76$)		
Whites – Intelligent to Unintelligent	Stereotype endorsement: White	-0.490
Blacks/African Americans – Intelligent to Unintelligent	Stereotype Endorsement: Black	-0.432
Hispanic Americans – Intelligent to Unintelligent	Stereotype Endorsement: Hispanic	-0.421

Big Five: Neuroticism [Reverse-coded]	Big Five: Neuroticism	-0.414
Whites – Hard-working to Lazy	Stereotype Endorsement: White	-0.397
Hispanic Americans – Compassionate to Insensitive	Stereotype Endorsement: Hispanic	-0.392
Factor 4 ($\alpha = 0.66$)		
Politicians are mainly in politics for their own benefit and not for the benefit of the community.	Distrustful complacency	0.478
Politicians are not really interested in what people like me think.	Right-wing populism	0.466
American society needs to be radically restructured.	System justification	-0.458
Most members of Congress are honest.	Distrustful complacency	0.441
I think that politicians usually do not tell us the true motives for their decisions.	Conspiracy	0.425
In general, I find society to be fair.	System justification	-0.417
Factor 5 ($\alpha = 0.79$)		
The notion of thinking abstractly is appealing to me.	Need for cognition	0.530
I like to have the responsibility of handling a situation that requires a lot of thinking.	Need for cognition	0.519
I really enjoy a task that involves coming up with new solutions to problems.	Need for cognition	0.502
The idea of relying on thought to make my way to the top appeals to me.	Need for cognition	0.489
I enjoy challenging the opinions of others. [Reverse-coded]	Conflict avoidance	-0.482
I find satisfaction in deliberating hard and for long hours.	Need for cognition	0.444
Factor 6 ($\alpha = 0.85$)		
Hispanic Americans – Intelligent to Unintelligent	Stereotype Endorsement: Hispanic	0.423
Asian Americans – Hard-working to Lazy	Stereotype Endorsement: Asian	0.416
Asian Americans – Intelligent to Unintelligent	Stereotype Endorsement: Asian	0.405
Hispanic Americans – Hard-working to Lazy	Stereotype Endorsement: Hispanic	0.394

Hispanic Americans – Compassionate to Insensitive	Stereotype Endorsement: Hispanic	0.372
Hispanic Americans – Honest to Dishonest	Stereotype Endorsement: Hispanic	0.359

Notes: Items sorted by descending absolute magnitude of factor loading for each factor.

This exercise selects 34 different items, but with one exception, they are items from at least three different scales for each new factor. Factor 1 loads most heavily onto four items from the social dominance (dominance) scale and one each from racial resentment and egalitarianism. Factor 2 draws from four scales: three items from the political intolerance battery, two from racial resentment, and one item from the Dirty Dozen narcissism scale. Note immediately that different items in the racial resentment scale load onto the two primary dimensions underlying all 191 survey items.

Factor 3 appears to measure, in part, a willingness to endorse stereotypes of multiple groups. It includes stereotype endorsement items for White, Black, and Hispanic people (five items) and one item from the Big Five neuroticism inventory. Note that the three items about intelligence for Whites, Blacks, and Hispanics scale in the same direction, meaning this is not a measure of relative group stereotypes (i.e., comparisons between groups such as Blacks and Whites) but is instead a willingness to stereotype positively or negatively across groups.

Factor 4 draws on four different scales: two items each from distrustful complacency and system justification and one item each from right-wing populism and conspiracy thinking.

Factor 5 is a notable exception to this pattern of drawing on at least 3 scales: Five of the selected items are from the original need for cognition scale and one is from the conflict avoidance scale (notably, an item related to challenging opinions of others).

Finally, scale 6 draws on four items each from the Hispanic stereotype endorsement battery and two from the Asian stereotype battery. As with factor 3, the shared stereotype items (intelligence and hardworking) have the same direction of loading, meaning that it is not the case that this factor measures a willingness to stereotype one of these groups vis-à-vis the other, but instead expressing correlated stereotypes of these groups. Two of the items about stereotypes of Hispanics are shared with scale 3, but with oppositely signed factor loadings.

Given space constraints it is not possible to fully review the items selected for each new factor scale (as a reminder, this is simply the first 6 most heavily-loading items for each new factor). However, examining the items chosen for the first two factors provides some general intuition for why one might be concerned that these items do not measure fixed traits independent of contemporary political conflict. For the first dimension, four of the items from the dominance dimension of the social dominance orientation scale are about opposition to improving the status of abstract groups, which could be interpreted, among other ways, as immigrants or racial minorities (or women), one is an item from the resentment scale about agreement with the statement that the relative status of Blacks versus Whites is due to lack of effort, a potential explanation for opposition to race-targeted policies, and the last item is agreement with a statement that Americans are too concerned about equality, potentially a blanket endorsement of opposition to policies targeted and traditionally underrepresented/disadvantaged groups. Similarly, the second factor draws on three items about views toward “foreigners” (it is unclear if this relates to legal immigrants, illegal immigrants, or citizens born abroad) and two items about the relative standing of Black Americans, so both are related to views about policies toward non-dominate groups.

More generally, (1) the pattern in which items from different scales load onto a single dimension and (2) items from the same scale load onto different dimensions raises important concerns about how independent and distinct these scales really are. The risk pointed to by the first result is that multiple scales proxy the same factor, meaning those scales are not measuring independent constructs but overlapping instances of the same underlying dimension. Because analyses using scales rarely account for other factors when assessing the predictive power of single scales, the false positives concern raised above is therefore more than hypothetical. The second result means that individual scales likely proxy different mixes of these different dimensions, rather than identifying truly distinct latent dimensions of differences across individuals. Overall, what appears to be evidence of multiple psychological antecedents of political behavior may instead reflect different labels of the same dimension (or a smaller subset of dimensions). Analysis that treats scales as distinct predictors therefore risks over-counting substantively similar variance.

How well do these items predict political choices and behavior?

We first construct the simple mean factor score for each of these 6 new scales using the six items we have selected for each new factor scale (i.e., we give all items equal weight, as with all the existing additive scales we study), reverse coding items as appropriate, and standardize each scale so that it has a mean 0 and standard deviation of 1. We then examine how these scores predict vote choice, partisanship, and ideological self-identification using regression analysis. OLS estimates appear in Table 2 (with parallel estimates using the unrotated factors in Appendix Table A7).

Table 2. Regression of political outcomes on standardized mean scores of six factors (scales).

	(1)	(2)	(3)	(4)	(5)	(6)
	Vote intention (0=Biden, 1=Trump, 0.5=Other)		7-point party ID (1=Strong D, 7=Strong R, 4=Ind./DK)		5-point ideological ID (1=Very lib., 5=Very conserv., 3=Moderate/DK)	
Factor 1	0.168***	0.156***	0.690***	0.655***	0.357***	0.356***
	(0.00947)	(0.0101)	(0.0451)	(0.0467)	(0.0229)	(0.0239)
Factor 2	0.205***	0.209***	0.894***	0.896***	0.503***	0.487***
	(0.00940)	(0.0107)	(0.0440)	(0.0510)	(0.0226)	(0.0265)
Factor 3	0.0172	0.0222	0.0963	0.120	0.154***	0.153***
	(0.0149)	(0.0149)	(0.0684)	(0.0686)	(0.0337)	(0.0343)
Factor 4	0.0302**	0.0288**	0.0604	0.0592	-0.0194	-0.0127
	(0.00941)	(0.00948)	(0.0433)	(0.0438)	(0.0206)	(0.0209)
Factor 5	-0.00629	-0.00508	-0.0511	-0.0601	-0.0556*	-0.0541*
	(0.00954)	(0.00974)	(0.0443)	(0.0462)	(0.0228)	(0.0241)
Factor 6	0.0149	0.0141	0.0520	0.0358	0.0842*	0.0783*
	(0.0149)	(0.0150)	(0.0688)	(0.0692)	(0.0339)	(0.0345)
Constant	0.474***	0.668***	3.786***	4.572***	3.028***	3.057***
	(0.00944)	(0.0443)	(0.0417)	(0.181)	(0.0202)	(0.0854)
Controls	No	Yes	No	Yes	No	Yes
N	1588	1588	2001	2001	2001	2001

Notes: OLS coefficients with robust standard errors in parentheses. Columns 2, 4, and 6 use controls for age, gender, income, race, and education. Responses in all columns collected for all panelists between December 11, 2023–February 29, 2024.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Factor 1 and Factor 2 are powerful predictors of all three of these outcomes. For example, a one standard deviation increase in the first-dimension scale score is associated with a 17-point swing from Biden to Trump (column 1), moving 0.7 steps on the 7-point party identification toward being a Republican (column 3), and roughly one-third of a unit on the 5-point ideology scale toward being a conservative (column 5). (The t-statistics of these estimates

are also large, ranging from 15 to 18 even in models that control for the other five factors.) Effects for factor 2 for the same outcomes are even larger: a 21-point to Trump vote swing, 0.9 units toward the Republican party, and 0.5 units of conservative ideology, with t-statistics ranging from 20 to 22. As a reminder, the scaling exercise was conducted without access to any of these outcomes, so this correlation is not induced by creating scales to maximize their predictive power but instead by the underlying correlation in the survey items used to create these factors, which also predict these outcomes. Note that these estimates are very similar in the even numbered column specifications that also control for other non-political demographics.

The effects of the remaining four factors are much smaller. Controlling for the other factors, the maximum coefficient (absolute value) for factor 4 is 0.06, for factor 5 it is 0.06, and for factor 6 it is 0.08. Factor 3 produces the only other statistically significant association with any of these three outcomes: A one standard deviation increase in factor 3 is associated with a 0.15 unit increase in conservative ideology, but the estimated effects for voting and partisanship are smaller (0.02 to 0.12) and insignificant. While those effects may be theoretically interesting, they are not generally statistically significant (after accounting for the other scales) and appear much less consequential for predicting these outcomes.

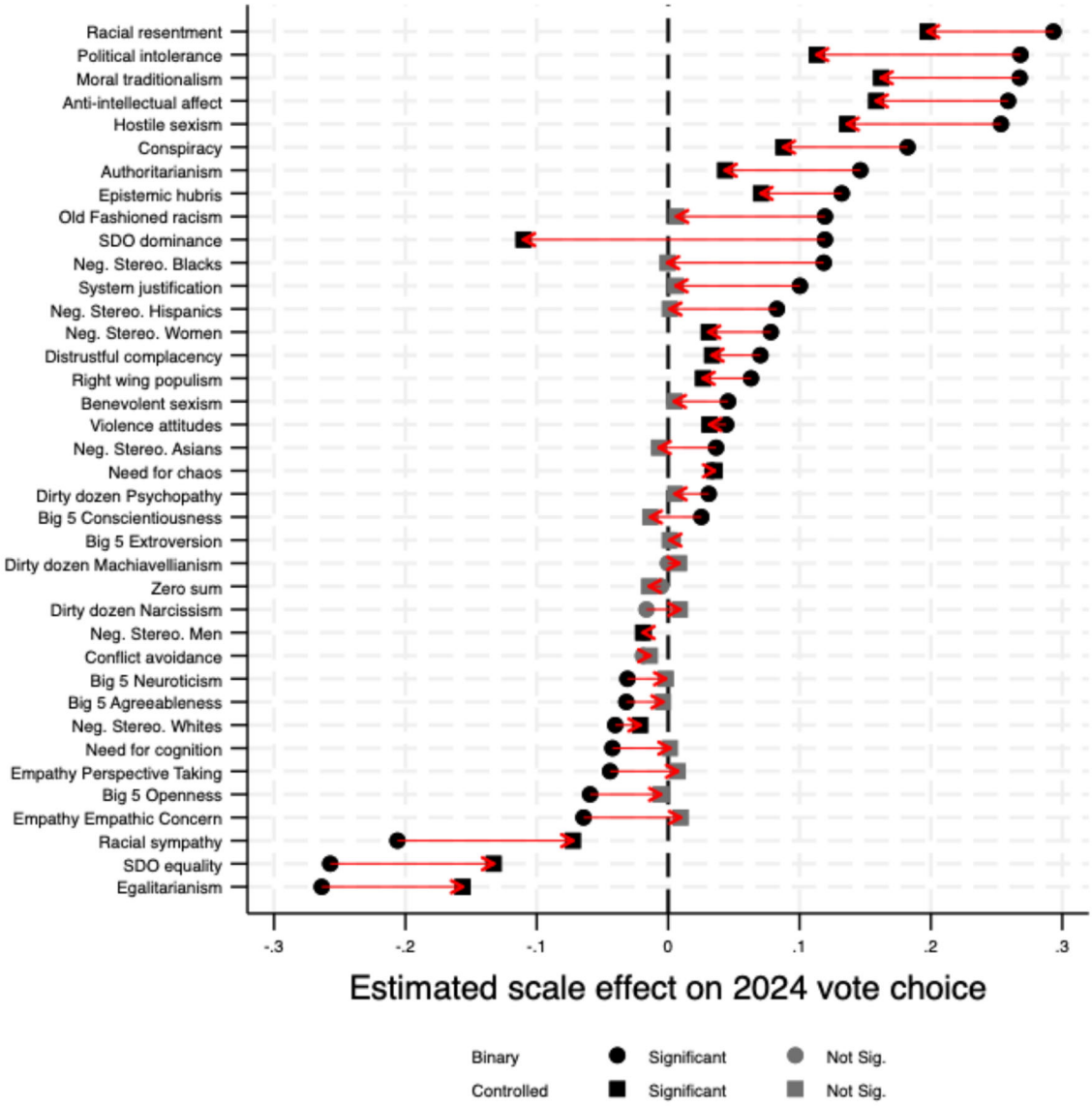
The predictive power of the first two latent dimensions (scales) reveal that these items capture the structure of contemporary political conflict with comparably large magnitudes. The remaining four factors likely capture narrower attitudinal tendencies and some degree of noise. In sum, the above regressions suggest that our novel factor scales recover two dominant dimensions that exhibit significant predictive validity. Thus, nominally independent scales may

simply reflect the variance of these correlates of overarching contemporary American political conflict.

This analysis yields two natural follow-on questions: (1) How much predictive power do the scales we study have after accounting for these two dominant factors? (2) If these two factors underlie a great deal of variance in the different survey items used to construct them, what is each scale measuring?

To address the first question, we conducted a series of OLS binary regressions where we predicted 2024 vote choice (the first outcome shown in Table 2) using each original scale individually. We then reran the same models after accounting for our operationalization of the first two factors identified above, where each dimension was constructed as a linear additive scale from the 6 individual survey items that loaded most heavily onto it. We plot the results of this analysis in Figure 6.

Figure 6. Effects of individual scales on vote choice, before and after controlling for two novel factors identified in factor analysis.



Notes: Dependent variable is vote choice (where 1=Trump, 0=Biden, 0.5=other). Controlled model adds factor 1 and factor 2 mean scales constructed using 6 items each. Among estimates significant in binary regression, (absolute value of) controlled estimate is on average 0.28 of binary estimate.

The vertical axis is each of the scales we study and the horizontal axis plots two quantities: The estimated coefficient for each scale from a binary regression where each scale is

used to predict vote choice (circles) and the same regression result when controlling for the two new factor scales (squares). Estimates that are significant are colored black and estimates that are not significant are colored gray. The arrows show the difference that controlling for the two first dimension factor scales has on the estimated effect of each scale on vote choice. The scales are ordered from top to bottom by the size of the binary estimates.

The results are stark: Controlling for the two factors we identified from our factor analysis of the underlying survey items tends to greatly reduce the estimated effect of each scale on vote choice. 32 of the estimates are significant in a binary regression and, of these, 18 (56%) are still significant after controlling for the two new primary factors. In nearly all cases, estimates shrink and are closer to 0 in the controlled regressions. This is a powerful result, in part, because we construct each measure of these new factors as a simple additive scale using only the 6 survey items that load most heavily on to each dimension rather than using all of the items that load onto each of these dimensions. Among estimates that were statistically significant in the binary regression, the (absolute value of the) controlled coefficient is on average only 38% of the (absolute value of the) binary estimate.[18]

One might be concerned that this test is stacked against finding that scale effects persist because a scale's items might be used in both the binary regression and in constructing the alternative factor scores. For this reason, we also conducted a similar exercise where we compared the predictive power of each scale alone to its predictive power when we factor

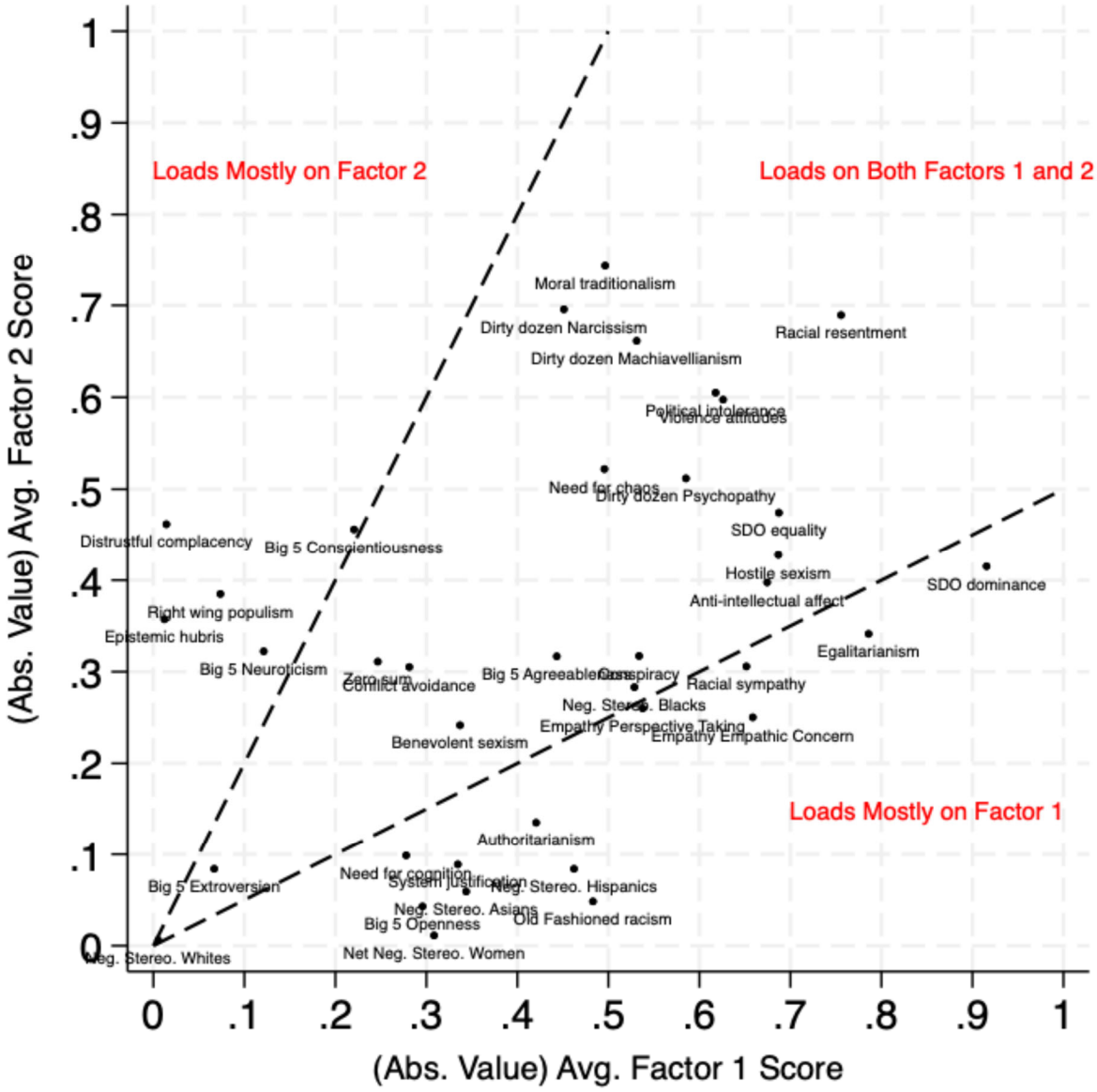
[18] In only two cases are the estimated effects of a scale larger (in an absolute sense) when controlling for the two factor scales, and in both cases these effects are significant. These are need for chaos and negative stereotype toward men. In both cases, the estimated effect of each scale is small.

analyzed all the survey items from *the other scale* to calculate novel factor 1 and factor 2 scores (i.e., we accounted for the shared variance across all other scales). In this design, a scale's items cannot directly affect the estimated factor 1 or factor 2 score. The results of this analysis provided even stronger evidence for the dimensionality concern: Among coefficients significant in a binary regression, the average absolute coefficient is only 32% of its original estimate once the common variance across all other scales is accounted for. See Appendix Figure A4 for a graphical display of this result.

Given that accounting for the common variance across scales tends to substantially reduce the estimated effect of each individual scale, this implies many scales also measure components of this shared first and/or second dimension. We therefore turn to answering our second question, which is what does each individual scale measure? It is most straightforward to present the result of this analysis graphically. In Figure 7, we display the relative weight that each scale as constructed receives, on average, for factor 1 and factor 2 in our item-level factor analysis (and include a parallel plot using the unrotated factors in Appendix Figure A5). To construct the figure, we first rescaled the individual factor loadings linearly so that the maximum absolute value factor loading for each factor was 1 to make it easier to compare across dimensions (i.e., if the largest absolute value individual factor loading was -0.6, we divided all factor loadings by 0.6). Then we then calculated for each scale the average of these rescaled factor loading across all the survey items in the scale for both the first and second dimension identified by our factor analysis. We then took the absolute value of this scale-level average, so positive values indicate that a scale on average loads highly positively or negatively onto each dimension.

We plot for each scale the absolute value of the average for dimension 1 (the horizontal axis) and dimension 2 (the vertical axis). Additionally, we plot 2 upward-facing diagonal lines for the functions $y = 2x$ and $y = 0.5x$ to distinguish scales that load most heavily onto factor 1 (lower right triangle), factor 2 (upper right triangle), or both (middle triangle). In the middle triangle, the (absolute) average factor score for the less important dimension is at least half of the (absolute) average factor score for the more important dimension.

Figure 7. Relative loadings of each scale for two underlying factors.



Notes: See text for details. Absolute value of average of factor 1 and factor 2 scores across items used to construct each scale, with each factor score rescaled linearly to have a maximum (or minimum) of 1 (-1).

Several important patterns are immediately apparent. First, many scales appear to have about the same weight for either factor 1 or factor 2. For example, on the first (horizontal) dimension, both egalitarianism and racial resentment have average loadings of between 0.75 and 0.8. Similarly, on the second (vertical) dimension, distrustful complacency,

conscientiousness, and SDO (equality) have average loadings between 0.4 and 0.5. These scales therefore appear to measure, at about the same rate, these underlying dimensions. This means that estimates using these scales capture the same "amounts" of each of these underlying dimensions.

Second, many scales measure relative similar "mixes" of the two dimensions. The easiest way to see this is to project a diagonal line outward from the origin, such as the lower dashed line that ends near SDO (dominance) at the near the coordinates (.9,.4). Points along the same linear function line capture dimension 1 and dimension 2 in similar (absolute) ratios. Several scales are near this line, including dominance, egalitarianism, anti-intellectual affect, racial sympathy, conspiracy thinking, and negative stereotypes about Blacks. We show above that these scales are correlated with one other, and this analysis provides some structure for those correlations: These 6 scales share underlying common variance ratios in the two key dimensions that predict many political outcomes and which emerged when factor analyzing all of these individual survey items without regard for scale label.

Third, many scales appear to be given roughly similar weight in both the first and second dimension. (These scales appear between the two dashed lines in the middle triangle of the figure.) For example, resentment, moral traditionalism, the three components of the Dirty Dozen, political intolerance, violence attitudes, need for chaos, SDO: equality, conflict avoidance, and zero sum, among others, appear to measure both factors to about the same degree. Relatedly, very few scales appear to be "pure" types. The notable exceptions (those scales that are far from the origin in one dimension but close to the origin in the other) include for factor 1 the scales clustered in the lower right portion of the figure and for factor 2 the

scales clustered in the upper left portion of the figure. For factor 1, these are net negative stereotypes about women, negative stereotypes about Asians and Hispanics, old-fashioned racism, system justification, and authoritarianism, among others. For factor two these are epistemic hubris, right-wing populism, and distrustful complacency. Importantly, these bundles of scales are distinct from one another, but in each bundle appear to draw to equal amounts on shared underlying variance in either factor.

Cumulatively, combined with the prior result that a simple six-item measure of each latent factor appears to dramatically diminish the predictive power of nearly all of these scales in explaining key political outcomes, it may be that the underlying latent trait space is much smaller than researchers have assumed. And, even if one disagrees with this assessment, it further demonstrates that the concern that few research designs are well powered to identify effects caused by any given scale when there is a possibility that another scale (or an omitted factor correlated with either underlying dimension that we identify) could also cause measured effects. Thus, even if one chooses to reject the dimensionality concern, the estimation problem is clearly present given the underlying observed correlations across scales and the observed pattern of scales mapping into these two shared dimensions.

Discussion and Conclusion

Political scientists use psychometric scales in their efforts to understand politics and test theories of human behavior. The number of scales researchers consider continues to grow and they are widely used in empirical research. But we lack systematic evidence about the independence of these scales from one another, leading to concerns that the estimated effects of these scales in predicting outcomes are biased due to their correlations with other scales and

the possibility that multiple scales are not theoretically independent but instead different operationalizations (measures) of shared underlying human traits. Both the former concern, which we label the estimation problem, and the latter concern, which we label the dimensionality problem, have important implications for our theoretical and empirical understanding of key patterns of politics. In short, we build from the tradition of Campbell and Fiske (1959) to argue that if a scale does not offer predictive power beyond the effect of other scales and factors common to multiple scales—*discriminant predictive validity*—a researcher has not demonstrated a proposed new trait is novel rather than a different measure of some previously identified trait(s).

In this paper we describe the theoretical basis for these concerns. We then document the widespread use of scales in contemporary political science research, including how researchers use scales and whether they conduct tests that would address the estimation and dimensionality concerns. These tests concern both statistical independence and construct validity.

Next, we describe the novel data we collected to assess empirically these concerns about estimation and dimensionality. For 2,001 respondents we asked all the individual survey items necessary to construct nearly 40 commonly used scales. We then assessed the correlations across scales and demonstrated the implications of the observed correlations for the possibility of biased estimates of scale effects, demonstrating empirically the estimation concern.

Additionally, we conducted exploratory factor analysis at both the scale- and survey-item level to assess the likelihood of the dimensionality concern. Contrary to the assumptions that

scales are independent, when considering a large number of scales (and their component items) simultaneously, we find that many scales appear to measure a much smaller number of latent dimensions—many fewer than assumptions about scale independence appear to warrant. Moreover, accounting for just the two factors with the greatest common variance substantially reduces the apparent predictive power of almost every individual scale, raising questions about whether tests of predictive validity in fact demonstrate a scale's empirical and theoretical importance.

Overall, contemporary empirical scholarship using survey scales is likely substantially underpowered to distinguish the effects of individual scales from other correlated scales, meaning empirical tests used to demonstrate scale importance and the point estimates from most extant analysis are likely less informative than researchers would prefer. Beyond this problem of statistical power, however, the dimensionality problem may be of even greater importance. Commonly used scales (including those appearing most frequently in top journals such as racial resentment, social dominance orientation, and authoritarianism) appear to measure a much smaller number of latent dimensions, or mixes of important dimensions, meaning that the proliferation of scales masks a much smaller number of shared dimensions of human differences.

Importantly, we are not arguing that seeking to understand dimensions of human differences is unimportant, theoretically or empirically. However, as implemented, we do believe extant work is insufficient to the estimation and dimensionality concerns we identify. At best, estimates of scale effects can be biased (and underpowered) in the absence of accounting for the effects of other independent but correlated scales (with theoretically motivated effects).

This means extant tests of predictive validity are likely not sufficient to demonstrate scale importance. At worst, binary tests of scale effects across scales are creating the impression of much greater variability in human traits that exaggerate the apparent dimensionality of such differences. Less biased predictive tests and claims about scale independence should, in the future, be performed considering the evidence we present about the correlations between individual scales and the correlations across individual scale items used to construct those scales.

In sum, our results indicate that the continued introduction of new scales in political science may not represent the discovery of new dimensions of human difference but rather different operationalizations of a smaller set of factors already captured by existing measures. Demonstrating that a scale has high internal consistency (i.e., high Cronbach's alpha) and predicts theoretically relevant outcomes is the current standard in most published work, but we argue it is insufficient to establish that a scale measures a novel construct. Specifically, future scale validation might require some combination of the following: (1) testing discriminant validity against a broad range of existing scales (not just a few conceptually similar measures); (2) demonstrating incremental predictive power after controlling for shared dimensions across existing scales; and (3) using exploratory factor analysis to show which scale items load onto novel versus pre-existing factors. Without such tests of *discriminant predictive validity*, the proliferation of scales risks creating an illusion of theoretical progress while relabeling already-identified psychological dimensions. Our matrices showing the correlations across 36 commonly used scales provide a starting point for researchers seeking to test a new construct, allowing them to identify which existing scales are most likely to confound their estimates and should

therefore be prioritized in validation testing. Absent this type of analysis, the estimation and dimensionality concerns we identify likely lead both to biased empirical estimates and a potentially misleading picture about the range of human differences measured by new scales.

We also note that a key challenge in undertaking these tests is that most political science studies are likely underpowered to distinguish among the effects of correlated scales, even if they have distinct effects. Our survey draws on 2,001 respondents, but given the high correlations between scales, this sample still yields false positives at very high rates and appropriate "horserace" tests in which a saturated model yields null effects across multiple scales is a likely outcome. While such work is likely difficult, failing to account for the patterns we demonstrate may bias our assessments of the breadth and importance of new and existing measures of human differences.

We close by noting some remaining challenges in our analysis. One is the length of the surveys we fielded. One might be concerned that respondents taking two long surveys would engage in satisficing behavior, suggesting the value of validating this analysis after instead fielding the component scales in multiple shorter surveys, while also fully randomizing scale order. Alternatively, one could compare our estimates between smaller subsets of scales to estimates derived from novel samples asked only those subsets. Similarly, future research might explore how response format options generate correlations across scales, perhaps by randomizing different response options to groups of scale (e.g., 5- versus 7-point agree-disagree options).

References

Amsalem, Eran, and Lior Sheffer. 2023. “Personality and the Policy Positions of Politicians.” *Political Psychology* 44(1): 119–38. doi:10.1111/pops.12821.

Ansolabehere, Stephen, Jonathan Rodden, and James M. Snyder. 2008. “The Strength of Issues: Using Multiple Measures to Gauge Preference Stability, Ideological Constraint, and Issue Voting.” *American Political Science Review* 102(2): 215–32. doi:10.1017/S0003055408080210.

Armendáriz Miranda, Paula. 2022. “Explaining Autocratic Support: The Varying Effects of Threat on Personality.” *Political Psychology* 43(6): 993–1007. doi:10.1111/pops.12794.

Bakker, Bert N., and Yphtach Lelkes. 2018. “Selling Ourselves Short? How Abbreviated Measures of Personality Change the Way We Think about Personality and Politics.” *The Journal of Politics* 80(4): 1311–25. doi:10.1086/698928.

Bakker, Bert N., Gijs Schumacher, and Matthijs Rooduijn. 2021. “The Populist Appeal: Personality and Antiestablishment Communication.” *The Journal of Politics* 83(2): 589–601. doi:10.1086/710014.

Banda, Kevin K., and Erin C. Cassese. 2022. “Hostile Sexism, Racial Resentment, and Political Mobilization.” *Political Behavior* 44(3): 1317–35. doi:10.1007/s11109-020-09674-7.

Barker, David C., Ryan Detamble, and Morgan Marietta. 2022. “Intellectualism, Anti-Intellectualism, and Epistemic Hubris in Red and Blue America.” *American Political Science Review* 116(1): 38–53. doi:10.1017/S0003055421000988.

Baron, Denise, Benjamin Lauderdale, and Jennifer Sheehy-Skeffington. 2023. “A Leader Who Sees the World as I Do: Voters Prefer Candidates Whose Statements Reveal Matching Social-Psychological Attitudes.” *Political Psychology* 44(4): 893–916. doi:10.1111/pops.12891.

Batista Pereira, Frederico. 2021. “Do Female Politicians Face Stronger Backlash for Corruption Allegations? Evidence from Survey-Experiments in Brazil and Mexico.” *Political Behavior* 43(4): 1561–80. doi:10.1007/s11109-020-09602-9.

Beattie, Peter, Rong Chen, and Karim Bettache. 2022. “When Left Is Right and Right Is Left: The Psychological Correlates of Political Ideology in China.” *Political Psychology* 43(3): 457–88. doi:10.1111/pops.12776.

Bonilla, Tabitha, Alexandra Filindra, and Nazita Lajevardi. 2022. “How Source Cues Shape Evaluations of Group-Based Derogatory Political Messages.” *The Journal of Politics* 84(4): 1979–96. doi:10.1086/720308.

Bruder, Martin, Peter Haffke, Nick Neave, Nina Nouripanah, and Roland Imhoff. 2013.

"Measuring Individual Differences in Generic Beliefs in Conspiracy Theories Across Cultures: Conspiracy Mentality Questionnaire." *Frontiers in Psychology* 4. doi:10.3389/fpsyg.2013.00225.

Cacioppo, John T., Richard E. Petty, and Chuan Feng Kao. 1984. "The Efficient Assessment of Need for Cognition." *Journal of Personality Assessment* 48(3): 306–7. doi:10.1207/s15327752jpa4803_13.

Campbell, Donald T., and Donald W. Fiske. 1959. "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix." *Psychological Bulletin* 56(2): 81–105. doi:10.1037/h0046016.

Cawvey, Matthew. 2023. "Extraversion and External Efficacy: The Moderating Role of Corruption." *Political Psychology* 44(1): 157–76. doi:10.1111/pops.12828.

Cepuran, Colin J. G., and Justin Berry. 2022. "Whites' Racial Resentment and Perceived Relative Discrimination Interactively Predict Participation." *Political Behavior* 44(2): 1003–24. doi:10.1007/s11109-022-09786-2.

Chinoy, Sahil, Nathan Nunn, Sandra Sequeira, and Stefanie Stantcheva. 2023. *Zero-Sum Thinking and the Roots of US Political Differences*. Cambridge, MA: National Bureau of Economic Research. doi:10.3386/w31688.

Choma, Becky, Gordon Hodson, Arvin Jagayat, and Mark R. Hoffarth. 2020. "Right-Wing Ideology as a Predictor of Collective Action: A Test Across Four Political Issue Domains." *Political Psychology* 41(2): 303–22. doi:10.1111/pops.12615.

Chudy, Jennifer. 2021. "Racial Sympathy and Its Political Consequences." *The Journal of Politics* 83(1): 122–36. doi:10.1086/708953.

Cronbach, Lee J. 1951. "Coefficient Alpha and the Internal Structure of Tests." *Psychometrika* 16(3): 297–334. doi:10.1007/BF02310555.

Cronbach, Lee J., and Paul E. Meehl. 1955. "Construct Validity in Psychological Tests." *Psychological Bulletin* 52(4): 281–302. doi:10.1037/h0040957.

Davis, Mark H. 1983. "Measuring Individual Differences in Empathy: Evidence for a Multidimensional Approach." *Journal of Personality and Social Psychology* 44(1): 113–26. doi:10.1037/0022-3514.44.1.113.

DeSante, Christopher D., and Candis Watts Smith. 2020. "Less Is More: A Cross-Generational Analysis of the Nature and Role of Racial Attitudes in the Twenty-First Century." *The Journal of Politics* 82(3): 967–80. doi:10.1086/707490.

Ditonto, Tessa. 2020. "The Mediating Role of Information Search in the Relationship Between Prejudice and Voting Behavior." *Political Psychology* 41(1): 71–88.

doi:10.1111/pops.12599.

Duckitt, John, and Chris G. Sibley. 2009. “A Dual-Process Motivational Model of Ideology, Politics, and Prejudice.” *Psychological Inquiry* 20(2–3): 98–109. doi:10.1080/10478400903028540.

Dynes, Adam M., Hans J. G. Hassell, and Matthew R. Miles. 2023. “Personality Traits and Approaches to Political Representation and Responsiveness: An Experiment in Local Government.” *Political Behavior* 45(4): 1791–1811. doi:10.1007/s11109-022-09800-7.

Eggers, Andrew C., Guadalupe Tuñón, and Allan Dafoe. 2024. “Placebo Tests for Causal Inference.” *American Journal of Political Science* 68(3): 1106–21. doi:10.1111/ajps.12818.

Ellis, Christopher, and Christopher Faricy. 2020. “Race, ‘Deservingness,’ and Social Spending Attitudes: The Role of Policy Delivery Mechanism.” *Political Behavior* 42(3): 819–43. doi:10.1007/s11109-018-09521-w.

Enders, Adam M., and Judd R. Thornton. 2022. “Racial Resentment, Electoral Loss, and Satisfaction with Democracy Among Whites in the United States: 2004–2016.” *Political Behavior* 44(1): 389–410. doi:10.1007/s11109-021-09708-8.

Feldman, Stanley, and Leonie Huddy. 2005. “Racial Resentment and White Opposition to Race-Conscious Programs: Principles or Prejudice?” *American Journal of Political Science* 49(1): 168–83. doi:10.1111/j.0092-5853.2005.00117.x.

Filindra, Alexandra, Noah J. Kaplan, and Beyza E. Buyuker. 2022. “Beyond Performance: Racial Prejudice and Whites’ Mistrust of Government.” *Political Behavior* 44(2): 961–79. doi:10.1007/s11109-022-09774-6.

Fornell, Claes, and David F. Larcker. 1981. “Evaluating Structural Equation Models with Unobservable Variables and Measurement Error.” *Journal of Marketing Research* 18(1): 39. doi:10.2307/3151312.

Georgarakis, George N. 2023. “Yikes! The Effect of Incidental Disgust and Information on Public Attitudes During the COVID -19 Pandemic.” *Political Psychology* 44(3): 493–513. doi:10.1111/pops.12865.

Glick, Peter, and Susan T. Fiske. 1996. “The Ambivalent Sexism Inventory: Differentiating Hostile and Benevolent Sexism.” *Journal of Personality and Social Psychology* 70(3): 491–512. doi:10.1037/0022-3514.70.3.491.

Goldsmith, Benjamin E., and Lars J. K. Moen. 2025. “The Personality of a Personality Cult? Personality Characteristics of Donald Trump’s Most Loyal Supporters.” *Political Psychology* 46(1): 225–43. doi:10.1111/pops.12991.

Goldstein, Susan B. 1999. “Construction and Validation of a Conflict Communication Scale.” *Journal of Applied Social Psychology* 29(9): 1803–32. doi:10.1111/j.1559-1816.1999.tb00153.x.

Goren, Paul. 2008. “The Two Faces of Government Spending.” *Political Research Quarterly* 61(1): 147–57. doi:10.1177/1065912907311881.

Goren, Paul, Harald Schoen, Jason Reifler, Thomas Scotto, and William Chittick. 2016. “A Unified Theory of Value-Based Reasoning and U.S. Public Opinion.” *Political Behavior* 38(4): 977–97. doi:10.1007/s11109-016-9344-x.

Gosling, Samuel D, Peter J Rentfrow, and William B Swann. 2003. “A Very Brief Measure of the Big-Five Personality Domains.” *Journal of Research in Personality* 37(6): 504–28. doi:10.1016/S0092-6566(03)00046-1.

Gross, Kimberly, and Julie Wronski. 2021. “Helping the Homeless: The Role of Empathy, Race and Deservingness in Motivating Policy Support and Charitable Giving.” *Political Behavior* 43(2): 585–613. doi:10.1007/s11109-019-09562-9.

Haselswerdt, Jake. 2022. “Who Benefits? Race, Immigration, and Assumptions About Policy.” *Political Behavior* 44(1): 271–318. doi:10.1007/s11109-020-09608-3.

Heide-Jørgensen, Tobias, Peter Thisted Dinesen, and Kim Mannemar Sønderskov. 2023. “Personality and Roots of Welfare State Support: How Openness to Experience Moderates the Influence of Self-Interest and Ideology on Redistributive Preferences.” *Political Behavior* 45(4): 1467–89. doi:10.1007/s11109-022-09775-5.

Helminen, Vilja, Hanna Wass, Anu Kantola, and Marko Elovainio. 2024. “Nordic Authoritarianism: Child-rearing Values and Political Behavior in a Multiparty Context.” *Political Psychology* 45(1): 91–111. doi:10.1111/pops.12915.

Henseler, Jörg, Christian M. Ringle, and Marko Sarstedt. 2015. “A New Criterion for Assessing Discriminant Validity in Variance-Based Structural Equation Modeling.” *Journal of the Academy of Marketing Science* 43(1): 115–35. doi:10.1007/s11747-014-0403-8.

Ho, Arnold K., Jim Sidanius, Nour Kteily, Jennifer Sheehy-Skeffington, Felicia Pratto, Kristin E. Henkel, Rob Foels, and Andrew L. Stewart. 2015. “The Nature of Social Dominance Orientation: Theorizing and Measuring Preferences for Intergroup Inequality Using the New SDO_7 Scale.” *Journal of Personality and Social Psychology* 109(6): 1003–28. doi:10.1037/pspi0000033.

Hopkins, Daniel J. 2021. “The Activation of Prejudice and Presidential Voting: Panel Evidence from the 2016 U.S. Election.” *Political Behavior* 43(2): 663–86. doi:10.1007/s11109-019-09567-4.

Jacoby, William G. 2008. “Comment: The Dimensionality of Public Attitudes toward

Government Spending." *Political Research Quarterly* 61(1): 158–61. doi:10.1177/1065912907309860.

Jardina, Ashley. 2021. "In-Group Love and Out-Group Hate: White Racial Attitudes in Contemporary U.S. Elections." *Political Behavior* 43(4): 1535–59. doi:10.1007/s11109-020-09600-x.

Jedinger, Alexander, and Axel M. Burger. 2020. "The Ideological Foundations of Economic Protectionism: Authoritarianism, Social Dominance Orientation, and the Moderating Role of Political Involvement." *Political Psychology* 41(2): 403–24. doi:10.1111/pops.12627.

Jonason, Peter K., and Gregory D. Webster. 2010. "The Dirty Dozen: A Concise Measure of the Dark Triad." *Psychological Assessment* 22(2): 420–32. doi:10.1037/a0019265.

Jöreskog, K. G. 1969. "A General Approach to Confirmatory Maximum Likelihood Factor Analysis." *Psychometrika* 34(2): 183–202. doi:10.1007/BF02289343.

Jung, Jae-Hee, and Scott Clifford. 2025. "Varieties of Values: Moral Values Are Uniquely Divisive." *American Political Science Review* 119(1): 462–78. doi:10.1017/S0003055424000443.

Karpowitz, Christopher F., Tyson King-Meadows, J. Quin Monson, and Jeremy C. Pope. 2021. "What Leads Racially Resentful Voters to Choose Black Candidates?" *The Journal of Politics* 83(1): 103–21. doi:10.1086/708952.

Kay, Aaron C., and John T. Jost. 2003. "Complementary Justice: Effects of 'Poor but Happy' and 'Poor but Honest' Stereotype Exemplars on System Justification and Implicit Activation of the Justice Motive." *Journal of Personality and Social Psychology* 85(5): 823–37. doi:10.1037/0022-3514.85.5.823.

Kevins, Anthony, and Barbara Vis. 2023. "Do Public Consultations Reduce Blame Attribution? The Impact of Consultation Characteristics, Gender, and Gender Attitudes." *Political Behavior* 45(3): 1121–42. doi:10.1007/s11109-021-09751-5.

Kinder, Donald R., and Lynn M. Sanders. 1996. *Divided by Color: Racial Politics and Democratic Ideals*. Chicago: University of Chicago Press.

Ksiazkiewicz, Aleksander, and Amanda Friesen. 2021. "The Higher Power of Religiosity Over Personality on Political Ideology." *Political Behavior* 43(2): 637–61. doi:10.1007/s11109-019-09566-5.

Lalot, Fanny, Dominic Abrams, Maria S. Heering, Jacinta Babaian, Hilal Ozkececi, Linus Peitz, Kaya Davies Hayon, and Jo Broadwood. 2023. "Distrustful Complacency and the COVID - 19 Vaccine: How Concern and Political Trust Interact to Affect Vaccine Hesitancy." *Political Psychology* 44(5): 983–1011. doi:10.1111/pops.12871.

Luttig, Matthew D. 2021. “Reconsidering the Relationship between Authoritarianism and Republican Support in 2016 and Beyond.” *The Journal of Politics* 83(2): 783–87. doi:10.1086/710145.

Masuoka, Natalie, Christian Grose, and Jane Junn. 2023. “Sexual Harassment and Candidate Evaluation: Gender and Partisanship Interact to Affect Voter Responses to Candidates Accused of Harassment.” *Political Behavior* 45(3): 1285–1307. doi:10.1007/s11109-021-09761-3.

Munis, B. Kal. 2022. “Us Over Here Versus Them Over There…Literally: Measuring Place Resentment in American Politics.” *Political Behavior* 44(3): 1057–78. doi:10.1007/s11109-020-09641-2.

Mußotter, Marlene. 2024. “On Nation, Homeland, and Democracy: Toward a Novel Three-factor Measurement Model for Nationalism and Patriotism. Evidence from Two Representative Studies.” *Political Psychology* 45(6): 903–21. doi:10.1111/pops.12945.

Oceno, Marzia, Nicholas A. Valentino, and Carly Wayne. 2023. “The Electoral Costs and Benefits of Feminism in Contemporary American Politics.” *Political Behavior* 45(1): 153–73. doi:10.1007/s11109-021-09692-z.

Ollerenshaw, Trent. 2024. “The Conditional Effects of Authoritarianism on COVID-19 Pandemic Health Behaviors and Policy Preferences.” *Political Behavior* 46(1): 233–56. doi:10.1007/s11109-022-09828-9.

Osborne, Danny, Yanshu Huang, Nickola C. Overall, Robbie M. Sutton, Aino Petterson, Karen M. Douglas, Paul G. Davies, and Chris G. Sibley. 2022. “Abortion Attitudes: An Overview of Demographic and Ideological Differences.” *Political Psychology* 43(S1): 29–76. doi:10.1111/pops.12803.

Pantazi, Myrto, Kostas Papaioannou, and Jan-Willem Van Prooijen. 2022. “Power to the People: The Hidden Link Between Support for Direct Democracy and Belief in Conspiracy Theories.” *Political Psychology* 43(3): 529–48. doi:10.1111/pops.12779.

Petersen, Michael Bang, Mathias Osmundsen, and Kevin Arceneaux. 2023. “The ‘Need for Chaos’ and Motivations to Share Hostile Political Rumors.” *American Political Science Review* 117(4): 1486–1505. doi:10.1017/S0003055422001447.

Peyton, Kyle, and Gregory A. Huber. 2021. “Racial Resentment, Prejudice, and Discrimination.” *The Journal of Politics* 83(4): 1829–36. doi:10.1086/711558.

Pizarro, José J., Huseyin Cakal, Lander Méndez, Larraitz N. Zumeta, Marcela Gracia-Leiva, Nekane Basabe, Ginés Navarro-Carrillo, et al. 2024. “Sociopolitical Consequences of COVID-19 in the Americas, Europe, and Asia: A Multilevel, Multicountry Investigation of Risk Perceptions and Support for Antidemocratic Practices.” *Political Psychology* 45(2): 407–33. doi:10.1111/pops.12930.

Prati, Francesca, Felicia Pratto, Fouad Zeineddine, Joseph Sweetman, Antonio Aiello, Nebojša Petrović, and Monica Rubini. 2022. “From Social Dominance Orientation to Political Engagement: The Role of Group Status and Shared Beliefs in Politics Across Multiple Contexts.” *Political Psychology* 43(1): 153–75. doi:10.1111/pops.12745.

Schulz, Anne, Philipp Müller, Christian Schemer, Dominique Stefanie Wirz, Martin Wettstein, and Werner Wirth. 2018. “Measuring Populist Attitudes on Three Dimensions.” *International Journal of Public Opinion Research* 30(2): 316–26. doi:10.1093/ijpor/edw037.

Simas, Elizabeth N., Scott Clifford, and Justin H. Kirkland. 2020. “How Empathic Concern Fuels Political Polarization.” *American Political Science Review* 114(1): 258–69. doi:10.1017/S0003055419000534.

Stephens-Dougan, Lafleur. 2023. “White Americans’ Reactions to Racial Disparities in COVID-19.” *American Political Science Review* 117(2): 773–80. doi:10.1017/S000305542200051X.

Strawbridge, Michael G., Heather Silber Mohamed, and Jennifer Lucas. 2025. “White Racial Resentment and Gender Attitudes: An Enduring Connection or an Artifact of the 2016 Election?” *Political Behavior* 47(2): 801–24. doi:10.1007/s11109-024-09970-6.

Thompson, Andrew Ifedapo, and Ethan C. Busby. 2023. “Defending the Dog Whistle: The Role of Justifications in Racial Messaging.” *Political Behavior* 45(3): 1241–62. doi:10.1007/s11109-021-09759-x.

Uscinski, Joseph E., and Joseph M. Parent. 2014. *American Conspiracy Theories*. Oxford University Press. doi:10.1093/acprof:oso/9780199351800.001.0001.

Van Prooijen, Jan-Willem, Talia Cohen Rodrigues, Carlotta Bunzel, Oana Georgescu, Dániel Komáromy, and André P. M. Krouwel. 2022. “Populist Gullibility: Conspiracy Theories, News Credibility, Bullshit Receptivity, and Paranormal Belief.” *Political Psychology* 43(6): 1061–79. doi:10.1111/pops.12802.

Vasilopoulos, Pavlos, and Justin Robinson. 2025. “Authoritarianism, Political Attitudes, and Vote Choice: A Longitudinal Analysis of the British Electorate.” *Political Behavior* 47(2): 503–27. doi:10.1007/s11109-024-09961-7.

Westwood, Sean J., and Erik Peterson. 2022. “The Inseparability of Race and Partisanship in the United States.” *Political Behavior* 44(3): 1125–47. doi:10.1007/s11109-020-09648-9.

Appendix

Table of Contents

Appendix A1. Survey fielding.

This study, administered by YouGov, included a group of 2,001 respondents who were interviewed specifically to measure the psychometric scales used in this paper. The survey was fielded February 2–15, 2024. It was split into two surveys that each took around 20 minutes to complete. Data were collected as part of the Stanford-Arizona-Yale Presidential Election Study (SAY24). YouGov's panel is an opt-in panel where respondents are invited to take surveys in return for "points" they can redeem for rewards. Samples are selected to be demographically diverse so that weighting can be performed to match targets, but in the case of this sample it is not designed to match national targets without weighting.

This survey includes a baseline sample of approximately 130,000 respondents in the United States recruited from YouGov's online survey panel. To ensure low attrition, respondents in the SAY24 panel tend to have longer histories with YouGov that reduces the likelihood they will drop out of the sample. Data in this manuscript are unweighted unless otherwise specified.

Table A2. List of scales, source paper, and brief description.

Scale (number of items in parentheses)	Source	Description
Benevolent sexism (4)	Glick & Fiske (1996)	Positive but patronizing beliefs about women
Conflict avoidance (5)	Goldstein (1999)	Discomfort with political disagreement
Hostile sexism (4)	Glick & Fiske (1996)	Overtly negative stereotypes and antagonism toward women
Old-fashioned racism (2)	GSS	Explicit racist beliefs
Racial resentment (4)	Kinder & Sanders (1996)	Prejudice linking anti-Black affect to beliefs about work ethic and deservingness
Racial sympathy (4)	Chudy (2017)	Empathetic concern for Black Americans' disadvantage and desire to address inequality
Stereotype endorsement: Asian (5)	ANES	Traits ascribed to Asian Americans
Stereotype endorsement: Black (5)	ANES	Traits ascribed to Black Americans
Stereotype endorsement: Hispanic (5)	ANES	Traits ascribed to Hispanic Americans
Stereotype endorsement: Men (5)	ANES	Traits ascribed to men
Stereotype endorsement: Illegal/Undoc. Immig. (5)[19]	ANES	Traits ascribed to undocumented immigrants
Stereotype endorsement: White (5)	ANES	Traits ascribed to White Americans
Stereotype endorsement: Women (5)	ANES	Traits ascribed to women
Need for cognition (18)	Cacioppo et al. (1984)	Tendency to enjoy and engage in effortful thought
Zero sum (4)	Chinoy et al. (2024)	Belief that society is zero-sum (gains for some necessarily mean losses for others)

[19] We collapse the five-item scales for "Stereotype Endorsement: Illegal immigrant" and "Stereotype Endorsement: Undocumented Immigrant" into one and randomize which wording respondents saw. Because this scale is rarely used, we omit it from our main analysis.

Anti-intellectual affect (5)	Barker et al. (2022)	Distrust or resentment toward intellectuals and experts
Conspiracy (5)	Bruder et al. (2013)	Endorsement of conspiracy thinking across domains
Distrustful complacency (3)	Adapted from Lalot et al. (2021)	Combination of distrust in politics and passive acceptance of status quo
Epistemic hubris (9)	Barker et al. (2022)	Overconfidence in one's knowledge and resistance to expert correction
Political intolerance (9)	Adapted from ALLBUS	Willingness to deny rights to disliked groups; standard intolerance battery
Right-wing populism (9)	Schulz et al. (2018)	Anti-elitism, people-centrism, and right-wing outgroups
System justification (8)	Kay & Jost (2003)	Motivation to defend and rationalize existing social and political arrangements
Big Five: Agreeableness (2)	Gosling et al. (2003)	TIPI Big Five short scale; measures cooperation, trust, and compassion
Big Five: Conscientiousness (2)	Gosling et al. (2003)	TIPI; measures organization, responsibility, and goal-directedness
Big Five: Extraversion (2)	Gosling et al. (2003)	TIPI; measures sociability and outgoingness
Big Five: Neuroticism (2)	Gosling et al. (2003)	TIPI; measures emotional stability versus negative affect
Big Five: Openness (2)	Gosling et al. (2003)	TIPI; measures creativity, curiosity, and preference for novelty
Dirty Dozen: Machiavellianism (4)	Jonason & Webster (2010)	Part of the Dirty Dozen short Dark Triad scale; manipulativeness and strategic exploitation
Dirty Dozen: Narcissism (4)	Jonason & Webster (2010)	Dirty Dozen; entitlement, self-centeredness, grandiosity
Dirty Dozen: Psychopathy (4)	Jonason & Webster (2010)	Dirty Dozen; callousness, lack of empathy, impulsivity
Empathy: empathic concern (7)	Davis (1983)	Other-oriented emotional responses

Empathy: perspective taking (7)	Davis (1983)	Tendency to appreciate others' viewpoints and motives
Need for chaos (9)	Peterson et al. (2023)	Desire to disrupt or destroy the existing social order, even at high costs
SDO: Dominance (8)	Ho et al. (2015)	Preference for group-based hierarchy through dominance
SDO: Equality (8)	Ho et al. (2015)	Opposition to group equality
Violence attitudes (2)	Uscinski & Parent (2014)	Support for use of violence in politics
Authoritarianism (8)	ANES	Child-rearing values battery; preference for obedience and conformity
Egalitarianism (4)	ANES	Belief in equality and government's role in promoting it
Moral traditionalism (2)	ANES	Preference for traditional moral and cultural norms

Table A3. Papers identified in literature sweep using scales as independent or moderating variables and whether they measure or control for other scales.

Paper	Scale(s)	Use(s)	Number of other scales measured/ controlled for
Ditonto 2020	Racial resentment	IV	0
Choma et al. 2020	Social dominance orientation	IV	0
Jedinger & Burger 2020	Social dominance orientation	IV	2 (Right-wing authoritarianism; Anti-egalitarianism)
Simas et al. 2020	Empathy: empathic concern	IV	3 (Empathy: perspective taking; Empathy: personal distress; Empathy: fantasy)
DeSante & Watts Smith 2020	Racial resentment	IV, moderator	0
Ellis & Faricy 2020	Racial resentment	IV	2 (Egalitarianism; Trust in government)
Karpowitz et al. 2021	Racial resentment	IV, moderator	0
Chudy 2021	Racial sympathy	IV, moderator	IV: 2 (Racial resentment; Limited government) Moderator: 2 (Racial resentment; Limited government)
Bakker et al. 2021	Big Five	IV, moderator	IV: 4 (Other Big Five dimensions) Moderator: 1 (Authoritarianism)
Luttig 2021	Authoritarianism	IV	0
Peyton & Huber 2021	Racial resentment; Stereotype Endorsement: Black	Moderator	Racial resentment: 1 (Stereotype Endorsement: Black) Stereotype Endorsement: Black: 1 (Racial resentment)
Gross & Wronski 2021	Racial resentment	IV	0

Ksiazkiewicz & Friesen 2021	Big Five	IV	4 (Other Big Five dimensions)
Hopkins 2021	Stereotype Endorsement: Black; Stereotype Endorsement: Hispanic; Stereotype Endorsement: White	IV	0
Jardina 2021	Racial resentment	IV	2 (White identity; Stereotype Endorsement: Black)
Pereira 2021	Benevolent sexism; Hostile sexism	Moderator	Benevolent sexism: 0 Hostile sexism: 0
Barker et al. 2022	Anti-intellectual affect	IV, moderator	IV: 4 (Educational elitism; Christian traditionalism; Generic populism; Authoritarianism) Moderator: 1 (Intellectual identity)
Bonilla et al. 2022	Racial resentment	Moderator	3 (Muslim American resentment; Linked fate; Group solidarity)
Haselwerdt 2022	Racial resentment	Moderator	1 (Anti-immigration)
Enders & Thornton 2022	Racial resentment	Moderator	0
Filindra et al. 2022	Stereotype Endorsement: Black	IV	2 (Authoritarianism; Trust in people)
Cepuran & Berry 2022	Racial resentment	IV	1 (Relative discrimination)
Munis 2022	Racial resentment	IV	2 (Place identity, populism)
Westwood & Peterson 2022	Racial resentment	Moderator	0
Banda & Cassese 2022	Hostile sexism; Racial resentment	IV	0
Prati et al. 2022	Social dominance orientation	IV	3 (political fairness, political self efficacy, perceived corruption)

Beattie et al. 2022	Social dominance orientation; Need for cognition	IV	29 (Schwartz values (10, see table note), Big Five (5), System justification, Cognitive reflection, Dogmatism, Personal need for security, Need for cognition, Cognitive reflection, Bullshit receptivity, Authoritarianism, Disgust sensitivity, Death anxiety, Threat sensitivity, Avoidant attachment, anxious attachment, Need for closure)
Pantazi et al. 2022	Conspiracy	IV	3 (Political cynicism, political powerlessness, belief in simple solutions)
Osborne et al. 2022	Hostile sexism; Benevolent sexism; Big Five	IV	0
Miranda 2022	Big Five	IV	2 (Interpersonal trust, perceived corruption)
van Prooijen et al. 2022	Conspiracy	IV	3 (Populism, Need for cognition, faith in intuition)
Stephens-Dougan 2023	Stereotype endorsement: Black; Stereotype endorsement: white	Moderator	0
Petersen et al. 2023	Need for chaos	IV	0
Oceno et al. 2023	Hostile sexism	IV	3 (Authoritarianism, ethnocentrism, racial resentment)
Kevins & Vis 2023	Hostile sexism	Moderator	0
Thompson & Busby 2023	Racial resentment	Moderator	0

Masuoka et al. 2023	Hostile sexism	Moderator	0
Heide-Jørgensen et al. 2023	Big Five: Openness	Moderator	4: (Big Five: Agreeableness; Big Five: Conscientiousness; Big Five: Extraversion; Big Five: Neuroticism)
Dynes et al. 2023	Big Five: Openness; Big Five: Agreeableness	IV	3 (Big Five: Conscientiousness; Big Five: Extraversion; Big Five: Neuroticism)
Amsalem & Sheffer 2023	Big Five	IV	0
Cawvey 2023	Big Five	IV	2 (Self efficacy, corruption)
Georgarakis 2023	Authoritarianism	Moderator	2 (Racial resentment, moral traditionalism)
Baron et al. 2023	Social dominance orientation	IV	4 (Authoritarianism, Egalitarianism, Nationalism, Populism)
Ollerenshaw 2024	Authoritarianism	IV, moderator	0
Helminen et al. 2024	Authoritarianism	IV	0
Pizarro et al. 2024	Social dominance orientation	Moderator	2 (Risk perception, Right wing authoritarianism)
Mußotter 2024	SDO	IV	0
Jung & Clifford 2025	Moral traditionalism	IV	21 (Schwartz values (10, see table note), Care, Fairness, Loyalty, Authority, Sanctity, Equality, Humanitarianism, Moral Traditionalism, Moral Tolerance, Individualism, Limited Governance)

Vasilopoulos & Robinson 2025	Authoritarianism	IV	0
Strawbridge et al. 2025	Moral traditionalism	IV	1 (Egalitarianism
Goldsmith & Moen 2025	Big Five	IV	9 (Other Big Five dimensions, Conservatism, Right-wing authoritarianism, Social dominance orientation, White identity, Populist voter)

Notes: Beatie et al. (2022) and Jung and Clifford (2025) control for "Schwartz values," which include the following items: Self-direction, Hedonism, Benevolence, Power, Universalism, Achievement, Security, Stimulation, Conformity, and Tradition.

Table A4. Unrotated factor loadings for exploratory factor analysis.

Scale	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Benevolent sexism	0.2939	-0.0926	-0.0272	0.4851	0.0901
Conflict avoidance	-0.2938	0.2909	0.1353	-0.1007	0.3563
Hostile sexism	0.6404	0.3590	0.1164	0.0228	-0.0763
Old Fashioned racism	0.4276	0.1536	-0.0435	-0.1928	0.0627
Racial resentment	0.6402	0.5160	0.0158	-0.0838	-0.0252
Racial sympathy	-0.5861	-0.1865	0.0232	0.2556	0.0964
Net anti-Asian (vs White) stereotype	0.3607	0.0380	-0.4905	-0.0041	0.3687
Net anti-Black (vs White) stereotype	0.4756	0.2370	-0.3252	-0.1119	0.3693
Net anti-Hispanic (vs White) stereotype	0.4512	0.1324	-0.4401	-0.0304	0.4339
Net anti-Woman (vs Men) stereotype	0.3134	-0.0002	-0.1913	-0.0736	0.0135
Need for thinking	-0.3053	0.1269	-0.1933	0.0155	-0.4028
Zero sum	0.2156	-0.2124	0.2206	0.3441	0.2733
Anti-intellectual affect	0.6142	0.3501	0.2816	0.2329	-0.0183
Conspiracy	0.4272	0.2456	0.3659	0.3698	0.0523
Distrustful complacency	-0.0294	0.2375	0.5763	-0.0915	0.0550
Epistemic hubris	0.0355	0.3862	0.1456	0.0824	-0.1435
Political intolerance	0.5935	0.4986	0.1211	-0.0492	-0.0143
Right wing populism	-0.0739	0.3584	0.4534	0.1858	0.1330
System justification	0.4076	0.0626	-0.4996	0.1426	-0.1447
Big 5 Agreeableness	-0.4018	0.4515	-0.1453	0.3092	0.0479
Big 5 Conscientiousness	-0.1887	0.4949	-0.1596	0.3082	-0.0617
Big 5 Extroversion	0.0667	-0.0434	-0.2013	0.1762	-0.2311
Big 5 Neuroticism	0.0947	-0.4013	0.2953	-0.2952	0.2168

Big 5 Openness	-0.2660	0.0752	-0.1868	0.3520	-0.2518
Dirty dozen Machiavellianism	0.4480	-0.6123	0.0530	0.0417	-0.0998
Dirty dozen Narcissism	0.3741	-0.5357	-0.1430	0.2320	-0.1161
Dirty dozen Psychopathy	0.5604	-0.5616	0.1081	-0.0573	-0.0773
Empathy Empathic Concern	-0.5597	0.2787	-0.0570	0.1260	0.0923
Empathy Perspective Taking	-0.4526	0.2514	-0.0933	0.0024	-0.0518
Need for chaos	0.5032	-0.5126	0.1613	0.1249	0.0071
SDO dominance	0.7456	-0.2840	-0.0259	0.2342	0.0181
SDO equality	-0.6201	-0.2898	0.0192	0.3508	0.2716
Violence attitudes	0.4528	-0.4258	0.0998	0.1751	-0.0356
Authoritarianism	0.4258	0.2123	-0.0528	0.2513	0.1031
Egalitarianism	-0.7519	-0.2511	0.0473	0.1545	0.2462
Moral traditionalism	0.4364	0.5792	0.0728	-0.0005	-0.0938

Figure A1. Scree plot for scale-level factor analysis.

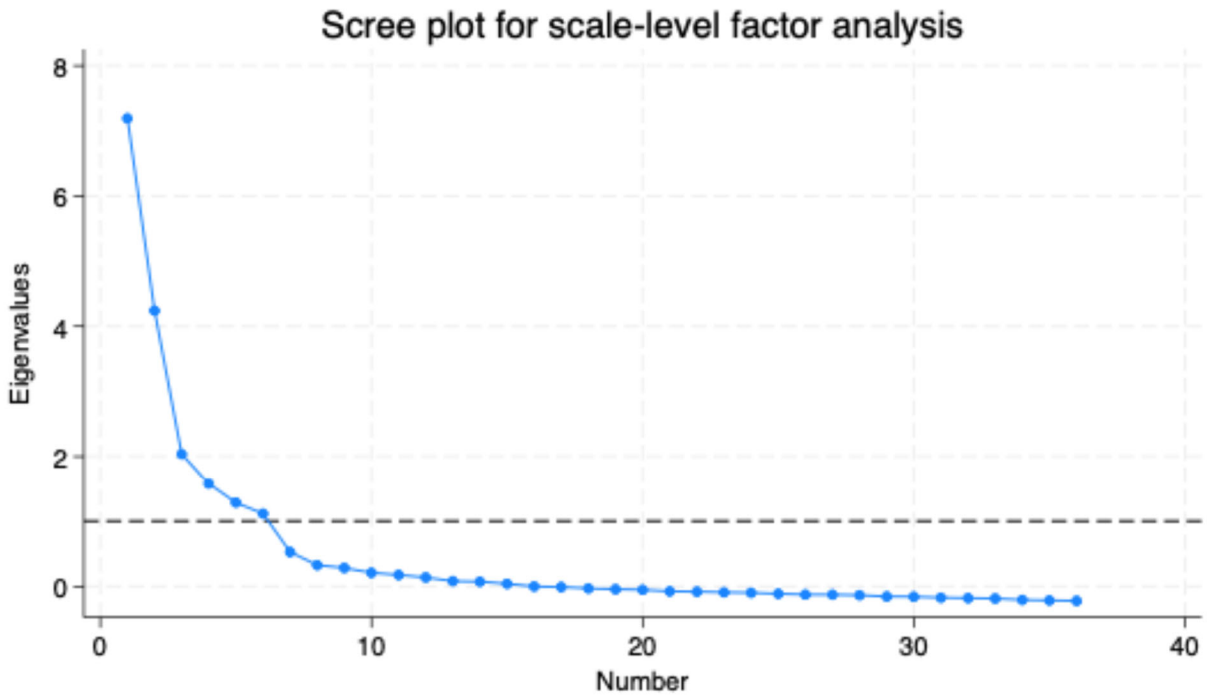
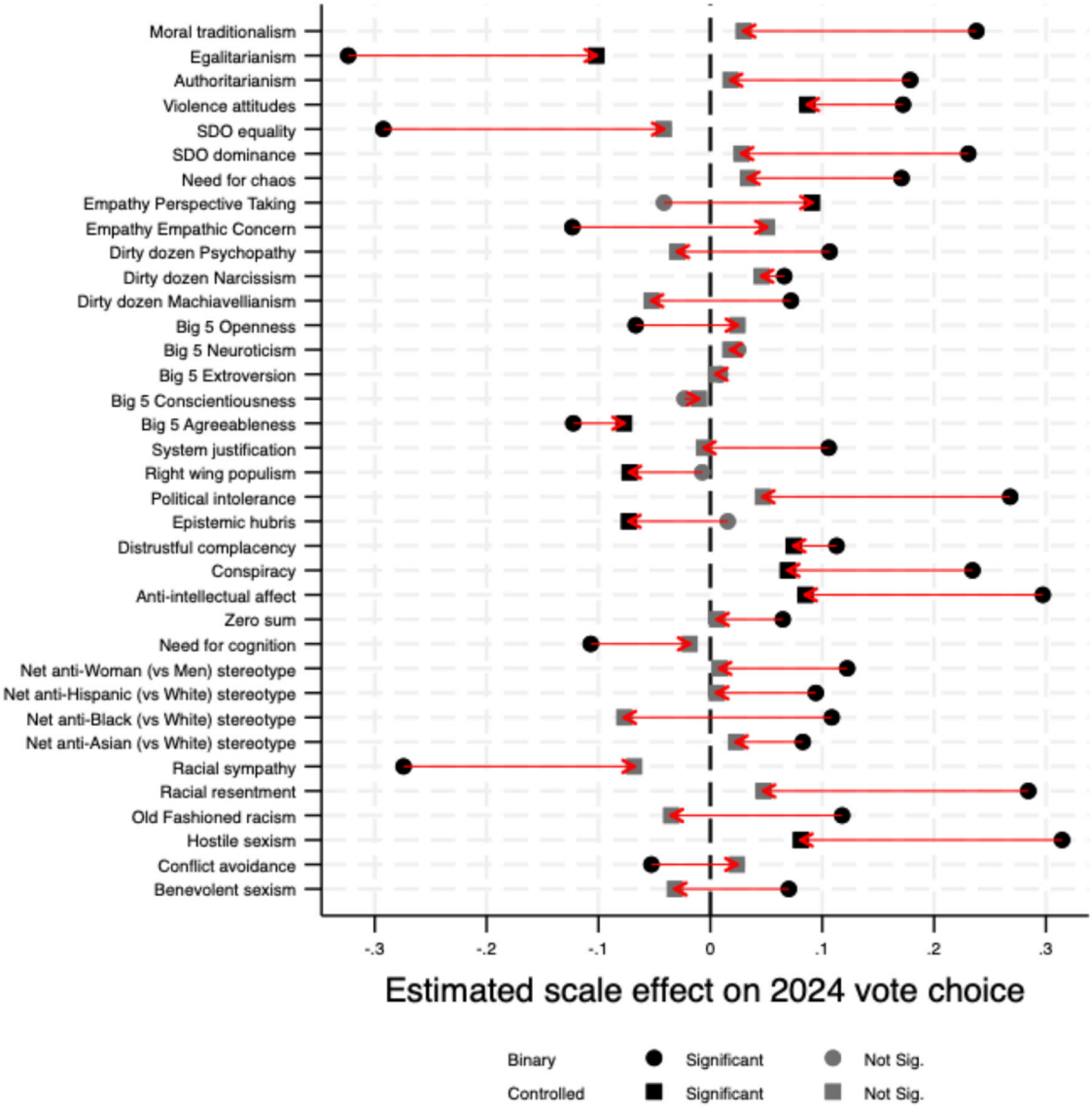


Figure A2. Estimated scale effect on 2024 vote choice.



Notes: Dependent variable is vote choice (where 1=Trump, 0=Biden, and 0.5=other). Controlled model adds all other scales. Among estimates significant in binary regression, controlled estimate is on average 0.31 of binary estimate.

Figure A3. Scree plot for item-level factor analysis.

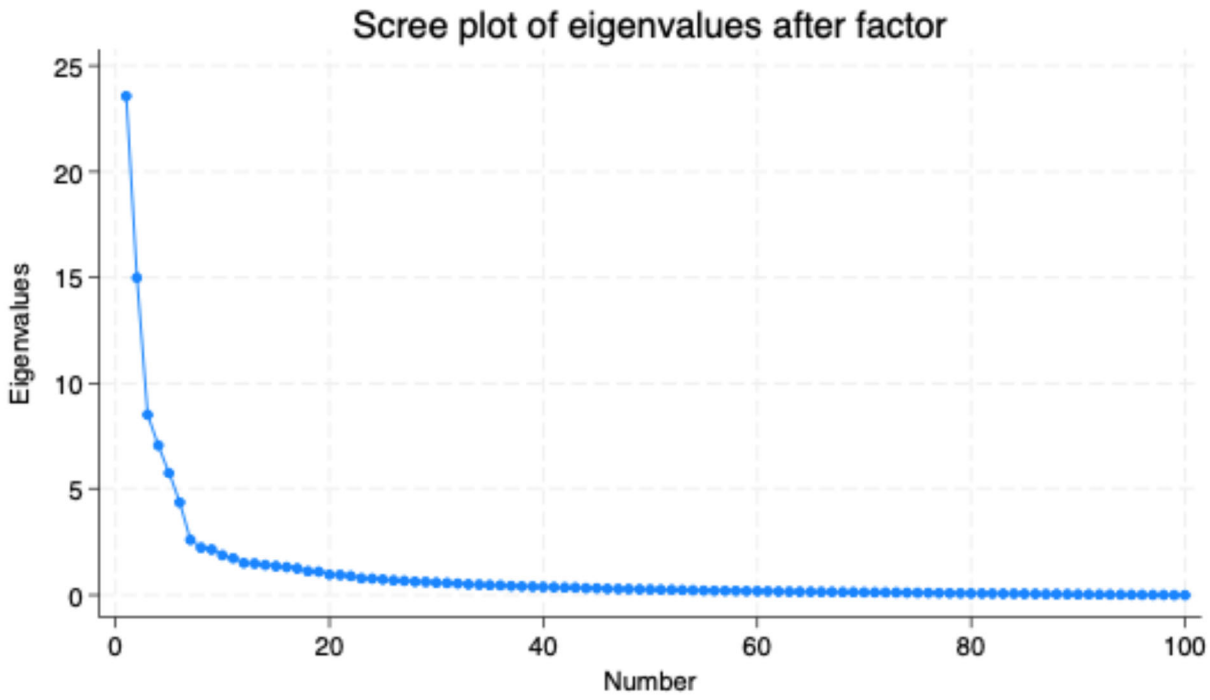


Table A5. Highest loading items for six factors (scales) from item-level factor analysis using orthogonal rotation.

Question wording	Source scale	Factor score
Factor 1		
I tend to seek prestige or status.	Dirty Dozen: Narcissism	0.733
I tend to expect special favors from others	Dirty Dozen: Narcissism	0.703
Sometimes I just feel like destroying beautiful things.	Need for chaos	0.703
Superior groups should dominate inferior groups.	Social dominance orientation: Dominance	0.69
I get a kick when natural disasters strike in foreign countries.	Need for chaos	0.683
In getting what your group wants, it is sometimes necessary to use force against other groups.	Social dominance orientation: Dominance	0.652
Factor 2		
We should strive to make incomes more equal.	Social dominance orientation: Equality	-0.804
We should do what we can to equalize conditions for different groups.	Social dominance orientation: Equality	-0.789
Increased social equality is beneficial to society.	Social dominance orientation: Equality	-0.764
Group equality should be our ideal.	Social dominance orientation: Equality	-0.733
We would have fewer problems if we treated different groups more equally.	Social dominance orientation: Equality	-0.729
If people were treated more equally in this country we would have many fewer problems.	Egalitarianism	-0.694
Factor 3		
I think that politicians usually do not tell us the true motives for their decisions.	Conspiracy	0.639
Politicians are not really interested in what people like me think.	Right-wing populism	0.617
Politicians are mainly in politics for their own benefit and not for the benefit of the community.	Distrustful complacency	0.611
Politicians in Washington very quickly lose touch with ordinary people.	Right-wing populism	0.579
Politicians talk too much and take too little action.	Right-wing populism	0.554
Most members of Congress are honest. [Reverse-coded]	Distrustful complacency	0.548
Factor 4		

This country would have many fewer problems if there were more emphasis on traditional family ties.	Moral traditionalism	0.494
Society is set up so that people usually get what they deserve.	System justification	0.474
Everyone has a fair shot at wealth and happiness.	System justification	0.468
The United States is the best country in the world to live in.	System justification	0.457
Good manners [is more important for a child to have than curiosity].	Authoritarianism	0.449
Most policies serve the greater good.	System justification	0.427
Factor 5 ($\alpha = 0.79$)		
Hispanic Americans – Intelligent to Unintelligent	Stereotype Endorsement: Hispanic	0.703
Hispanic Americans – Compassionate to Insensitive	Stereotype Endorsement: Hispanic	0.646
Black Americans – Intelligent to Unintelligent	Stereotype Endorsement: Black	0.638
Hispanic Americans – Hard-working to Lazy	Stereotype Endorsement: Hispanic	0.617
Asian Americans – Compassionate to Insensitive	Stereotype Endorsement: Asian	0.613
Hispanic Americans – Honest to Dishonest	Stereotype Endorsement: Hispanic	0.611
Factor 6 ($\alpha = 0.85$)		
I really enjoy a task that involves coming up with new solutions to problems.	Need for cognition	0.681
I like to have the responsibility of handling a situation that requires a lot of thinking.	Need for cognition	0.661
The idea of relying on thought to make my way to the top appeals to me.	Need for cognition	0.614
The notion of thinking abstractly is appealing to me.	Need for cognition	0.613
I would prefer a task that is intellectual, difficult, and important to one that is somewhat but does not require much thought.	Need for cognition	0.558
I prefer my life to filled with puzzles that I must solve.	Need for cognition	0.541

Notes: Items sorted by descending absolute magnitude of factor loading for each factor.

Table A6. Complete loadings for first six dimensions of exploratory analysis.

Scale	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
sc1_need_for_chaos_1_rc	0.4725	-0.5006				
sc1_need_for_chaos_2_rc	0.3983	-0.4242				
sc1_need_for_chaos_3_rc	0.4305	-0.3171				
sc1_need_for_chaos_4_rc	0.3980			0.3087		
sc1_need_for_chaos_5_rc	0.3710			0.3335		
sc1_need_for_chaos_6_rc	0.3851	-0.5146				
sc1_need_for_chaos_7_rc	0.4630	-0.5359				
sc1_need_for_chaos_8_rc		-0.3082	-0.3381			
sc1_need_for_chaos_9_rc			-0.3313			
sc1_distrustful_complacency_2_rc				0.4781		
sc1_distrustful_complacency_3_rc		0.3613		0.4406		
sc1_distrustful_complacency_4_rc				0.3179		
sc1_benevolent_sexism_2_rc	0.3646					
sc1_benevolent_sexism_4_rc		-0.3199				
sc1_hostile_sexism_1_rc	0.6325					
sc1_hostile_sexism_3_rc	0.5137					
sc1_hostile_sexism_2_rc	0.4103	0.3904				
sc1_hostile_sexism_4_rc	0.3417	0.4424				
sc1_need_for_thinking_1_rc					0.4084	
sc1_need_for_thinking_2_rc					0.5190	
sc1_need_for_thinking_6_rc					0.4442	
sc1_need_for_thinking_10_rc					0.4894	
sc1_need_for_thinking_11_rc			0.3689		0.5024	

sc1_need_for_thinking_13_rc				0.4437
sc1_need_for_thinking_14_rc				0.5299
sc1_need_for_thinking_15_rc				0.4416
sc1_need_for_thinking_18_rc				0.4018
sc1_need_for_thinking_3_rc	-0.3864	0.3140		0.4119
sc1_need_for_thinking_4_rc	-0.3819	0.3682		0.3816
sc1_need_for_thinking_5_rc	-0.3966	0.3622		0.3562
sc1_need_for_thinking_7_rc	-0.3741	0.3388		0.3430
sc1_need_for_thinking_9_rc				0.3124
sc1_need_for_thinking_12_rc	-0.4340			0.3073
sc1_need_for_thinking_17_rc	-0.3076			
sc1_sdo_dominance_1_rc	0.5917			
sc1_sdo_dominance_2_rc	0.5452	-0.3592		
sc1_sdo_dominance_3_rc	0.6245	-0.3560		
sc1_sdo_dominance_4_rc	0.5613	-0.3282		
sc1_sdo_dominance_5_rc	0.6906			
sc1_sdo_dominance_6_rc	0.6844			
sc1_sdo_dominance_7_rc	0.6910			
sc1_sdo_dominance_8_rc	0.6726			
sc1_sdo_equality_1_rc	-0.4480		0.3042	
sc1_sdo_equality_2_rc	-0.4125	-0.4260		
sc1_sdo_equality_3_rc	-0.4433			
sc1_sdo_equality_4_rc	-0.4886	-0.4656		
sc1_sdo_equality_5_rc	-0.5405	-0.4156		
sc1_sdo_equality_6_rc	-0.5194	-0.3671		

sc1_sdo_equality_7_rc	-0.4589	-0.4884	0.3457	
sc1_sdo_equality_8_rc	-0.4886			
sc1_racial_resentment_1_rc	0.5876	0.3806		
sc1_racial_resentment_2_rc	0.4491	0.5841		
sc1_racial_resentment_3_rc	0.4058	0.5675		
sc1_racial_resentment_4_rc	0.6462			
sc1_system_justification_1_rc	0.3075		-0.4170	
sc1_system_justification_2_rc		-0.3566	-0.3984	
sc1_system_justification_6_rc	0.5282			
sc1_system_justification_8_rc	0.5043			
sc1_system_justification_3_rc		0.3232	-0.4585	
sc1_system_justification_7_rc		-0.3160	-0.3883	
sc1_zero_sum_1_rc		-0.3971		
sc1_zero_sum_3_rc	0.5573			
sc1_zero_sum_4_rc		-0.3365	0.3749	
sc1_violence_attitudes_1_rc	0.4149	-0.3522		
sc1_violence_attitudes_2_rc	0.4503	-0.4231		
sc1_authoritarianism_3_rc	0.3356			-0.3050
sc1_authoritarianism_4_rc	0.3728			
sc1_authoritarianism_5_rc	0.3346			
sc1_authoritarianism_7_rc	0.3105			
sc1_epistemic_hubris_1_rc		0.4351		
sc1_epistemic_hubris_5_rc		0.3829		
sc1_racial_sympathy_1_rc	-0.5080			
sc1_racial_sympathy_2_rc	-0.5018			

sc1_racial_sympathy_3_rc	-0.3768			
sc1_racial_sympathy_4_rc	-0.4145	-0.3197		
sc1_right_wing_populism_1_rc		0.4212		0.3880
sc1_right_wing_populism_2_rc				0.3253
sc1_right_wing_populism_3_rc				0.4074
sc1_right_wing_populism_4_rc		0.3134		0.4662
sc1_right_wing_populism_5_rc		0.3667		0.3911
sc1_right_wing_populism_6_rc			0.3498	
sc1_right_wing_populism_7_rc			0.3733	
sc1_right_wing_populism_8_rc			0.3246	
sc1_right_wing_populism_9_rc			0.3409	
sc1_dirt_doz_narcissism_1_rc		-0.3646		
sc1_dirt_doz_narcissism_2_rc		-0.3906		
sc1_dirt_doz_narcissism_3_rc	0.4535	-0.5484		
sc1_dirt_doz_narcissism_4_rc	0.3891	-0.5019		
sc1_dirt_doz_psychopathy_1_rc	0.5254	-0.3787		
sc1_dirt_doz_psychopathy_2_rc	0.4944	-0.3526		
sc1_dirt_doz_psychopathy_3_rc	0.4240	-0.4587		
sc1_dirt_doz_machiavelli_1_rc		-0.3787		
sc1_dirt_doz_machiavelli_2_rc	0.4181	-0.4619		
sc1_dirt_doz_machiavelli_3_rc		-0.3575		
sc1_dirt_doz_machiavelli_4_rc	0.5095	-0.5185		
sc1_conspiracy_1_rc				0.3198
sc1_conspiracy_2_rc		0.3050		0.4245
sc1_conspiracy_3_rc	0.4614			

sc1_conspiracy_4_rc	0.4992		0.3264		
sc1_conspiracy_5_rc	0.4267		0.3244	0.3066	
sc1_egalitarianism_1_rc	-0.4449				
sc1_egalitarianism_4_rc	-0.5011	-0.3487			
sc1_egalitarianism_2_rc	-0.6449				
sc1_egalitarianism_3_rc	-0.5814				
sc1_anti_intellectual_1_rc	0.3126		0.3362		
sc1_anti_intellectual_2_rc	0.5323				
sc1_anti_intellectual_3_rc	0.4718	0.3341			
sc1_anti_intellectual_4_rc	0.6036	0.3806			
sc1_anti_intellectual_5_rc	0.4099	0.3017			
sc1_empathy_perspective_1_rc	-0.3395				
sc1_empathy_perspective_4_rc	-0.4036				
sc1_empathy_empathic_concer_2_rc	-0.3554				
sc1_empathy_empathic_concer_4_rc	-0.5025				
sc1_empathy_empathic_concer_5_rc	-0.5075				
sc1_conflict_avoidance_2_rc	-0.3356	0.3355			-0.3230
sc1_conflict_avoidance_4_rc					-0.4821
sc1_stereot_endorse_white_1_rc			-0.3968		
sc1_stereot_endorse_white_2_rc			-0.3094	0.3178	
sc1_stereot_endorse_white_3_rc			-0.4902		
sc1_stereot_endorse_white_4_rc			-0.3432	0.3382	
sc1_stereot_endorse_white_5_rc			-0.3785	0.3195	

sc1_stereot_endorse_black_1_rc	0.3853		-0.3591	0.3294
sc1_stereot_endorse_black_2_rc	0.3076	0.3029		
sc1_stereot_endorse_black_3_rc	0.3745		-0.4321	0.3311
sc1_stereot_endorse_black_4_rc	0.3225		-0.3264	
sc1_stereot_endorse_black_5_rc	0.4363		-0.3622	0.3144
sc1_stereot_endorse_hispan_1_rc	0.3685		-0.3763	0.3938
sc1_stereot_endorse_hispan_2_rc			-0.3166	0.3171
sc1_stereot_endorse_hispan_3_rc	0.3728		-0.4205	0.4234
sc1_stereot_endorse_hispan_4_rc			-0.3528	0.3591
sc1_stereot_endorse_hispan_5_rc	0.3550		-0.3924	0.3717
sc1_stereot_endorse_asian_1_rc			-0.3692	0.4161
sc1_stereot_endorse_asian_2_rc			-0.3564	0.3417
sc1_stereot_endorse_asian_3_rc			-0.3799	0.4052
sc1_stereot_endorse_asian_4_rc			-0.3907	0.3365
sc1_stereot_endorse_asian_5_rc			-0.3906	0.3560
sc1_political_intolerance_1_rc	0.4627	0.3686		
sc1_political_intolerance_2_rc	0.6411			
sc1_political_intolerance_3_rc	0.4979	0.3335		
sc1_political_intolerance_4_rc	0.5625			
sc1_political_intolerance_8_rc	0.6392			
sc1_political_intolerance_5_rc	0.3498	0.4680		
sc1_political_intolerance_6_rc	0.3078	0.6489		
sc1_political_intolerance_7_rc		0.6439		

sc1_political_intolerance_9_rc		0.6260			
sc1_moral_traditionalism_1_rc		0.5474			
sc1_moral_traditionalism_2_rc	0.4922	0.4171	0.3112		
sc1_old_fashioned_racism_1_rc	0.3320				
sc1_old_fashioned_racism_2_rc	0.3355				
sc1_big_5_openness_rc			0.3099		0.3044
sc1_big_5_agreeableness_re_rc		0.3308			
sc1_big_5_agreeableness_rc	-0.3414		0.3844		
sc1_big_5_conscientiousnes_rc		0.3277			
sc1_big_5_conscientiousness_rc			0.3758		
sc1_big_5_neuroticism_reve_rc			-0.4139		
sc1_big_5_neuroticism_rc				0.3249	

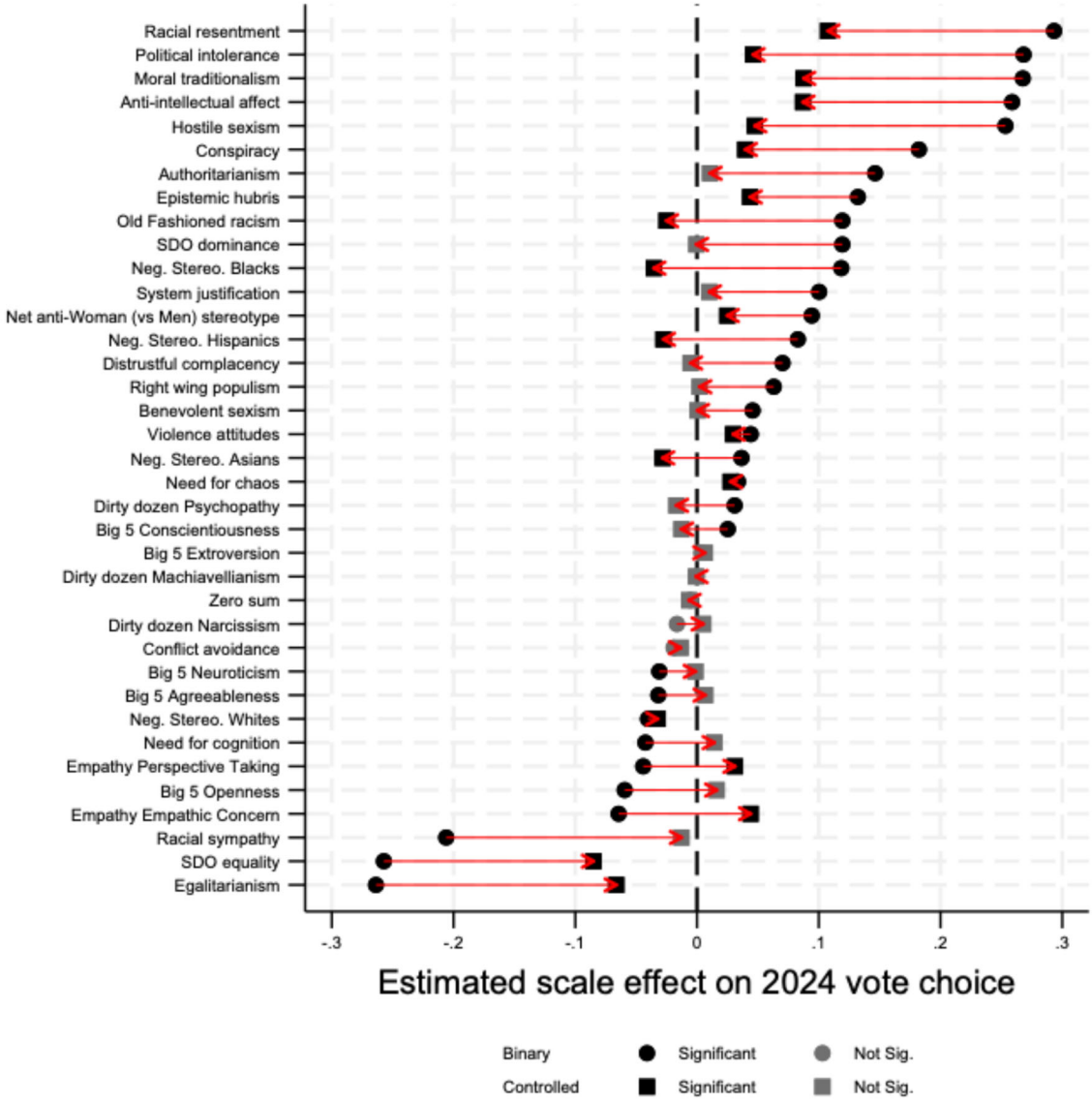
Table A7. Regression of political outcomes on standardized mean scores of six factors (scales) using orthogonal rotation.

	(1)	(2)	(3)	(4)	(5)	(6)
	Vote intention (0=Biden, 1=Trump, 0.5=Other)		7-point party ID (1=Strong D, 7=Strong R, 4=Ind./DK)		5-point ideological ID (1=Very lib., 5=Very conserv., 3=Moderate/DK)	
Factor 1	-0.00537	-0.0144	-0.0519	-0.0768	-0.0877***	-0.0662*
	(0.0114)	(0.0131)	(0.0534)	(0.0630)	(0.0245)	(0.0300)
Factor 2	0.207***	0.207***	0.865***	0.836***	0.411***	0.405***
	(0.00948)	(0.00973)	(0.0462)	(0.0472)	(0.0236)	(0.0241)
Factor 3	0.0778***	0.0739***	0.324***	0.311***	0.148***	0.146***
	(0.00985)	(0.00969)	(0.0451)	(0.0448)	(0.0213)	(0.0215)
Factor 4	0.126***	0.125***	0.601***	0.616***	0.436***	0.422***
	(0.00996)	(0.0103)	(0.0467)	(0.0482)	(0.0225)	(0.0236)
Factor 5	0.00708	0.00292	-0.00196	-0.0352	-0.00860	-0.0185
	(0.00998)	(0.00992)	(0.0476)	(0.0471)	(0.0226)	(0.0229)
Factor 6	-0.00169	0.00657	-0.0406	-0.0255	-0.0587**	-0.0494*
	(0.00922)	(0.00956)	(0.0429)	(0.0453)	(0.0212)	(0.0230)
Constant	0.473***	0.675***	3.786***	4.583***	3.028***	3.077***
	(0.00916)	(0.0436)	(0.0406)	(0.186)	(0.0193)	(0.0886)
Controls	No	Yes	No	Yes	No	Yes
N	1588	1588	2001	2001	2001	2001

Notes: OLS coefficients with robust standard errors in parentheses. Columns 2, 4, and 6 use controls for age, gender, income, race, and education. Responses in all columns collected for all panelists between December 11, 2023–February 29, 2024.

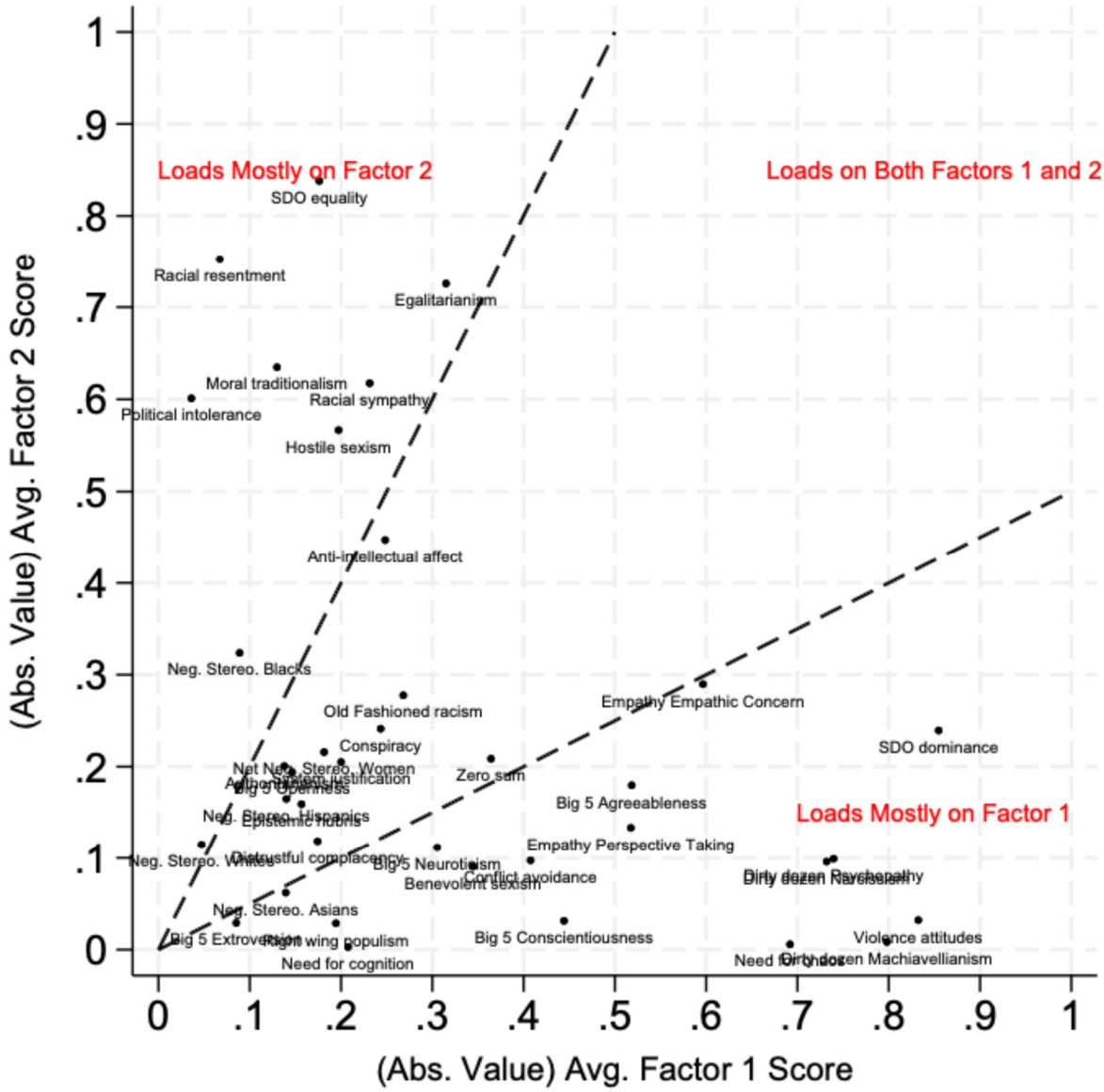
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figure A4. Estimated scale effect on 2024 vote choice excluding own-scale items from factors.



Notes: Dependent variable is vote choice (where 1=Trump, 0=Biden, and 0.5=other). Controlled model adds Factor 1 and Factor 2 constructed from all other scales' items, excluding the focal scale's items. Among estimates significant in binary regression, controlled estimate is on average 0.32 of binary estimate.

Figure A5. Relative loadings of each scale for two underlying factors using orthogonal rotation.



Notes: Absolute value of average of factor 1 and factor 2 scores across items used to construct each scale, with each factor score rescaled linearly to have a maximum (or minimum) of 1 (-1).