

Common Psychometric Scales Are Not Exogenous to Partisanship: Evidence from a Two-Wave Panel

Alan S. Gerber^{*1}, Gregory A. Huber², Mackenzie Lockhart³, Jack D. Walker II⁴

This Draft: 10 July 2026

Political science uses attitudinal and social-psychological scales like racial resentment, right-wing populism, and social dominance orientation to explain important outcomes. These analyses assume that measured scale scores are not caused by the same political forces that drive outcomes and that they are fixed over the short term. We test these assumptions for 36 commonly used scales using a nationally representative two-wave U.S. panel study fielded six months apart during the 2024 presidential campaign. We find that partisanship strongly predicts directional changes in most scales, with partisan gaps widening over the campaign. We also estimate the test-retest reliability (stability) of each scale against key benchmarks and find very different levels of stability. Our results raise important concerns about the exogeneity of these scales as implemented in surveys.

Word count: 6198 (inclusive of text, footnotes, references, and tables/figures)

*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 Opus 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 Jennifer Jerit, Cindy Kam, and Tali Mendelberg for helpful comments.

A growing body of political science and related research uses psychometric scales as independent (as well as moderating, mediating, and dependent) variables to explain political outcomes. The frequent use of these scales is a striking feature of contemporary observational and experimental analysis in peer-reviewed scholarship.¹ For example, Rush et al. (2025) find that attitudes about the January 6th insurrection are predicted by a racial antipathy scale. Similarly, Winter (2022) uses a conjoint experiment to show that the hostile sexism scale affects vote choice. And Simas et al. (2020) test the effect of the interaction between the empathic concern scale and partisanship on party affect.

For these estimates to be unbiased, measured scales must not be caused by the same political forces that affect outcomes. Thus, analyses using these scales often rest on an assumption (stated or otherwise) about their independence from short-term political influences and stability over time (see Alwin & Krosnick [1991] for foundational discussions of these issues). Importantly, if one uses a scale to predict an outcome like vote choice, for example, it should not be the case that any other factor that also affects vote choice, such as partisanship, also predicts changes in the measured scale.² If partisanship affects a scale score, and partisanship also affects the outcome of interest, then the scale-outcome correlation is confounded. Moreover, even controlling for contemporaneous partisanship still risks bias if the functional form of this relationship is not correctly specified or in the presence of random measurement error. In light of this concern, stability of measured scale scores over time is often treated as a proxy for this independence: If a measured scale is

¹ See Appendix Table A.1 for a recent use case in the political science literature for each of the 36 scales we study.

² Similar concerns apply if a scale is used to explain treatment heterogeneity across respondents.

stable, it reduces concerns that the predictive power of a scale is due to some other factor that predicts both the contemporaneous scale score and the outcome of interest.³

Unfortunately, we lack systematic and comparative evidence about the measured independence from political influence and stability of most widely used psychological scales. Some of the scales we examine here have been used in surveys for decades without clear benchmarks. Researchers often interpret psychometric scales as measuring enduring traits, but if measured scale scores fluctuate systematically over time, it raises questions about interpretation and risks bias in causal inference. Scales aim to measure underlying traits, but instability in measured scores may reflect genuine trait change, random or systematic measurement error, or both.

A growing body of longitudinal work has begun to challenge the assumption that specific psychological dispositions are temporally stable and causally prior to political attitudes (e.g., Boston et al. 2018 [Big Five]; Bakker et al. 2021 [Big Five, authoritarianism]; Luttig 2021 [authoritarianism]; Vasilopoulos & Robinson 2025 [authoritarianism]), yet these papers examine only a small number of scales. This work grapples with the fact that changes in measured traits may reflect changes in those traits and/or measurement issues. But, for researchers who use additive scale scores constructed from cross-sectional survey data, like most political science research, issues of endogeneity and measurement are intertwined because measured scales are affected by both factors.

Our primary analysis tests whether there are systematic directional changes in 36 measured scales over the course of the 2024 United States presidential campaign using panel

³ Note, however, that even a completely stable measured scale score might still be endogenous to prior political influences. We do not take up that critique here, but this concern is raised in other work (e.g., Osborne & Sibley 2020).

data from a 1,400-person two-wave survey conducted six months apart. We deliberately focus on both changes in measured scale scores over time and test-retest correlations of summed scale scores (stability), rather than estimates of the latent traits they measure, because nearly all political science work uses additive scales calculated from cross-sectional data. Our survey is unique both in the variety of scales measured in the same sample and the length of time between waves, allowing us to examine many scales using the same time period and population. The pattern of changes in psychological scale scores helps us understand the nature of these scales as implemented: If changes can be predicted by respondent characteristics, especially political ones, it suggests these scales as used respond to political context and do not measure perfectly stable traits. In that case, cross-sectional estimates of the correlation between a scale and an outcome (or similar heterogeneous effects analysis) are vulnerable to concerns about omitted-variable bias stemming from factors that predict both the outcome and the measured scale. Contrary to the key assumption that measured scales are independent of political influences, we find that partisanship strongly predicts directional changes over the course of the 2024 presidential campaign for the majority of the scales we measure, with scale scores becoming more aligned with pre-existing partisanship. The magnitudes of these changes are large for some scales.

This finding demonstrates that concerns about endogeneity are not purely hypothetical: Measured changes in commonly used scales appear to reflect partisan dynamics (partisanship itself or any factor correlated with partisanship), meaning these measures do not appear independent of party identification.⁴ This calls into question the causal interpretation of analysis using these scales, an increasingly common practice in the literature. We note two

⁴ We note that we cannot empirically test whether wave 1 scale scores are already similarly contaminated.

mechanisms that may explain this pattern. First, scale items may capture elements of contemporary political divisions that become more salient during campaigns, making partisan considerations more accessible when respondents answer. Second, parties may send signals about positions related to scale content, leading partisans to adopt or intensify views consistent with their party's messaging (Lenz 2012; Engelhardt 2021). Although we cannot distinguish between these mechanisms with our data, both would suggest that measured scale scores are influenced by prior views and the political environment in which they are collected.

Additionally, we calculate the test-retest reliability (stability) of each scale we study compared to key benchmarks (partisanship and ideology) and examine which factors predict the degree of variability in a respondent's measured scale scores over time. We show that scales exhibit vastly different levels of stability over time: some are almost as stable as partisanship while others fluctuate greatly. In general, key demographic and political variables do little to predict the magnitude of absolute variability in scale scores over time.

These results indicate that even scales with high test-retest reliability show systematic partisan-correlated directional shifts. Put simply, high stability and partisan endogeneity coexist. Whether these shifts reflect genuine attitude change, partisan-motivated responding (e.g., cheerleading), or changing interpretation of scale items, the implication for cross-sectional analysis is the same: Measured scale scores gathered during a political campaign cannot be treated as exogenous to the partisan forces that also drive the outcomes these scales are used to explain. Importantly, this paper does not resolve the direction of causality between partisanship and psychometric scale scores. But researchers who use cross-sectional scale scores (measured proximate to an election) to predict political outcomes bear the burden of articulating why those scores should be treated as causally prior to partisanship rather than assuming independence.

Data and Methods

We identified 39 of the most frequently used attitudinal and social-psychological scales in recent political science research. Appendix Table A.1 lists those scales, includes their original source, a recent use case in political science, and a brief description. We describe scale selection, scoring, construction, and question wording fully in the appendix (A and G). Note that for our analysis we replace negative stereotype endorsement measures with net stereotype endorsements to mirror actual usage in most political science work. For Asians, Blacks, and Hispanics, these measures are therefore negative stereotypes relative to negative stereotypes of Whites, while for women the comparison is to men. This leaves 36 scales plus our three benchmarks (below).

In addition to the batteries listed in Appendix Table A.1, we identified three-point and seven-point party identification (branched question format) and five-point ideology as useful benchmarks for assessing the comparative stability of these scales. Party identification is often theorized to be a psychological identity and has been found to be stable over time (e.g., Schickler & Green 1997). It thus serves as a valuable benchmark for a diverse set of scales. Scholars find ideology moderately stable but decidedly less so than party identification (Ansolabehere et al. 2008). Most of the other scales we study are constructed from a greater number of items, which may reduce random measurement error (although the branching format of the partisanship battery may enhance its stability (Krosnick & Berent 1993)).

For each scale, we construct an additive scale following original coding rules, normalize to a mean of 0 and standard deviation of 1, and restrict analysis to respondents who completed all items for a given scale in both waves. (This is how these scales are constructed in the work we reviewed, but we note our results are substantively the same without normalization.) Further details on scale construction appear in Appendix A. In Appendix F,

we also discuss results using factor analysis from wave 1 responses to construct one-dimensional factor scores for each scale in place of additive scales.

Panel data

Our core empirical goal is to estimate the changes in, and stability of, commonly measured psychometric scales over time. To do so, we use panel data to compare individuals' responses to each scale six months apart. Data were collected as part of the [removed for peer review] administered by YouGov. Approximately 1,400 respondents were interviewed at two points in time at least 172 days apart specifically to measure the psychometric scales used in this paper. Many scale items were reverse-coded.⁵ This sample is designed to be nationally representative using weights provided by YouGov. For ease of presentation, we present unweighted analysis in the main text with parallel weighted analysis in the appendix. More information about survey design and implementation is in the appendix.

Wave 1 was fielded February 2–15, 2024 ($n=2001$). Wave 2 was fielded August 5–27, 2024, to all individuals who completed wave 1 ($n=1381$). Sample demographics for those who completed both waves are listed in Appendix Table B.1. Approximately 69% of those respondents who completed wave 1 also completed wave 2. In Appendix Table B.2 we examine the bivariate relationship between selected demographics and whether a respondent completed the wave 2 survey and find some important differences. Specifically, young and Black respondents were much less likely to complete both waves.⁶

⁵ Reverse-coding protects against acquiescence bias. More generally, because of reverse-coding, inattention would tend to depress rather than inflate the partisan differences reported below by making respondents appear more moderate.

⁶ One might worry that differential attrition by age biases our findings, since older respondents are more likely to have crystallized partisanship. However, psychological traits

The timing of these survey waves was purposive. Many political science studies of elections measure these kinds of scales at least as proximate to the election as our wave 2 study (e.g., the 2024 pre-election wave of the CES was conducted in October). If wave 2 scale scores are systematically different from wave 1 scores it would provide evidence that the political environment surrounding an election affects those measured scales. Our wave 1 survey was therefore fielded early in 2024, prior to most key campaign events, while our wave 2 survey was conducted during the heat of the summer campaign and after Harris had replaced Biden as the presumptive Democratic nominee.

Results

Our core analysis assesses whether there was secular change in measured scale scores during this six-month window. The time between these survey waves included the core of the 2024 general election campaign and a change in the Democratic nominee for president. Given this context, we are interested in evaluating whether any of these scales exhibit systematic change over time.

Predictors of directional change

For each scale we estimate the following OLS model with robust standard errors:

$$\Delta_i = \beta_0 + \beta_1 scale1_i + \beta_2 age_i + \beta_3 educ_i + \beta_4 gender_i + \beta_5 race_i + \beta_6 pid_i + \beta_7 newsint_i + \varepsilon_i,$$

where Δ_i is the directional change in scale scores between wave 1 and wave 2. We include the measured wave 1 scale score to account for the possibility that people with lower (higher)

are also thought to crystallize with age, which would bias against finding partisan-correlated change in scores among an older sample. Our weighted analysis available in the appendix, which rebalances the sample on key demographics, yields substantively similar results.

scale scores can move more in the opposite direction over time.⁷ We also include the standard seven-point partisanship measure because we are interested in whether more Democratic (Republican) individuals have different changes in measured scale scores than opposing partisans.⁸

Results for the demographic and political variables that are statistically significant for each scale are listed in Table 1. We mark estimates that are statistically significant after a Bonferroni adjustment with an asterisk. We estimate 63 statistically significant coefficients (29% of the 216 estimated coefficients), of which 28 (13%) remain significant after the Bonferroni correction. Estimates are most frequently significant, even after adjustment, for three covariates: age, gender, and partisanship.

⁷ A cross-lagged panel model, regressing wave 2 scale score on wave 1 partisanship while controlling for wave 1 scale score, is algebraically similar to our change-score specification. Note also that a random-intercept cross-lagged panel model (RI-CLPM) requires three or more waves and is therefore inestimable with our data.

⁸ The remaining covariates are coded as follows: age (years), education (binary, college graduate or not), gender (binary, female identifying or not), race (binary, White or not), and political interest (binary, follow what is going on most of the time or not).

Table 1. Statistically significant coefficients from linear regression models that predict the directional change in each scale between surveys using demographic variables in a regression model with all covariates included simultaneously.

	Est. Age Effect	Est. Educ. Effect	Est. Gender Effect	Est. Race Effect	Est. PID7 Effect	Est. News Interest Effect
Benevolent sexism (4)		-0.091		-0.200*	0.022	
Conflict avoidance (5)	0.003		0.143*			
Hostile sexism (4)			-0.091		0.080*	
Old-fashioned racism (2)	0.003	-0.091	-0.151*		0.048*	
Racial resentment (4)	0.002				0.035*	-0.060
Racial sympathy (4)			0.136*		-0.052*	
Net SE: Anti-Asian (5)				0.123	0.037	
Net SE: Anti-Black (5)					0.063*	
Net SE: Anti-Hispanic (5)				0.191*	0.056*	
Net SE: Anti-women (5)			-0.154*	0.136	0.032	
Need for cognition (18)						0.101
Zero sum (4)		-0.129				
Anti-intellectual affect (5)		-0.069			0.060*	
Conspiracy (5)					0.046*	
Distrustful complacency (3)					0.060*	
Epistemic hubris (9)					0.052*	0.179*
Political intolerance (9)	0.002				0.062*	
Right-wing populism (9)					0.056*	
System justification (8)						
Big 5: Agreeableness (2)	0.005		0.101			
Big 5: Conscientiousness (2)					0.029	
Big 5: Extraversion (2)						
Big 5: Neuroticism (2)						
Big 5: Openness (2)	-0.003				-0.029	
Dirty Dozen:						
Machiavellianism (4)	-0.006*					
Dirty Dozen: Narcissism (4)	-0.007*		-0.084			
Dirty Dozen: Psychopathy (4)	-0.007*		-0.091			
Empathy: Empathetic concern (7)			0.106		-0.029	
Empathy: Perspective taking (7)					-0.034	
Need for chaos (9)	-0.006*					-0.142
SDO: Dominance (8)	-0.003				0.025	
SDO: Equality (8)			0.066		-0.063*	
Violence attitudes (2)	-0.009*			-0.111		
Authoritarianism (8)	0.003	-0.077			0.038*	
Egalitarianism (4)				-0.090	-0.087*	
Moral traditionalism (2)	0.003				0.097*	

Note: Number of scale items in parentheses. “SE” refers to stereotype endorsement. Listed coefficients are statistically significant at $p < 0.05$, those marked with an asterisk are significant after Bonferroni correction. Complete regression estimates are available upon request.

Column 1 shows that older people had more negative changes in measured negative personality traits (all three components of the dirty dozen, need for chaos, and violence attitudes) than younger people. Column 3 shows that women became more conflict avoidant, became more liberal on old-fashioned racism and racial sympathy, and expressed less anti-women negative stereotypes.

The greatest number of significant estimates are for partisanship (column 5), however. Across 36 scales, 24 (67%) of the partisanship estimates are significant, including 16 (44%) after our Bonferroni adjustment.⁹ Importantly, in all cases where these estimates are statistically significant, the estimated partisanship coefficient is in the direction of the bivariate relationship between partisanship and the same scale measured at wave 1, meaning that Republicans and Democrats diverge more over time on many of the same scales that they were divided on at the beginning of the campaign. For example, Republicans are more different from Democrats later in the campaign on scales related to key issues discussed in the campaign including attitudes about race (e.g., old-fashioned racism, etc.), gender (e.g., hostile/benevolent sexism), group stereotypes (all race and gender groups), anti-intellectualism, conspiracy endorsement, distrustful complacency, epistemic hubris, intolerance, right-wing populism, and myriad personality measures. This divergence along

⁹ Some work shows that short scales tend to be less reliable (e.g., Ansolabehere et al. 2008; Bakker & Leites 2018), raising the concern that noisier change scores mechanically inflate partisan coefficients. However, the average significant partisanship effect among two-item scales (0.051) is comparable in magnitude to the average among scales with four or more items (0.058), suggesting short scales are not driving our findings.

partisan lines across many scales is important because it shows these commonly used scales change in the partisan environment of an electoral campaign with partisans becoming more polarized over time.¹⁰

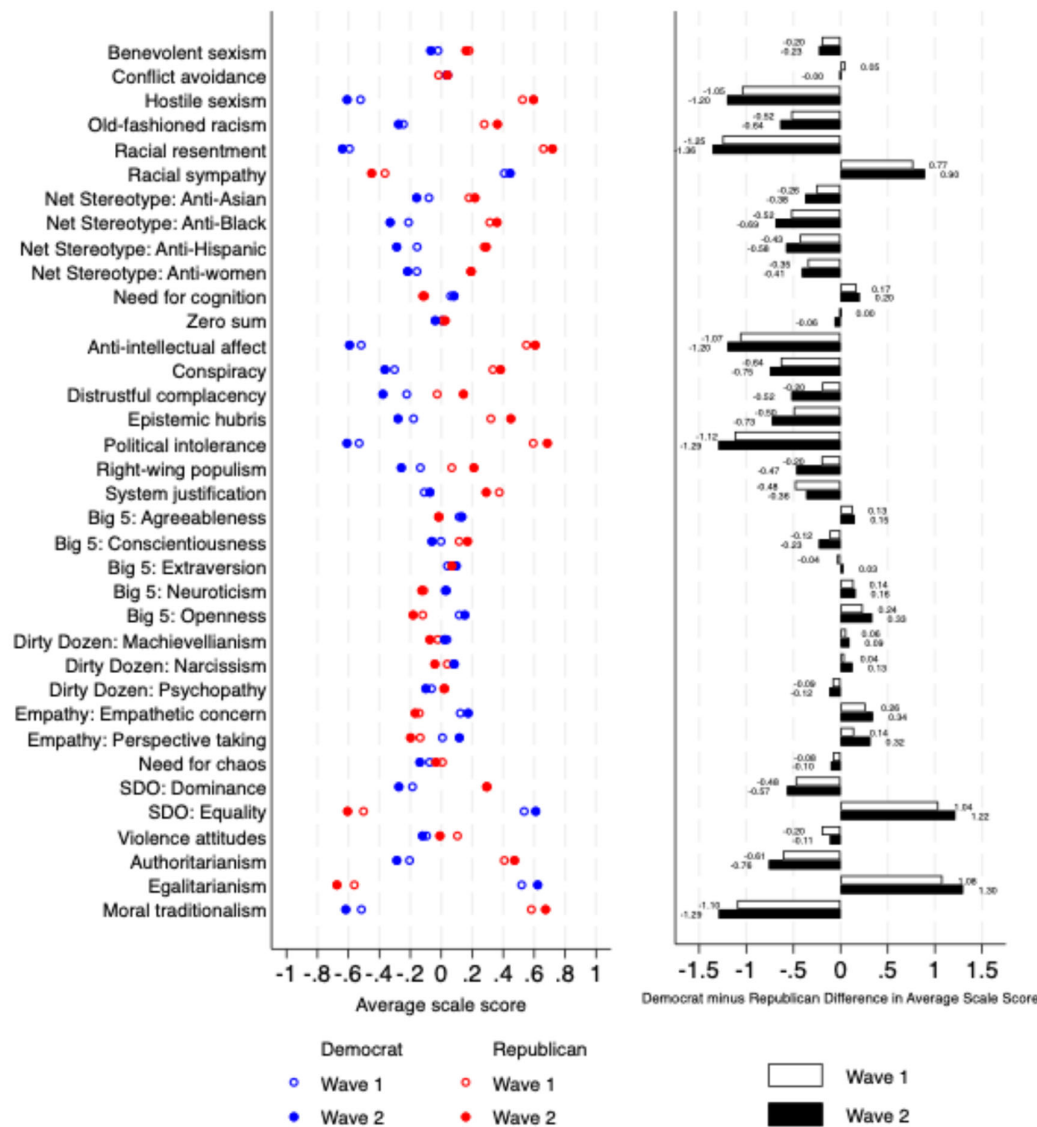
Given the importance of this result, which strongly suggests that many measured scale scores are affected by the combination of (pre-existing) partisanship and the political environment, we examine its robustness and importance. To provide a sense of the magnitude and sources of partisan differences across scales, the left panel of Figure 1 displays the average scale scores by partisanship in each survey wave. We exclude independents and leaners to examine the contrasts among partisans. We note that Democrats and Republicans occupy distinct positions across most scales, with particularly large gaps shown between partisans on hostile sexism, racial resentment, and anti-intellectual affect, among others. In most cases, changes between waves for Democrats and Republicans are symmetric (i.e., both groups change in opposite directions).

The right panel summarizes these differences by displaying the difference between Democratic and Republican averages (Democrat minus Republican) in each wave, thereby collapsing the partisan differences into a single quantity per scale per wave. Positive values indicate higher average scores among Democrats, while negative values indicate higher

¹⁰ Pietryka and MacIntosh (2022) show that some scales (e.g., authoritarianism, egalitarianism, etc.) exhibit differential item functioning (DIF) by partisanship, suggesting items measure different constructs across partisan subgroups. DIF is related to our finding but distinct. In a static analysis, DIF suggests scales measure different traits across parties. Our analysis is consistent with this but also shows over-time variability. Even if DIF is stable over time, it would not predict partisan-correlated *changes* in scale scores because the same measurement bias applies to a given respondent at both waves.

average scores among Republicans. Note that 32 of 36 partisan gaps grow between waves in the direction that reinforces the magnitude of wave 1 differences. Together, these raw differences confirm the earlier regression analysis and provide descriptive context with respect to the magnitudes of change estimated in the models reported in Table 1.

Figure 1. Average scale scores by (2-point) partisanship and differences by wave.



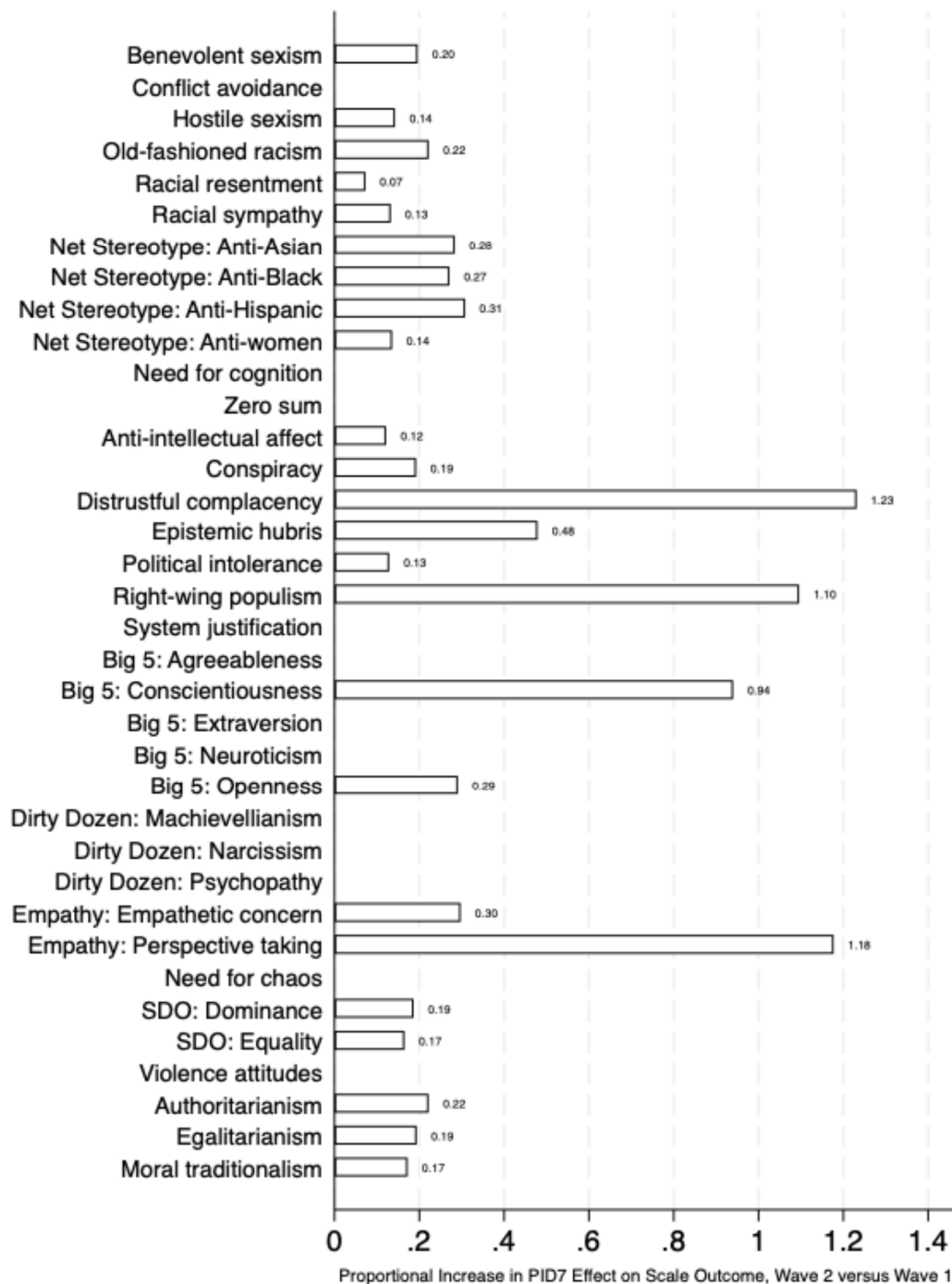
Note: *N* ranges from 818–877 for all scales. Partisanship excludes leaners and independents and was measured in wave 1.

Figure 2 turns from descriptive differences to assess the relative predictive role of partisanship in explaining changes in scale scores between waves. For each scale where partisanship significantly predicted change in the Table 1 specification, we regressed scale scores on seven-point partisanship separately by wave and calculated the proportional increase in the partisanship coefficient from wave 1 to wave 2 (defined as the change in the coefficient divided by its wave 1 value). Values greater than zero indicate that partisan identity became a stronger predictor of scale responses in wave 2.

The figure shows considerable variation in how partisanship predicted scale score response between waves. At one extreme, distrustful complacency shows a proportional increase of 1.23, meaning the estimated effect of partisanship on this scale more than doubled between waves.¹¹ At the other end are scales where the gap was already large in wave 1; scales like racial resentment (0.07) show more modest increases in partisanship's predictive power, though partisanship still became more predictive over time (and is highly correlated with scale scores in both waves). Figure 2 highlights scales in which partisanship increasingly structures scale score response, even when raw partisan gaps remain relatively modest.

¹¹ For distrustful complacency, the partisanship coefficient from wave 1 is 0.046 and it is 0.103 from wave 2. Thus the proportional increase is 1.23 as $(0.103-0.046)/0.046=1.23$.

Figure 2. Relative predictive power of (seven-point) partisanship by wave.



Note: *N* ranges from 1279–1370 for all scales. Proportional increase calculated as (PID coefficient wave 2 minus coefficient wave 1)/coefficient wave 1; see footnote 5 for example calculation. Estimates shown only for scales where PID7 significantly predicts changes in scale scores over time. Complete regression estimates are available upon request.

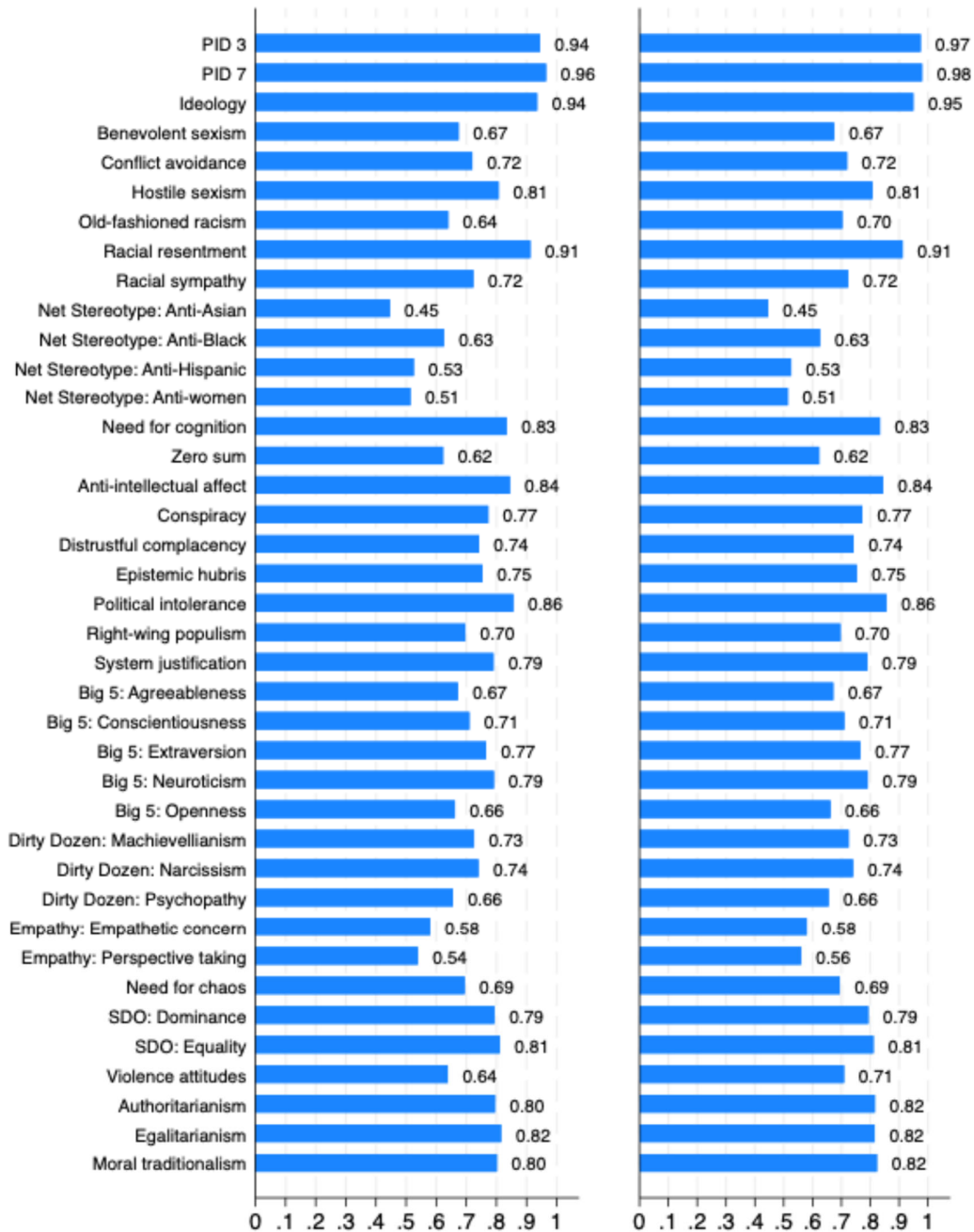
Turning to robustness, Appendix Table C.1 reports the results of additional tests of our partisanship finding, where we list all statistically significant coefficients for our original model alongside alternative specifications (e.g., including just wave 1 scale scores and partisanship, replacing seven-point partisanship with three-point partisanship, etc.). Briefly, only the estimate for benevolent sexism is not robust to any of our alternative modeling choices.

Test-retest reliability of scales

We proceed to the estimated test-retest reliability (stability) of each psychometric scale in our study as measured by the Pearson correlation (left panel) and the polychoric correlation (right panel) of scores between survey waves (a six-month period), shown in Figure 3.¹² Scatter plots of the underlying data for each scale appear in Appendix Figure E.1.

¹² We compute both correlations on the constructed additive scale scores (not individual items). In general, polychoric correlations have been shown to better account for the “lumpiness” of ordinal data from Likert scales than Pearson correlations (Holgado-Tello et al. 2008).

Figure 3. Pearson (left) and polychoric (right) correlation coefficients (reliability) between wave 1 and wave 2 by scale (additive scales, not standardized).



Note: *N* ranges from 1279–1370 for all scales. For PID3, PID7, and Ideology, *N*=1381. Time between waves is six months.

Party identification (polychoric correlation = 0.97-0.98) and ideology (0.95) are highly stable (reliable). No scale reaches this level, but racial resentment (0.91), political intolerance (0.86), and anti-intellectual affect (0.84) come closest, indicating strong consistency in survey responses over time. In contrast, scales such as net Asian stereotype

endorsement (0.45), net women stereotype endorsement (0.51), and net Hispanic stereotype endorsement (0.53) show the lowest reliability. These data provide a baseline for understanding which measured scales are more similar over time.¹³

Predictors of individual-level stability

Finally, we examine whether key demographic and political variables predict the absolute value of the difference in scale scores between wave 1 and 2 for each scale in a manner similar to our earlier directional analysis. We do not have strong *ex ante* notions about these likely relationships. We use wave 1 measures for all variables to predict changes to rule out the possibility that, for example, partisanship changes over time because of wave 1 scale scores.

We describe these specifications in detail in Appendix D. Appendix Table D.1 reports results for the demographic and political variables that are statistically significant for each scale, of which there are relatively few (24 of 216 coefficients [11% at $p < 0.05$], with three surviving a Bonferroni correction to adjust for multiple hypothesis testing). Substantively, statistically significant coefficients are signed inconsistently, but to the degree that any patterns emerge, stronger partisans have more over-time variability in their measured scale scores for five scales (of which one estimate remains significant with adjustment) and those who are older have less over time variability for eight scales (of which one estimate remains significant with adjustment; there is greater variability for two scales). Overall, it does not appear that these demographic or political variables, apart from age, consistently predict the stability of measured scale scores over time.

¹³ As a reminder, random measurement error will tend to depress correlations and a small correlation alone is not an indication that an underlying trait (rather than its measure) is unstable.

Discussion

The use of psychometric scales in political science (and related disciplines) has become ubiquitous. Despite their popularity, there has been, up to this point, little systematic study of their independence from political events and relative reliability (stability). In this paper, we bring novel survey data to bear to test underlying assumptions about these features and find three important results.

First, and perhaps most importantly, we show that partisanship strongly predicts change in measured scale scores over the course of the 2024 presidential election, with the changes that occur tending to increase existing partisan differences. Our analyses reveal that Republicans (Democrats) tended to increase over time on scales that Republicans (Democrats) began higher on (and vice versa for scales on which partisans are low). Overall, we estimate partisanship has a statistically significant correlation with change in measured scale scores for 24 (67%) of the scales we fielded, effects that are significant even with multiple-comparison adjustments for 16 (44%) of these scales. These observations suggest that shifts in psychometric scale responses often move in ways that accentuate existing partisan gaps and that contemporaneously measured scales are not stable measures of dispositions but instead vary meaningfully across political context. Across our other predictors of directional change, we also find some interesting results for both age and gender.

Second, the stability (test-retest reliability) of these scales varies substantially, though it is generally high. Compared to our party identification benchmark (0.97-0.98 polychoric correlation coefficient), we note wide variation. Responses to the racial resentment scale, for example, appears comparably stable between our two survey waves (0.91), while net anti-Asian stereotype endorsement shows the lowest stability (0.45). Notably, however, even scales with high stability still differ systematically with partisanship.

Third, we find that differential (absolute) instability in scale scores is uncommon. Absolute deviations in scale scores over time are generally not predicted by demographic and political variables including education, partisanship, and age. While we see small differences by scale, there is no overall pattern.

Researchers should find parts of these results comforting and others worrying. That reliability (stability) is high and relatively evenly distributed is encouraging, and the comparison of stability across scales using the results presented here will allow future research to accurately judge the scales they use. On the other hand, that partisanship should predict directional changes so starkly is consistent with the idea that attitudes and political variables are endogenous to one another, making causal inference using these scales substantially more challenging.

Whether these findings reflect genuine trait instability or measurement error properties is difficult to separate with two-wave panel data. Test-retest correlations reflect both sources of variation, which is precisely what applied researchers confront when using summed scale scores in their analyses. Future work could investigate this by studying whether differences in stability are driven by measurement error or true differences in the stability of underlying traits. Studying attitudes using three waves, for example, makes it straightforward to non-parametrically estimate the share of variance due to measurement error compared to instability of the underlying concept (Kustov et al. 2021). If the issue is measurement error, methods such as instrumental variables or using an error-in-variables model hold promise. If the issue is the instability of underlying concepts, however, then researchers should use that information to temper the inferences they draw from cross-sectional correlations.

Regardless of which effect is at work (or both), these results suggest strong caution is warranted if one wants to interpret the effect of these scales (as measured proximate to an

election) on outcomes as reflect the effect of those traits. Without clear and strong assumptions, these measured scales do not appear independent of the influence of partisanship.

References

- Alwin, Duane F., and Jon A. Krosnick. 1991. "The Reliability of Survey Attitude Measurement: The Influence of Question and Respondent Attributes." *Sociological Methods & Research* 20 (1): 139–81. <https://doi.org/10.1177/0049124191020001005>.
- 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. <https://doi.org/10.1017/S0003055408080210>.
- Armaly, Miles T., and Adam M. Enders. 2022. "'Why Me?' The Role of Perceived Victimhood in American Politics." *Political Behavior* 44 (4): 1583–609. <https://doi.org/10.1007/s11109-020-09662-x>.
- 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. <https://doi.org/10.1086/698928>.
- Bakker, Bert N., Yphtach Lelkes, and Ariel Malka. 2021. "Reconsidering the Link Between Self-Reported Personality Traits and Political Preferences." *American Political Science Review* 115 (4): 1482–98. <https://doi.org/10.1017/S0003055421000605>.
- 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. <https://doi.org/10.1017/S0003055421000988>.
- Boston, Joshua, Jonathan Homola, Betsy Sinclair, Michelle Torres, and Patrick D. Tucker. 2018. "The Dynamic Relationship between Personality Stability and Political Attitudes." *Public Opinion Quarterly* 82 (S1): 843–65. <https://doi.org/10.1093/poq/nfy001>.
- 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. <https://doi.org/10.3389/fpsyg.2013.00225>.
- Bullock, John G. 2011. "Elite Influence on Public Opinion in an Informed Electorate." *American Political Science Review* 105 (3): 496–515. <https://doi.org/10.1017/S0003055411000165>.
- 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. https://doi.org/10.1207/s15327752jpa4803_13.
- Carsey, Thomas M., and Geoffrey C. Layman. 2006. "Changing Sides or Changing Minds? Party Identification and Policy Preferences in the American Electorate." *American Journal of Political Science* 50 (2): 464–77. <https://doi.org/10.1111/j.1540-5907.2006.00196.x>.
- Chinoy, Sahil, Nathan Nunn, Sandra Sequeira, and Stefanie Stantcheva. 2023. *Zero-Sum Thinking and the Roots of US Political Differences*. No. W31688. National Bureau of

- Economic Research. <https://doi.org/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. <https://doi.org/10.1111/pops.12615>.
- Chudy, Jennifer. 2021. "Racial Sympathy and Its Political Consequences." *The Journal of Politics* 83 (1): 122–36. <https://doi.org/10.1086/708953>.
- Davis, Mark H. 1983. "Measuring Individual Differences in Empathy: Evidence for a Multidimensional Approach." *Journal of Personality and Social Psychology* 44 (1): 113–26. <https://doi.org/10.1037/0022-3514.44.1.113>.
- Douglas, Karen M., Joseph E. Uscinski, Robbie M. Sutton, et al. 2019. "Understanding Conspiracy Theories." *Political Psychology* 40 (S1): 3–35. <https://doi.org/10.1111/pops.12568>.
- Enders, Adam M., Joseph E. Uscinski, Casey A. Klofstad, et al. 2021. "Do Conspiracy Beliefs Form a Belief System? Examining the Structure and Organization of Conspiracy Beliefs." *Journal of Social and Political Psychology* 9 (1): 255–71. <https://doi.org/10.5964/jspp.5649>.
- Engelhardt, Andrew M. 2021. "Racial Attitudes Through a Partisan Lens." *British Journal of Political Science* 51 (3): 1062–79. <https://doi.org/10.1017/S0007123419000437>.
- Fawzi, Nayla. 2019. "Untrustworthy News and the Media as 'Enemy of the People?' How a Populist Worldview Shapes Recipients' Attitudes toward the Media." *The International Journal of Press/Politics* 24 (2): 146–64. <https://doi.org/10.1177/1940161218811981>.
- Gerber, Alan S., Gregory A. Huber, David Doherty, Conor M. Dowling, Connor Raso, and Shang E. Ha. 2011. "Personality Traits and Participation in Political Processes." *The Journal of Politics* 73 (3): 692–706. <https://doi.org/10.1017/S0022381611000399>.
- 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. <https://doi.org/10.1037/0022-3514.70.3.491>.
- Goldman, Seth K., and Daniel J. Hopkins. 2020. "Past Place, Present Prejudice: The Impact of Adolescent Racial Context on White Racial Attitudes." *The Journal of Politics* 82 (2): 529–42. <https://doi.org/10.1086/706461>.
- Goldstein, Susan B. 1999. "Construction and Validation of a Conflict Communication Scale." *Journal of Applied Social Psychology* 29 (9): 1803–32. <https://doi.org/10.1111/j.1559-1816.1999.tb00153.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. [https://doi.org/10.1016/S0092-6566\(03\)00046-1](https://doi.org/10.1016/S0092-6566(03)00046-1).
- Green, Donald P., and Oliver Hyman-Metzger. 2024. "The Vietnam Draft Lottery and Whites' Racial Attitudes: Evidence from the General Social Survey." *American*

- Political Science Review*, December 17, 1–8.
<https://doi.org/10.1017/S0003055424001266>.
- Ho, Arnold K., Jim Sidanius, Nour Kteily, et al. 2015. “The Nature of Social Dominance Orientation: Theorizing and Measuring Preferences for Intergroup Inequality Using the New SDO_r Scale.” *Journal of Personality and Social Psychology* 109 (6): 1003–28. <https://doi.org/10.1037/pspi0000033>.
- Holgado-Tello, Francisco Pablo, Salvador Chacón-Moscoso, Isabel Barbero-García, and Enrique Vila-Abad. 2010. “Polychoric versus Pearson Correlations in Exploratory and Confirmatory Factor Analysis of Ordinal Variables.” *Quality & Quantity* 44 (1): 153–66. <https://doi.org/10.1007/s11135-008-9190-y>.
- Homola, Jonathan, Miguel M. Pereira, and Margit Tavits. 2020. “Legacies of the Third Reich: Concentration Camps and Out-Group Intolerance.” *American Political Science Review* 114 (2): 573–90. <https://doi.org/10.1017/S0003055419000832>.
- Jonason, Peter K., and Gregory D. Webster. 2010. “The Dirty Dozen: A Concise Measure of the Dark Triad.” *Psychological Assessment* 22 (2): 420–32.
<https://doi.org/10.1037/a0019265>.
- Jost, John T., Mahzarin R. Banaji, and Brian A. Nosek. 2004. “A Decade of System Justification Theory: Accumulated Evidence of Conscious and Unconscious Bolstering of the Status Quo.” *Political Psychology* 25 (6): 881–919.
<https://doi.org/10.1111/j.1467-9221.2004.00402.x>.
- 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. <https://doi.org/10.1037/0022-3514.85.5.823>.
- Kinder, Donald R., and Lynn M. Sanders. 1996. *Divided by Color: Racial Politics and Democratic Ideals*. American Politics and Political Economy. University of Chicago Press.
- Krosnick, Jon A., and Matthew K. Berent. 1993. “Comparisons of Party Identification and Policy Preferences: The Impact of Survey Question Format.” *American Journal of Political Science* 37 (3): 941. <https://doi.org/10.2307/2111580>.
- Kustov, Alexander, Dillon Laaker, and Cassidy Reller. 2021. “The Stability of Immigration Attitudes: Evidence and Implications.” *The Journal of Politics* 83 (4): 1478–94.
<https://doi.org/10.1086/715061>.
- Lalot, Fanny, Dominic Abrams, Maria S. Heering, et al. 2023. “Distrustful Complacency and the COVID -19 Vaccine: How Concern and Political Trust Interact to Affect Vaccine Hesitancy.” *Political Psychology* 44 (5): 983–1011.
<https://doi.org/10.1111/pops.12871>.
- Lalot, Fanny, Dominic Abrams, and Giovanni A. Travaglino. 2021. “Aversion Amplification in the Emerging COVID -19 Pandemic: The Impact of Political Trust and Subjective Uncertainty on Perceived Threat.” *Journal of Community & Applied Social Psychology* 31 (2): 213–22. <https://doi.org/10.1002/casp.2490>.

- Landry, Alexander P., James N. Druckman, and Robb Willer. 2024. "Need for Chaos and Dehumanization Are Robustly Associated with Support for Partisan Violence, While Political Measures Are Not." *Political Behavior* 46 (4): 2631–55. <https://doi.org/10.1007/s11109-024-09934-w>.
- Laustsen, Lasse, and Michael Bang Petersen. 2017. "Perceived Conflict and Leader Dominance: Individual and Contextual Factors Behind Preferences for Dominant Leaders." *Political Psychology* 38 (6): 1083–101. <https://doi.org/10.1111/pops.12403>.
- Lenz, Gabriel S. 2012. *Follow the Leader? How Voters Respond to Politicians' Policies and Performance*. Chicago Studies in American Politics. The University of Chicago Press.
- Lunz Trujillo, Kristin. 2022. "Rural Identity as a Contributing Factor to Anti-Intellectualism in the U.S." *Political Behavior* 44 (3): 1509–32. <https://doi.org/10.1007/s11109-022-09770-w>.
- Luttig, Matthew D. 2021. "Reconsidering the Relationship between Authoritarianism and Republican Support in 2016 and Beyond." *The Journal of Politics* 83 (2): 783–87. <https://doi.org/10.1086/710145>.
- Mutz, Diana C., and Byron Reeves. 2005. "The New Videomalaise: Effects of Televised Incivility on Political Trust." *American Political Science Review* 99 (1): 1–15. <https://doi.org/10.1017/S0003055405051452>.
- Osborne, Danny, and Chris G. Sibley. 2020. "Does Openness to Experience Predict Changes in Conservatism? A Nine-Wave Longitudinal Investigation into the Personality Roots to Ideology." *Journal of Research in Personality* 87 (August): 103979. <https://doi.org/10.1016/j.jrp.2020.103979>.
- Pérez, Efrén, Crystal Robertson, and Bianca Vicuña. 2023. "Prejudiced When Climbing Up or When Falling Down? Why Some People of Color Express Anti-Black Racism." *American Political Science Review* 117 (1): 168–83. <https://doi.org/10.1017/S0003055422000545>.
- 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): 1486–505. <https://doi.org/10.1017/S0003055422001447>.
- Peyton, Kyle, and Gregory A. Huber. 2021. "Racial Resentment, Prejudice, and Discrimination." *The Journal of Politics* 83 (4): 1829–36. <https://doi.org/10.1086/711558>.
- Pietryka, Matthew T., and Randall C. MacIntosh. 2022. "ANES Scales Often Do Not Measure What You Think They Measure." *The Journal of Politics* 84 (2): 1074–90. <https://doi.org/10.1086/715251>.
- Rush, Tye, Chelsea Jones, Michael Herndon, and Matt A. Barreto. 2025. "Catalysts of Insurrection: How White Racial Antipathy Influenced Beliefs of Voter Fraud and Support for the January 6th Insurrection." *The Journal of Race, Ethnicity, and Politics* 10 (1): 5–25. <https://doi.org/10.1017/rep.2025.9>.
- Schickler, Eric, and Donald Philip Green. 1997. "The Stability of Party Identification in

- Western Democracies: Results from Eight Panel Surveys.” *Comparative Political Studies* 30 (4): 450–83. <https://doi.org/10.1177/0010414097030004003>.
- 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. <https://doi.org/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. <https://doi.org/10.1017/S0003055419000534>.
- Stephens-Dougan, LaFleur. 2021. “The Persistence of Racial Cues and Appeals in American Elections.” *Annual Review of Political Science* 24 (1): 301–20. <https://doi.org/10.1146/annurev-polisci-082619-015522>.
- Uscinski, Joseph E., and Joseph M. Parent. 2014. *American Conspiracy Theories*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199351800.001.0001>.
- 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. <https://doi.org/10.1007/s11109-024-09961-7>.
- Winter, Nicholas J. G. 2023. “Hostile Sexism, Benevolent Sexism, and American Elections.” *Politics & Gender* 19 (2): 427–56. <https://doi.org/10.1017/S1743923X22000010>.