Acquiescence Bias and Criterion Validity: Problems and Potential Solutions for Agree/disagree Scales

Scott Clifford
Professor
Department of Political Science
Texas A&M University

Andrew M. Engelhardt
Assistant Professor
Department of Political Science
Stony Brook University

Abstract. Scholars frequently measure dispositions like populism, conspiracism, racism, and sexism by asking survey respondents whether they agree or disagree with statements exemplifying those constructs. This agree/disagree format is known to introduce acquiescence bias, which can inflate endorsement rates. The evidence is less clear, however, on how acquiescence affects criterion validity. Though many survey researchers are aware of the problem of acquiescence, we show that political scientists frequently overlook it. First, across 160 tests we show that acquiescence bias can severely distort the magnitude of relationships between constructs and even produce sign errors. Second, an internal meta-analysis indicates acquiescence bias varies in size across samples and constructs. Third, we evaluate balanced scales and measurement models as potential solutions and find that estimated relationships between constructs are sensitive to both measurement choices. We conclude with survey design advice and recommendations for redesigning several popular measures of political and psychological dispositions.

Word Count: 8998

Data Availability: The code and data to reproduce the article's analyses are available on the *Political Behavior* Dataverse page https://doi.org/10.7910/DVN/RYHYCI

Scholarly investigations into democratic backsliding, the success of rightwing populist parties, and relationships between conspiratorial thinking and lax uptake of the COVID-19 vaccine and political violence have provided invaluable insights into pressing social and political challenges. Recent, consequential work on troubling aspects of the public's political beliefs and attitudes uses survey measures to uncover that people who hold sexist or racist views, who are prone to conspiracy belief, or who endorse populist sentiments are all more likely to support political violence (Armaly and Enders 2022; Kalmoe and Mason 2022; Uscinski et al. 2021).

Although rapidly advancing, much of this research faces a common measurement challenge – acquiescence bias. Measures frequently ask respondents how much they agree or disagree with a statement exemplifying the concept, including "Quite a few of the people running the government are crooked." This question format is easy to write and efficient to administer. Yet, it is prone to acquiescence bias – the tendency of some respondents to agree with a statement regardless of its substantive content. Acquiescence bias can affect descriptive statistics *and* inflate or deflate estimated relationships between constructs (Campbell, Siegman, and Rees 1967; Plieninger 2016; Leiton 2021).

While a standard topic in questionnaire design treatments (Krosnick and Presser 2010), evidence supplied for acquiescence bias's substantive consequences may have led applied researchers to either overlook or underestimate them. This may be because treatments are limited in scope. Recent assessments of acquiescence bias's effects for construct endorsement rates focus on conspiracy theories and interpretations of substantive consequences may depend on the correction employed (compare Clifford, Kim, and Sullivan 2020 with Hill and Roberts 2023). Similarly, whether one believes acquiescence bias affects correlations between constructs may depend on whether one values evidence focused on efficacy or hostile sexism (compare Lelkes

and Weiss 2015 with Archer and Clifford 2021). Combined with methodologists spending more time systematically studying implications for scale measurement properties (Leiton 2021) or construct endorsement rates (Schuman and Presser 1981), a clear picture of how acquiescence might bias estimated correlations between political constructs is lacking. As we report below, a systematic literature review finds scholars often correlate agree/disagree scales but rarely attempt to address acquiescence bias at either the design or analysis stage.

We use four original studies to show how acquiescence bias affects relationships between constructs measured on agree/disagree scales and how much substantive conclusions change after implementing recommended solutions, complementing recent work on debiasing attitude prevalence assessments (Hill and Roberts 2023). We offer four results. First, an internal meta-analysis of 160 acquiescence bias tests indicates that correlations between scales phrased in the same direction are consistently larger than scales phrased in opposing directions; the average difference in correlations is 0.26, reaching 0.35 in a study using a popular online survey firm. Our meta-analysis also suggests that acquiescence bias varies across samples and constructs. Second, we show how simple response scale choices can affect regression results by altering both the magnitude and sign of coefficient estimates, complicating hypothesis testing. Third, we find balanced scales (which consist of an equal number of positively and negatively worded items) do not completely remove acquiescence bias, as explicitly modeling acquiescence bias consistently changes the magnitude – and sometimes even the direction – of the relationship between variables. Fourth, we show how an "ideal" approach pairing both balanced scales and measurement models produces markedly different conclusions compared to the "standard" approach correlating only aligned scales. Much like concerns with content overlap should give

scholars pause when interpreting the correlation between two constructs (e.g., Malka, Lelkes, and Holzer 2017), so too should shared method variance (see also Joel et al 2025).

Acquiescence Bias: Causes and Consequences

While survey researchers have long known that survey respondents will more likely agree with a statement than disagree with its opposite (Campbell, Siegman, and Rees 1967; Schuman and Presser 1981), and questionnaire design reviews describe how acquiescence bias affects descriptive claims (Krosnick and Presser 2010), applied researchers appear to proceed as if it has muted substantive consequences or that solutions are hard to implement. Following public and scholarly concern with democratic stability and social cohesion, we conducted a systematic literature review of measurement practices for a cluster of constructs we name "anti-establishment" dispositions.[1] For scales using an agree/disagree format, we coded whether the scale contained an equal number of positively and negatively worded items (balanced) or not (unbalanced). Of the 108 scales, 44% are unbalanced. However, this figure rises to 90% after excluding racial resentment, suggesting that most researchers use a balanced scale only when it has been established as canonical. Further, 83% of the unbalanced scales are *completely* unbalanced, containing no reversed items. While experiments or event studies often include these

[1] Using GoogleScholar, we canvassed *The American Political Science Review*, *American Journal of Political Science*, *Journal of Politics*, *Political Behavior*, and *Public Opinion Quarterly* for articles published 2016-2024 that included one of the following as a focal independent or dependent variable and measured using agree/disagree scales: anti-democratic attitudes, conspiratorial thinking, need for chaos, political violence, populism, hostile sexism, and racial resentment (Appendix 1).

measures, 20% of the time scholars correlate these with other agree/disagree outcomes. Of these correlates, 73% are also unbalanced. Further, authors using agree/disagree scales rarely mention acquiescence (6% of all articles). These findings suggest limited concern over acquiescence bias.

These choices matter because acquiescence bias has multiple effects. While applied researchers may recognize inflated construct endorsement rates, acquiescence also biases estimates of the relationship between constructs using agree/disagree scales (Campbell, Siegman, and Rees 1967; Leiton 2021). If all items measuring two constructs are phrased in the same direction, then acquiescence bias inflates the estimated correlation. If the two sets of items are phrased in opposing directions, however, acquiescence bias deflates the estimated correlation (Green and Citrin 1994). For example, Archer and Clifford (2021) found that a short version of hostile sexism including only positively-worded items was strongly positively related to a set of positively-worded items regarding feelings of control over one's life (r = .40), but weakly and negatively related to a set of negatively-worded items designed to measure the same construct (r = -.04).

This problem extends to cases where researchers estimate the relationship between an agree/disagree scale and a construct measured using a different format. This is because acquiescent-responding tendencies vary across individuals. For example, less educated respondents may be more prone to acquiescent responding (say, through increased satisficing), biasing the relationship between education and the relevant construct in a negative direction. Of course, bias can come in more complicated forms as a function of multiple measured and unmeasured variables, making its nature difficult to predict, as mixed results regarding acquiescence bias's predictive validity consequences indicate (Archer and Clifford 2021; Clifford, Kim, and Sullivan 2020; Hill and Roberts 2023; Kuru and Pasek 2016).

Acquiescence bias has multiple causes, making its consequences broad-based. Perhaps most obviously, satisficing or insufficient effort can cause acquiescence. The cognitive effort required to reject statements or translate responses from a latent scale (e.g., "not important") to the response scale (e.g., "slightly disagree") facilitates satisficing (Gilbert, Tafarodi, and Malone 1993). Common survey design features like item grids can exacerbate satisficing. Respondents assume grid content is closely related, a particular problem for included "reversed" or "negatively keyed" items where agreement indicates rejecting the underlying construct (Couper et al. 2013; Neuert, Roßmann, and Silber 2023, Berinsky et al. 2019). But while satisficing tendencies can vary within and across surveys (Berinsky, Margolis, and Sances 2013), acquiescent response style appears more like a personality trait related to individual or cultural characteristics (Billiet and Davidov 2008; Wetzel et al. 2016; Hibbing et al. 2019; Lechner et al. 2019; Rammstedt, Kemper, and Borg 2013). Thus, survey design and individual differences contribute to acquiescence.

Coping With Acquiescence Bias

Survey researchers have proposed many potential solutions to acquiescence bias. Perhaps the best known among applied researchers involve question design. A common solution involves using construct-specific, or item-specific, scales (Krosnick and Presser 2010; Pasek and Krosnick 2010; Saris et al. 2010). Rather than querying respondents' agreement with a statement, the researcher instead develops response options specific to the construct. Instead of asking whether respondents agree that "women seek to gain power by getting control over men," a construct-specific question asks *how often* women seek to gain power by getting control over men (e.g., Archer and Clifford 2021). Respondents then select among answer choices capturing this behavior's *frequency* (e.g., never, rarely), not how much they agree with a statement about a

particular frequency of that behavior. This format consistently reduces acquiescence bias (e.g., Saris et al. 2010), though does not eliminate all response bias (Lelkes and Weiss 2015; Liu, Lee, and Conrad 2015).

Scholars have also proposed forced-choice formats. This design avoids acquiescence by asking respondents to choose between two options instead of expressing their endorsement of a single claim (Schuman and Presser 1981; but see Ray 1990). For example, a popular measure of authoritarianism has respondents choose between trait pairs with one indexing authoritarianism (Engelhardt, Feldman, and Hetherington 2023). Others have developed a conspiracy belief measure where respondents select which of two claims, one conspiratorial and one conventional, is most likely true (Clifford, Kim, and Sullivan 2020).

While alternative scale formats are a promising solution, survey researchers have also proposed methods for addressing acquiescence bias that maintain the convenience and efficiency of agree/disagree scales. These recommendations involve *balanced scales* that include equal numbers of positively worded (PW) and negatively worded (NW) (or reversed) items (Cloud and Vaughan 1970). But while straightforward, our systematic literature review suggests this is infrequently implemented. Under the assumption that acquiescence bias equally affects both types of items, this bias will decrease, ideally to 0. Balanced scales should thus reduce bias in descriptive claims and criterion validity assessments. This logic has led many researchers to create balanced sets of agree/disagree items. For example, the popular racial resentment scale includes two positively worded statements, where agreement indicates high racial resentment, and two negatively worded statements, where agreement indicates *low* racial resentment (Kinder and Sanders 1996).

However, this solution faces several criticisms. First, some argue that balanced scales inadvertently pool particularly acquiescent respondents to the middle of the response dimension, regardless of where they belong (for discussion, see Saris et al. 2010). More generally, it is unclear how well balanced scales purge acquiescence bias (Leiton 2021). Assumptions about acquiescence bias's connection to items and prevalence across respondents, required for unbiased mean or covariance assessments, are frequently untenable (Savalei and Falk 2014).

A second and related concern is that negatively worded items may introduce additional forms of measurement error. In particular, NW items may be more cognitively difficult to answer, leading to greater error, possibly even affecting responses to later items. Researchers typically construct NW items using negations (e.g., "not") or antonyms (e.g., liking "simple" vs. "complex" tasks). Each strategy poses potential problems. Survey researchers generally recommend avoiding negations because respondents must first decide how they feel about the statement and then reverse the response (Saris et al. 2010). This extra step increases the likelihood of respondent errors. Additionally, satisficing may cause a respondent to miss the negation entirely. Using antonyms can be problematic because two seemingly opposing statements may not actually represent polar opposites but instead represent two substantive factors (Weijters and Baumgartner 2012). For example, a respondent might agree that they "like complex problems" and also that they "like simple problems" (Zhang, Noor, and Savalei 2016). While these two statements may be used as positively and negatively worded items to measure Need for Cognition, they may simply capture two distinct constructs.

Recognizing these multiple sources of response variance in items, psychometricians have proposed methods to separate trait and method variance, including acquiescence (Primi et al. 2019; Savalei and Falk 2014), but our systematic literature review suggests these approaches

may be unfamiliar to applied researchers in political science. One approach, ipsatization, addresses acquiescence by calculating respondents' average across all constructs using the same response scale and then subtracting this grand mean from scale-specific scores (Cattell 1944; Cunningham, Cunningham, and Green 1977). Such scores, however, are interpersonally incomparable and, by creating a dependence between scored scales, produce improper covariances (Chan 2003).

Another approach applies structural equation modeling (SEM) to balanced scales. In the single construct case, researchers estimate two factors: a trait factor and a method factor. The method factor isolates common variance in items due to response style with the trait factor containing the remaining common variance, assumed to capture the construct of interest (Savalei and Falk 2014; Leiton 2021). This approach addresses assumptions that acquiescence bias averages out across positively and negatively worded items, required for using balanced scales alone, and potential response styles unique to negatively worded items which may add additional method error variance. While this approach appears to outperform other common factor analytic approaches at removing acquiescence bias, the empirical consequences of accounting for acquiescence this way depend on a construct's susceptibility to acquiescent responding.

Overall, short of abandoning the agree/disagree format, research suggests scholars can mitigate acquiescence bias by using balanced scales, particularly paired with SEM. Yet, doing so requires developing negatively worded items and expertise in structural equation modeling. Further, these methods make assumptions about the nature of bias across items and constructs that are difficult to test. But as we show below, these costs might be worth paying given the potential empirical consequences of ignoring acquiescence.

Acquiescence Bias and Anti-Establishment Beliefs

We assess systemically acquiescence bias's criterion validity consequences and the sufficiency of balanced scales, alone or paired with SEM, by focusing on a cluster of concepts related to anti-establishment beliefs and attitudes and some primary correlates. We first briefly review some measures of each concept and discuss prominent examples and then discuss how acquiescence bias may affect evidence connecting these concepts.

Anti-Democratic Attitudes

No standard battery exists to capture anti-democratic attitudes, but common items record responses on agree/disagree scales.[2] Bartels (2020), for example, used how much respondents agreed that "Strong leaders sometimes have to bend the rules in order to get things done." Other studies probe support for partisan ends over democratic processes, asking respondents for instance "[Democrats/Republicans] should redraw districts to maximize their potential to win more seats in federal elections, even if it may be technically illegal" (Voelkel et al 2023). Scholars commonly correlate such scales with other constructs measured with unbalanced agree/disagree scales (Sprong et al. 2019; Young, Jackson, and Boyon 2023). While at least one author directly addresses acquiescence via a measurement model (Bartels 2020), this practice is rare.

Conspiracy Beliefs

Conspiratorial beliefs are measured in myriad ways but often on an agree/disagree scale (cf. Brotherton, French and Pickering 2013; Bruder et al 2013). A popular measure (e.g., Golec

[2] Cross-national surveys like the Asian, Arab, and Americas Barometers also offer useful operationalizations but often use agree/disagree scales.

de Zavala and Federico 2018), includes an item proposing: "Much of our lives are being controlled by plots hatched in secret places" (Uscinski, Klofstad, and Atkinson 2016). Agreement with this and three other items indicates greater conspiracy belief. Other scholars measure agreement with specific conspiracies. Prominently, Oliver and Wood (2014) record respondents' agreement with seven conspiracies, like whether US officials allowed the 9/11 attacks to happen. While these authors discuss acquiescence bias inflating endorsement rates, they do not address consequences for criterion validity. This is particularly important given that they measure their focal independent variables with unbalanced agree/disagree scales. Critically, recent work demonstrates that acquiescence bias in questions about specific conspiracy beliefs tends to inflate endorsement rates and can bias correlations with other constructs (Clifford, Kim, and Sullivan 2020; Hill and Roberts 2023). However, we are unaware of evidence for measures of conspiratorial predispositions.

Need for Chaos

Petersen, Osmundsen, and Arceneaux (2023) introduce the Need for Chaos–"a desire for a new beginning through the destruction of order and established structures" (1489)–measured by agreement with items like "I think society should be burned to the ground." Recognizing concerns with acquiescence, the authors added two negatively-worded items to an original 7-item scale and took several steps to evaluate their results' robustness.[3] Nonetheless, the revised scale remains only partially balanced, and some scholars use the initial unbalanced scale, often correlating it with other unbalanced agree/disagree scales. For example, scholars have correlated

[3] For example, they compared correlations after removing respondents potentially more prone to acquiescent responding, and analyzed a reversed dependent variable.

an unbalanced NfC scale with other unbalanced scales (Bartusevičius et al 2021; Bor, Jørgensen, and Petersen 2023; Farhart et al 2023). Frequently, analyses find that scales constructed from items sharing response scale direction exhibit the strongest correlations. While these findings are often interpreted as substantive evidence for the strongest predictors, shared acquiescence bias may partially explain them.

Populism

Interest in populist beliefs has motivated scholars to develop several scales recording respondents' agreement with statements like "Politics is ultimately a struggle between good and evil." Recently, Castanho Silva et al. (2020) assessed seven alternative scales, three of which use an agree/disagree format with no reversed items (Akkerman, Mudde, and Zaslove 2014; Elchardus and Spruyt 2014; Schulz et al. 2018). These scales have been used in various applied studies (e.g., Aytac, Carkoglu, and Elci 2021; Spruyt, Keppens, and Van Droogenbroeck 2016) and scholars have even combined conspiracy belief and populism measures into an "anti-establishment" beliefs measure that excludes negatively-worded items (Uscinski et al. 2021).

Populism measures thus manifest in many forms but often as unbalanced agree/disagree scales that are then correlated with other unbalanced scales. For example, recent work examined the correlation between an unbalanced measure of populism and a partially balanced measure of support for political violence (Piazza 2023). Further, this author examined four possible mediators, two of which were measured with unbalanced agree/disagree scales. These latter two mediators turned out to be the strongest, a finding perhaps influenced by shared acquiescence bias. This is just one example of what seems to be a common practice within this literature.

Support for Political Violence

Support for political violence has garnered attention recently, particularly following the January 6 insurrection. An early study developed a five-item scale that recorded agreement with statements like "Sometimes the only way to stop bad government is with physical force" (Kalmoe 2014). One of the five items is negatively-worded, capturing endorsement of non-violence in politics. Recently, scholars have used similar two- to three-item scales that omit negatively-worded items (Armaly, Buckley, and Enders 2022; Armaly and Enders 2022; Uscinski et al. 2021). Critics note these measures likely inflate estimates of support for violence and may overstate correlations with other constructs (Westwood et al. 2022).

One recent and prominent example predicts support for political violence with Christian Nationalism, perceived victimhood, and white racial identity (Armaly, Buckley, and Enders 2022). It measured all constructs with completely unbalanced agree/disagree scales, except for Christian Nationalism, which contains one negatively worded item (out of six). All constructs strongly correlated with support for political violence, with Christian Nationalism's relationship the weakest. Acquiescence likely biases these correlations, but to different degrees across scales depending on their measurement.

Hostile Sexism

We now turn to two well-established constructs scholars have used to predict anti-establishment beliefs. The first, hostile sexism, is arguably the most popular and most theoretically relevant aspect of sexism used in political science (Schaffner 2021), including predicting anti-establishment orientations (Kalmoe and Mason 2022). Hostile sexism measures agreement with statements like "women are too easily offended" (Glick and Fiske 1996). Of 11 items, three are reversed (e.g., "Feminists are making entirely reasonable demands of men"), which the original authors included specifically to combat acquiescence bias. However, scholars

commonly use shortened versions of the scale, particularly in political science. Some abbreviated scales might end up unbalanced at the questionnaire development stage due simply to space constraints (Cassese and Holman 2019). Other work has systematically developed and validated at least two shortened scales, but both remove all of the negatively-worded items explicitly because they loaded on a second factor (Rollero, Glick, and Tartaglia 2014; Schaffner 2021). This suggests a strong method factor in responses (Green and Citrin 1994), but removing the reversed items then confounds the method factor with the substantive factor. These and other problems motivated scholars to develop an item-specific version of the scale (Archer and Clifford 2021), though it has not yet been widely adopted.

Racial Resentment

Racial resentment features prominently in the measurement of racial attitudes (Kinder and Sanders 1996; Kam and Burge 2018) and frequently relates to anti-establishment dispositions (Armaly and Enders 2022; Kalmoe and Mason 2022). Its typical operationalization is a balanced four-item scale. Importantly for our present endeavor, scholars find that symbolic racism, of which racial resentment is one variation, contains two subdimensions (Tarman and Sears 2005), with the forward and reversed pairs loading on separate factors.[4] These item pairs may therefore be associated due to both trait and common method.

Various theories suggest these constructs – anti-democratic attitudes, need for chaos, populism, conspiracy belief, support for violence, hostile sexism, and racial resentment – relate to each other. While considerable research thus assesses these relationships, our systematic

[4] Method effects cannot explain this fully. Items measured on different response scales have factor loadings paralleling the four conventional items (Tarman and Sears 2005).

literature review indicates scholars typically measure these constructs with agree/disagree scales containing no reversed items. Consequently, research examining relationships between these constructs may be inflating estimates due to acquiescence bias that is shared across scales, as we show below.

Overview of Studies and Tests

We assess acquiescence bias's criterion validity consequences and potential solutions with four separate studies using the aforementioned constructs. While some study details differ, combined they let us test key questions relevant to applied researchers.

Our core test relies on scales that include both positively and negatively worded items.[5] If the scales capture only respondents' beliefs and attitudes, rather than acquiescence, then the two positively-worded and negatively-worded scale halves should be interchangeable. For example, the correlation between hostile sexism and support for political violence should be the same, regardless of whether we use the positively- or negatively-worded items from the validated scale. In contrast, we expect that when both constructs are measured such that agreement with the item indicates high values of the construct (scale directions are matched), the estimated relationship will be more positive than when the directions of the scales do not match. We estimate acquiescence bias's magnitude with the difference between the estimated correlations in these

[5] Negatively worded items are not necessarily direct opposites or negations of a specific positively worded item but aim to capture the same construct. See the racial resentment items (Appendix A2).

two scenarios (Campbell, Siegman, and Rees 1967). This approach assumes that the positively-worded and negatively-worded subscales equally capture the intended construct.[6]

We collected responses to several well-established and widely used scales across our studies (Castanho Silva et al. 2018; Bartels 2020; Uscinski, Klofstad, and Atkinson 2016; Glick and Fiske 1996; Schaffner 2021; Kinder and Sanders 1996; Petersen, Osmundsen, and Arceneaux 2023; Kalmoe 2014). Some scales include no reversed items, limiting the tests we can conduct. However, we can still compare these scales to the two halves of other balanced scales. In two studies, we create new reversed items to balance all included scales. Table 1 summarizes scale inclusion and type (balanced, unbalanced) across studies. The bottom row displays the number of paired comparisons each study contributes.

Table 1: Summary of Measures Across Studies

Study	1 (Student)	2 (Dynata)	3 (Mturk)	4 (Bovitz)
Conspiratorial Predispositions	U	U	B*	B*
Need for Chaos			B*	B*
Support for Violence	U		B*	B*
Populism		B	B	B
Anti-Democratic		U		B*
Hostile Sexism	B	B	B	B
Racial Resentment	B		B	B
Paired comparisons	8	8	60	84

U: unbalanced. B: balanced. B: new scale items added to create balanced scale. Survey item wording included in appendix 2.*

Combined, the four studies and seven constructs give us 160 pairs of correlations where the two constructs are held constant, but one correlation is aligned and the other is unaligned. The difference between these two correlations is our estimate of the effects of acquiescence bias. For example, in Study 2 respondents completed a balanced hostile sexism scale and a completely

[6] The appendix (4 and 5) offers evidence that NW items and subscales do not systematically differ from PW items by introducing new error (e.g., cognitive burden). Instead, differences appear scale-specific. This points to the importance of continued scale validation and refinement.

unbalanced conspiratorial predispositions scale. These two scales contribute one difference in correlations by comparing the correlation between conspiratorial predispositions and the positively worded hostile sexism items and between conspiratorial predispositions and the negatively worded hostile sexism items. When both scales are completely balanced, we calculate four correlations and four differences between aligned (positive-positive, negative-negative) and unaligned (positive-negative, negative-positive) pairs.[7] Our analysis excludes all comparisons between scale pairs where both are unbalanced.

We follow these tests by considering applied consequences and solutions. Our systematic literature review revealed scholars often correlate unbalanced agree/disagree scales, and rarely correct for acquiescence bias. We probe the consequences of these choices by first offering a set of results plausibly reached based simply on different measurement choices using agree/disagree scales. We then evaluate the import of combining balanced scales with SEM as a solution to acquiescence bias, investigating how much design-stage and analysis-stage choices might matter for information gleaned from correlating constructs.

Study 1

Subjects were drawn from introductory and required political science courses at a large public university in Fall 2022 (N = 1,271). Respondents completed questions for unrelated projects before answering our scales.

Unique to this study, respondents completed only the positively- or negatively-worded sets of scales including reversed items, with this assigned on a random basis. For example, for

[7] To illustrate, Study 4 contributes 84 paired comparisons from its seven balanced scales: $\frac{7!}{2!(7-2)!} \times 4 = 84.$

hostile sexism, respondents received either the three reversed items from the scale, or three items randomly selected from the eight positively worded. All scale items were placed in a single grid and used the same seven-point response scale. All grids appeared on separate survey pages.

Study 2

Dynata recruited respondents for a 15-minute survey that was fielded from Dec. 2-11, 2022. The intended sample was N=1,000, balanced to census figures on sex, age, race, and region. Due to a programming error, several demographic groups were oversampled and the final sample size was 1,318 (see Appendix 2). To minimize the prevalence of satisficing, we placed an attention check in a grid early in the survey. Respondents who failed the check were not allowed to complete the survey. This is a conservative design choice in that the excluded respondents should be among the most susceptible to displaying acquiescence bias. Like study 1, respondents first completed questions for unrelated projects before answering our various scales. To examine if scale presentation moderates the degree of acquiescence bias, we randomized whether the populism scale was presented in two separate grids for the positively and negatively-worded items or a single grid that intermixed all items. We did the same for the hostile sexism scale. All grids appeared on individual pages. Grid format does not affect our results (Appendix 6).

Study 3

Respondents were recruited from the Mechanical Turk platform using CloudResearch to screen out fraudulent respondents. The study was fielded on Aug. 30, 2023 (N=993). We measured satisficing with an instructed response item placed in a grid and two open-ended questions. We retain all respondents for our primary analyses below, but report robustness

checks in Appendix 9 using attentive respondents.[8] We introduce in this study new reversed items for included unbalanced scales to construct fully balanced scales that would allow us to better assess the impact of acquiescence bias.

Study 4

We contracted with Bovitz to sample 1000 respondents from their Forthright panel with completes matched to Census benchmarks on: gender, race, ethnicity, age, and region. The survey was fielded Oct. 23-29, 2023. We again measured satisficing using an instructed response item and two open-ends. We analyze the full sample but report robustness checks with attentive respondents in Appendix 9.[9] We again introduce reversed items to create completely balanced scales when needed and replace two poor performing reversed items from Study 3.

Creating new scale items in studies 3 and 4 raises the possibility our new items introduce additional measurement error and perhaps exaggerate our results. We find these reversed items validly measure the existing constructs in all but four instances (Appendix 4). The scales we subsequently analyze drop these invalid items.

Acquiescence Bias Can Have Large Effects on Scale Correlations

[8] Respondents were highly attentive: 86% passed all three and 95% passed at least two of the three. Our substantive findings hold using those passing all three checks. Passing our open-ended checks required providing responses generally related to the question (e.g., What word or phrase best describes your political views?).

[9] Respondents were highly attentive: 81% passed all three and 95% passed at least two of the three. Our substantive findings persist using only those passing all checks.

To estimate the difference in correlations, we use a bootstrapping approach that retains the structure of the data. This is particularly important for Studies 2-4 in which respondents completed both halves of all balanced scales. We draw a sample from each dataset of the same size with replacement, calculate each correlation and difference in correlations, repeating this process 1,000 times. We use a normal approximation to determine 95% confidence intervals.

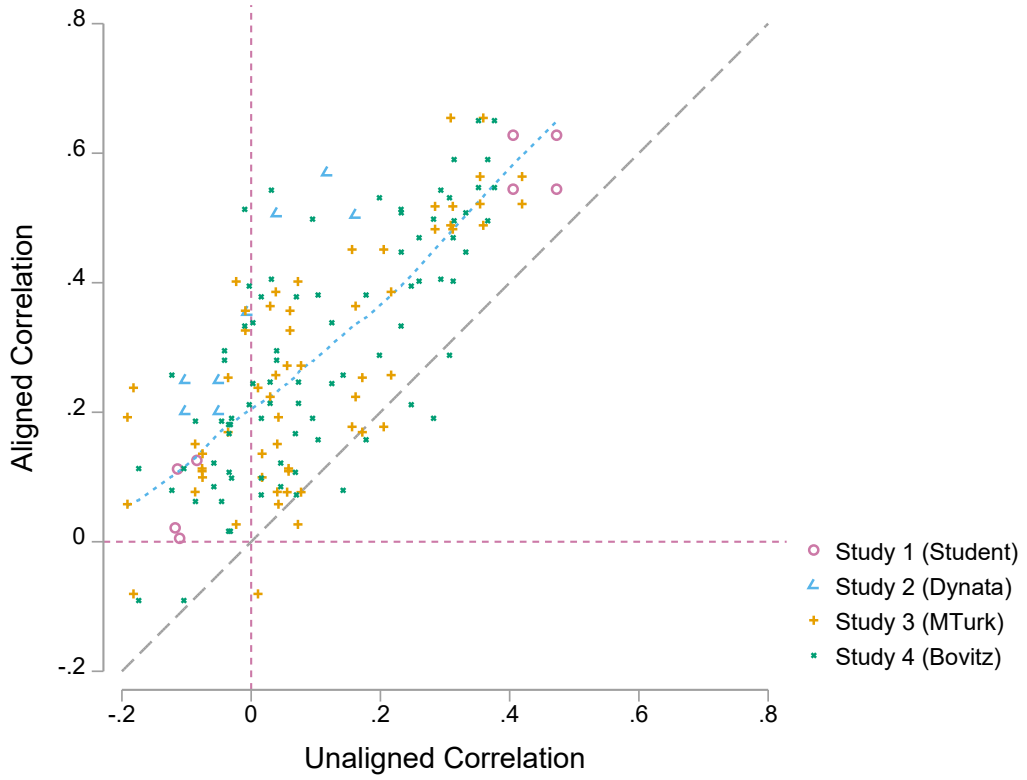
Figure 1 plots the 160 correlation pairs. The y-axis displays the correlation when the two constructs are aligned (e.g., both positively-worded) and the x-axis displays the correlation when the two constructs are unaligned (e.g., one positively-worded and one negatively-worded). If acquiescence bias is completely absent, then the data should fall along the positive diagonal. If acquiescence bias is present, then the data should fall above the positive diagonal, indicating that the correlation between constructs is larger when the scales are aligned than when they are unaligned.

Consistent with expectations, 94% of the data falls above the positive diagonal, indicating consistent acquiescence bias. For each pair of correlations, the difference is significantly larger than zero in 84% of the cases. To illustrate, consider one rather stark example from Study 2 comparing conspiratorial predispositions and hostile sexism. When using the positively worded hostile sexism items, such that agreement indicates high values of both constructs, the two correlate strongly ($r = .51$, $p < .001$). However, when instead using the negatively worded hostile sexism items, this correlation is nearly zero ($r = .04$, $p = .157$) for a difference in correlations of 0.47.[10] The average difference in correlations, across all studies and constructs, is 0.19. This

[10] Recall, survey programming means this includes only respondents passing an instructional attention check embedded in a question grid.

effect is large. Comparing it to the results of a systematic review of more than 6,000 effect sizes in social psychology (Lovakov and Agadullina 2021) we find that shifting from a correlation of 0.10 (average unaligned) to 0.29 (average aligned) represents a shift from the 20th percentile to the 60th percentile. We also find that, of the 160 pairs, 31% differ in sign. Simple measurement choices can lead researchers to make size and sign errors.

Figure 1: Difference in Scale Correlations by Item Alignment



Note: each estimate represents the correlation between a pair of measures. The x-axis shows unaligned correlations (PW/NW, NW/PW) and the y-axis shows aligned correlations (PW/PW, NW/NW). The gray line represents identical correlations regardless of measurement alignment. The blue line is a loess curve showing the average deviation away from identical correlations.

Meta-Regression of Differences in Correlations

We examine variation in acquiescence bias's magnitude with a meta-regression framework where the dependent variable is the difference between each pair of correlations (for

a related application, see Kertzer 2022). Meta-regression allows us to model the size of the difference in correlations (weighted by precision) as a function of covariates, including data source and constructs involved. Following standard practices, we use a random effects model with the Knapp-Hartung modification.

We first model the outcome as a function of dummy variables for each study, with Study 4 the omitted condition. Average levels of acquiescence bias do not significantly differ between Studies 1, 3, and 4, with the effect size ranging 0.16–0.19. However, in Study 2 the average difference in correlations is significantly larger (0.35), equivalent to shifting from the 5th percentile to the 70th percentile of effect sizes in social psychology. While the magnitude of acquiescence bias clearly can differ substantially across studies, the reason is not immediately clear. We took additional caution in this study to exclude inattentive respondents. Thus, sample source matters, but not predictably.

We now expand the model above by including dummy variables for each scale, using racial resentment as the omitted category. This approach assumes additive effects.[11] Each dummy variable indicates how much more the construct increases acquiescence bias relative to racial resentment. Doing so also characterizes whether sample differences hold after controlling for construct, as constructs appear inconsistently across studies.

Figure 2 plots the coefficients from the expanded meta-regression. Estimated sample differences closely track the basic model, suggesting that differences in construct inclusion did not drive the results. Among the constructs, populism, sexism, and violence all have coefficients

[11] Appendix 7 reports substantively similar results from an alternative approach using dummies for each construct pair.

near zero, indicating that they function similarly to racial resentment. In contrast, anti-democratic views, conspiratorial predispositions, and need for chaos are all statistically significant, contributing approximately a 0.10 increase in the difference in coefficients.

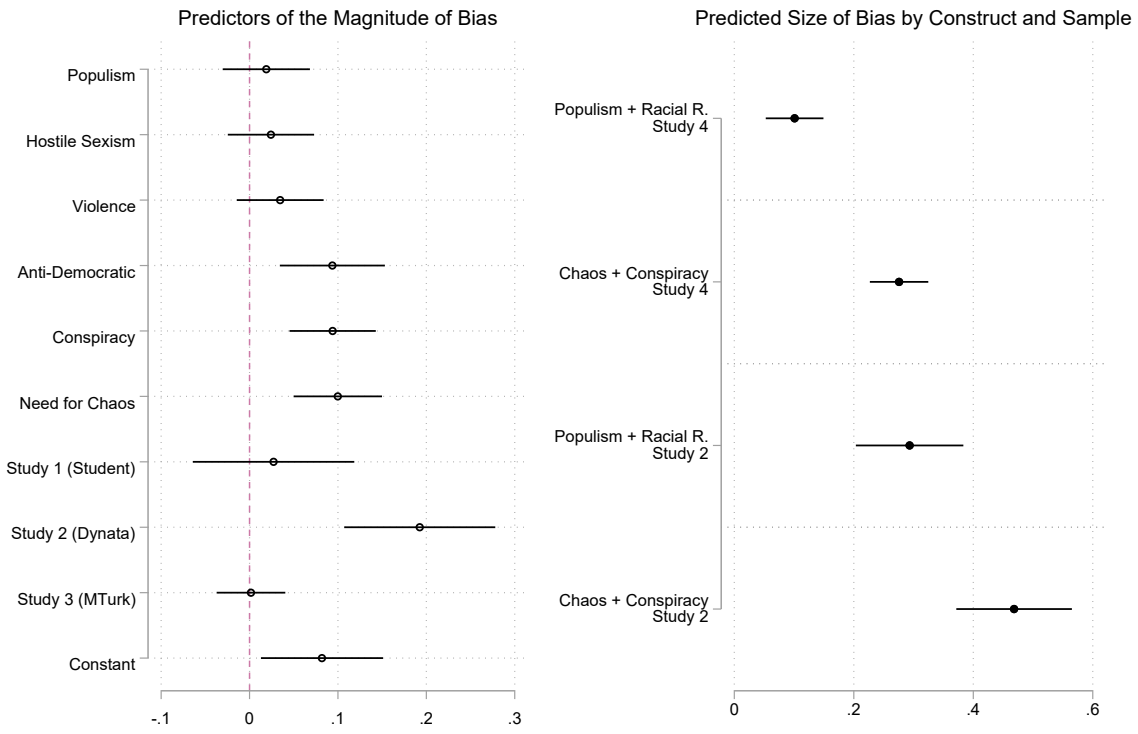
To illustrate how study features can collectively influence acquiescence bias's severity, Figure 2's right-hand panel displays predicted values from the model as a function of the variables involved and the sample used. The top prediction is the best-case scenario – two constructs (populism and racial resentment) that generate the least bias, and the sample that generates the least bias. In this scenario, the predicted bias is a difference in correlations of approximately 0.10. However, as shown by the third estimate, this bias increases to 0.29 when studying the same variables in a lower-quality sample. Both estimates shift considerably when instead studying the relationship between conspiratorial predispositions and need for chaos. Even in a high-quality sample, the predicted bias is 0.28. But, troublingly, that estimate increases to 0.47 in a low-quality sample.

Naturally, one might ask why some constructs exhibit more acquiescence bias. We can only speculate, but three of the constructs least prone to acquiescence (racial resentment, hostile sexism, and populism) have undergone extensive validation (Hussey and Hughes 2020). Additionally, some of the more bias-prone constructs tend to be more abstract or have higher conceptual bandwidth (for discussion, see Bakker and Lelkes 2018), such as Need for Chaos. Both construct bandwidth and scale development may contribute to higher measurement error and the level of acquiescence bias. However, further research is needed.

We also wonder about sample differences in bias. Many scholars would be tempted to treat the Dynata and Bovitz samples similarly, given that both quota sample from opt-in panels, and distinguish them from student and MTurk samples. However, one of the quota samples

stands out as having worse acquiescence. While the natural explanation is satisficing, we used stricter screening of inattentive respondents here. Thus, sample clearly matters, though not in easily predicted or explained ways.

Figure 2: Meta-Regression Predicting the Magnitude of Acquiescence Bias



Note: the left panel shows coefficients from a meta-regression predicting the difference in correlations between aligned and unaligned measure. The right panel shows predicted values based on combinations of the sample and constructs involved.

Bias in Correlations with Alternative Scale Formats

Our analysis has focused on correlating pairs of scales measured using agree/disagree scales. However, our systematic literature review showed researchers often correlate agree/disagree scales with alternative formats. We briefly examine whether acquiescence bias has implications for correlations between agree/disagree scales and alternative formats by relying on Study 2 and three multi-item scales that were not measured using an agree/disagree scale

(none of our other studies include such scales). The first is expressive partisanship, a four-item scale measured on a four-point, fully-labeled scale with endpoints "always" and "never" (Huddy, Mason, and Aarøe 2015). We use this as a measure of partisan identity strength and ignore direction (e.g., Democrat). The second is trait aggression, a four-item scale measured on a six-point, fully-labeled scale with endpoints "completely true for me" and "completely false for me" (Kalmoe 2014). The third non-agree/disagree scale is support for political aggression. Each respondent evaluated up to three vignettes describing actions taken by a governor, then rated their support for five aggressive actions against that governor (e.g., throw eggs, punch in the face) using a fully-labeled, five-point scale ranging from "strongly support" to "strongly oppose." Respondents' scores average across all five items for all three vignettes.

To assess acquiescence bias's potential consequences, we estimate the correlation between each of the three non-agree/disagree scales and each half of the study's two balanced scales – hostile sexism and populism (note that the reversed items were recoded so higher values always indicate more sexism/populism). Consider two illustrative examples. Partisan identity strength correlates with hostile sexism either positively (r = .20) or negatively (r = -.14), depending on the set of items measuring hostile sexism. Trait aggression either positively (r = .16) or negatively (r = -.14) correlates with populism, depending on the populism scale used. On average, across the six possible comparisons, the average difference in correlations between the positively and negatively-worded scales is 0.34. Notably, the meta-analytic average for pairs of agree/disagree scales in Study 2 is 0.35, suggesting that the bias between agree/disagree and non-agree/disagree scales may be just as large as the bias between two agree/disagree scales. This raises important questions about whether these same patterns extend to correlations that do not

involve agree/disagree scales at all and the precise mechanisms behind these effects, questions for future research.

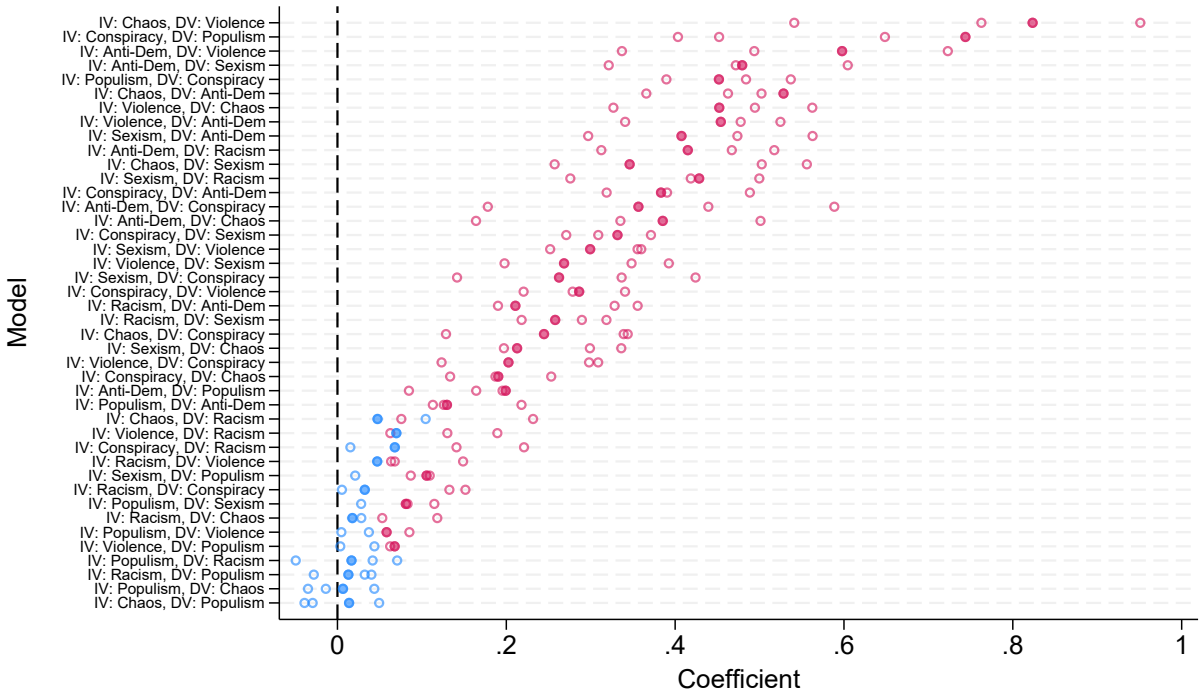
Acquiescence Bias's Myriad Substantive Consequences

We next use Study 4 to demonstrate acquiescence bias's applied consequences. Our systematic literature review revealed applied researchers correlating agree/disagree scales will often correlate two balanced scales, a balanced scale with a completely unbalanced scaled (PW items only), or two completely unbalanced scales (PW items only). Following theory and practice, we ask how such choices might affect estimated relationships. We estimate regression models using every pairwise combination of our seven variables, allowing all variables to serve as both independent and dependent variables. In each model, we control for partisanship, ideology, race, ethnicity, age, and gender. To examine the effects of measurement choices, we estimate four versions of each model: both variables measured with balanced scales, both measured with scales of only positively-worded items, or one balanced scale and the other positively-worded. This produces four coefficients for each of the 42 models.

To summarize our results, we focus on how much the coefficient point estimates vary by measurement choices and how this variability compares to sampling variability. Figure 3 shows the four coefficient estimates for each of the 42 models. The average coefficient size is 0.27, but within each of the 42 sets of models, the average difference between the smallest and largest coefficient is 0.18, or about 67% of the size of the average coefficient. Put differently, within each of the 42 sets of models, the standard deviation of the coefficient estimates is 0.08. By contrast, the average standard error for each coefficient is 0.03. In other words, the variation due to measurement choices is nearly three times as large as the variation due to sampling error.

Additionally, out of the 42 model sets, in 10 cases (24%), measurement choices affect whether the coefficient is statistically significant and in 5 cases (12%) the coefficient's sign. This analysis suggests that common measurement choices can have substantively important effects on conclusions reached in applied research.

Figure 3. How Measurement Choices Affect Regression Coefficients



Note: figure displays coefficients from regression models predicting the DV as a function of the IV and control variables. Each row represents a different pair of IV and DV, listed on the left-hand side. Within each row, the estimates correspond to models relying on one of four different measurement practices: both variables measured with balanced scales, both measured with only positively worded items, IV is balanced and DV measured with only PW items, and IV is measured with only PW items and DV is balanced. Filled markers represent the preferred model in which both scales are balanced. Red estimates are statistically significant, while blue estimates are not.

Measurement Models Help Address Acquiescence Bias Beyond Balanced Scales

Considering these substantive consequences, we next evaluate how correcting agree/disagree scales for acquiescence bias matters. Our systematic literature review revealed that scholars correlate myriad agree/disagree scales and rarely use recommended approaches to

address acquiescence bias at the analysis stage. We focus on the most common technique to attempt to purge correlations of acquiescence: structural equation modeling (SEM). The conventional strategy estimates separate trait (e.g., populism) and method (e.g., item type) factors (Billiet and McClendon 2000; Savalei and Falk 2014).

Our demonstration accomplishes two goals. First, we (re)introduce applied researchers to this technique. Second, we consider whether a design-stage choice of balanced scales sufficiently addresses acquiescence by comparing results with and without SEM corrections. We make no claim that observed changes reveal the true magnitude of acquiescence bias; rather, these analyses offer information on how controlling acquiescence at the analysis stage can alter substantive conclusions, conditional on meeting SEM assumptions like linearity and normality (Appendix 2). We focus on Studies 2-4, as respondents answered all scale items, enabling model estimation. It is worth emphasizing that we start with balanced scales in nearly all cases. Thus, even before attempting to statistically purge acquiescence bias via SEM, using balanced scales (when available) should reduce this bias (Paulhus 1991; Savalei and Falk 2014; Leiton 2021; cf. Messick 1991; McClendon 1991).

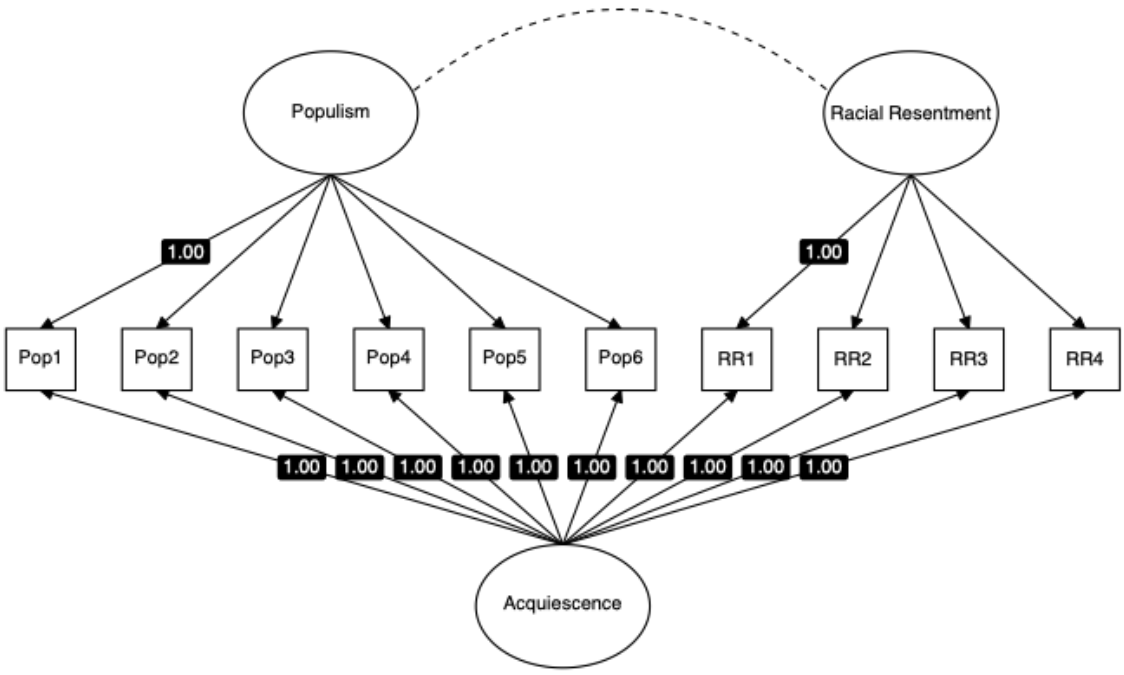
For each scale pair we estimate two models.[12] The first estimates a trait factor for each construct. The second model adds a method factor which we define using the items from both

[12] Appendix 4 reports scale validation information. Our approach makes assumptions about the relationship between scale items and the latent variable potentially inconsistent with a concept's conceptualization, with criterion validity consequences (Wuttke et al 2020). But with our interest in consequences from typical measurement and analysis decisions, and alternative scoring methods uncommon, we proceed accordingly.

scales (Bolt and Newton 2011; Savalei and Falk 2014). We allow items to load freely on their respective trait factors but fix each item's loading on the method factor to 1 (with all items recoded so higher values are pro-trait). This assumes acquiescence contributes equally to each item and enables model identification.

Figure 4 shows this second model visually using populism and racial resentment as an example. The items measuring populism and racial resentment are assigned to load on separate constructs, seen in the arrows directed from trait to item. All items are also assigned to load on a single construct, seen in the arrows directed from acquiescence to the items. Labeled paths all denote factor loadings fixed to 1 for identification. Unlabeled paths denote parameters to be estimated including factor loadings and the correlation between populism and racial resentment. Variance in the observed items is then a function of trait (populism or racial resentment) plus acquiescence bias. Left unlabeled, but assumed present in all observed items, is idiosyncratic measurement error that the model removes.

Figure 4: Visual Representation of Method Factor Model



Using both constructs to estimate a method factor facilitates more precisely recovering variance from acquiescent responding than estimating separate trait and method factors for a single construct (Bolt and Newton 2011).[13] We also fix the correlation between the method

[13] Effectively controlling for acquiescence using a method factor requires reversed items. Either the scale contains some or other items exist in the questionnaire with common method but different substance. This approach therefore likely underestimates how much acquiescence affects correlations between completely unbalanced scales because the method factor cannot clearly identify acquiescence. But it can still control for variation that affects all items to the same degree which arguably partially captures acquiescence (Weijters, Geuens, and Schillewaert 2010).

factor and each trait factor to 0.[14] Differences in construct correlations across models show whether corrections are consequential and, because most inputs are balanced scales, offer information on whether they sufficiently mitigate acquiescence bias's validity consequences.

Figure 5 shows the relationships between each pair of constructs when estimated without a method factor (x-axis) and with a method factor (y-axis).[15] The figure shows that including a method factor alters the estimated correlation and to varying degrees. The average absolute change across studies and construct pairs is 0.06 (min=.01, max=0.24, SD=0.06).[16] It is smallest in study 4 (0.04) and largest in study 2 (0.14). As important, direction of change varies: 52% of pairs see *stronger* relationships after including a method factor, a pattern consistent across studies (smallest: 50% in study 2; largest: 53% in study 3). Put concretely, in study 2, the relationship between hostile sexism and populism flips between negative and significant (r=-0.09; no method factor) to positive and significant (r =0.09; method factor). In studies 3 and 4, hostile sexism's correlation with racial resentment remains positive but increases from 0.50 to 0.63 and 0.43 to 0.56 after including a method factor. But hostile sexism's associations are not

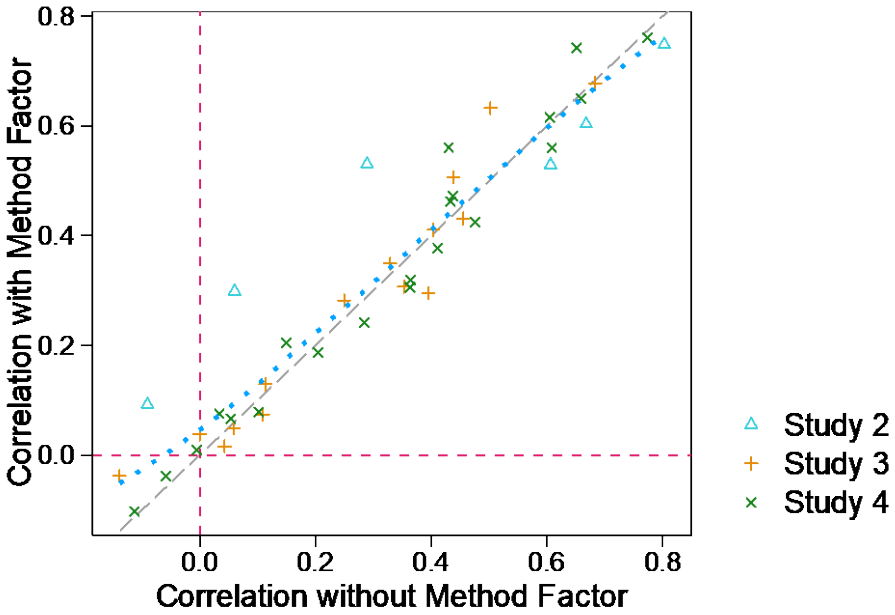
[14] We find allowing the trait and method factors to correlate our does not substantively affect our results. Nor does defining the method factor with all study agree/disagree items, not just those in the two constructs being considered, matter. Parsimony favors the present approach.

[15] All models correct for random error, so the estimated relationships in the baseline models excluding a method factor can differ from the bivariate correlations reported above.

[16] The average difference for correlations between balanced scales is .04 (SD=.04), with these changes somewhat more likely to be increases (54%) than decreases.

universally underestimated—its correlation with need for chaos decreases in studies 3 and 4 (0.40 to 0.30 and 0.36 to 0.31).

Figure 5: Construct pair correlations estimated via SEM with and without corrections for method effects.



Note: scale correlations from SEMs that vary in inclusion of method factor. The gray line represents identical correlations regardless of method factor estimation. The blue line is a loess curve showing the average deviation away from identical correlations.

What explains these differences in the magnitude of correlations and how much acquiescent responding may have affected the naïve correlation? The meta-regression suggests possible construct-specific contributions for acquiescence (see also Ferrando, Morales-Vives, and Lorenzo-Seva 2016). Alternatively, measure quality might matter. Although construct assessment overwhelmingly produced well-fitting CFA models with large, standardized factor loadings across studies, some scales faced problems. While model fit usually remained acceptable (though not ideal, Brown 2015), some scales had items with standardized factor loadings below 0.50 (Appendix 4). But we find no unique connection to positively- or negatively-worded items.

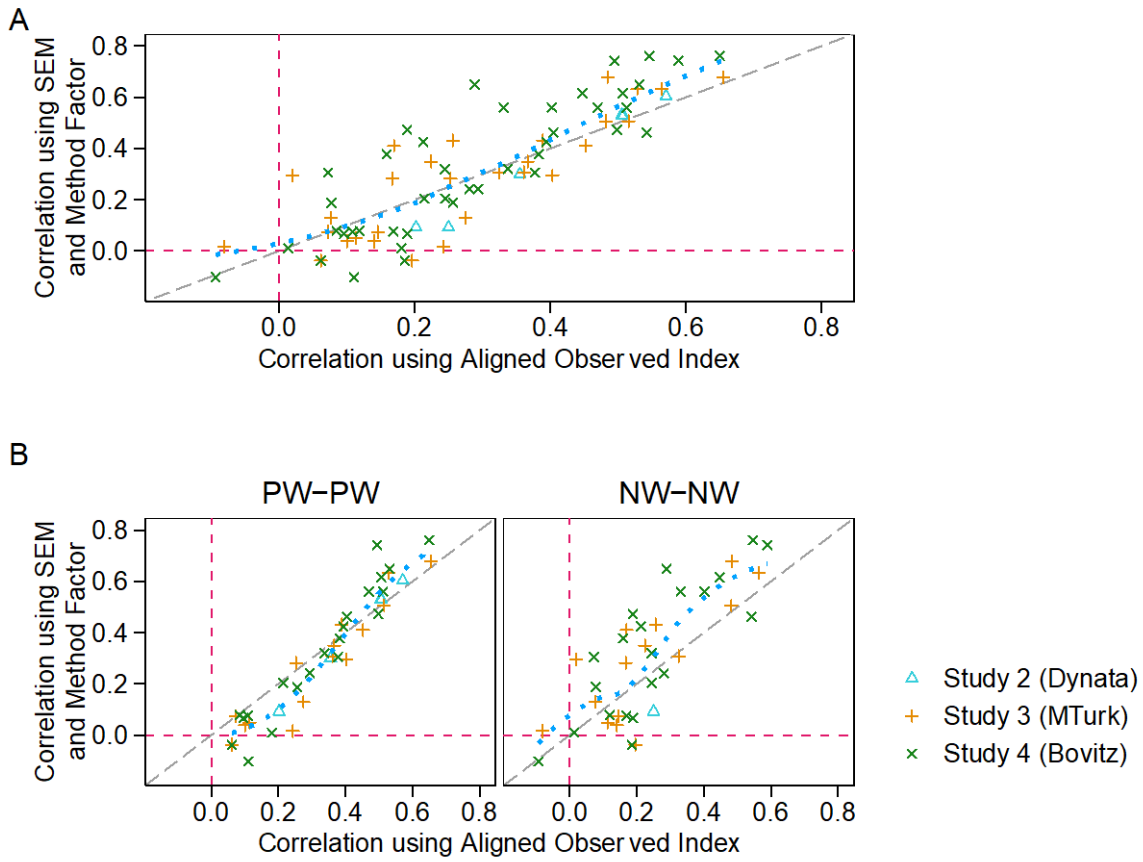
Even when correlating balanced scales, statistically accounting for acquiescence bias can meaningfully change both the magnitude of estimated relationships and the sign of these estimates compared to a model controlling for random error but not acquiescence. We suspect that adding information from balanced, rather than unbalanced, scales improves estimates by better identifying the method factor, which future work can investigate.

Design- and Analysis-Stage Decisions Shape Substantive Conclusions

We conclude by addressing how much ignoring acquiescence bias at *both* the design and analysis stage and correlating only aligned scales might influence conclusions. We consider how much the correlation between two variables changes when using additive indexes comprised only of directionally-aligned items (e.g., PW-PW), the typical analytical approach our systematic literature review revealed, or balanced scales and estimated as SEMs including a method factor.

Figure 6A visualizes this difference in correlations. The x-axis reports the correlation between two constructs using the aligned additive indices while the y-axis reports the correlation using the balanced scales and estimated via SEM to control acquiescence bias. Correlations calculated with the latter approach increase and decrease relative to the estimate obtained from the aligned additive indices, and to different degrees. Forty-eight percent increase from the aligned correlation. Further, the average absolute difference in correlations is 0.10 (min=0.002, max=0.36, SD=0.08). With the average absolute correlation across construct pairs about 0.30, a 0.10-point change constitutes moving from the 60th percentile to either the 40th or 75th percentiles of observed effect sizes (Lovakov and Agadullina 2021).

Figure 6: Correlation Differences by Estimation Strategy and Scale Direction



Note: scale correlations for construct pairs using the typical approach of aligned, summed scales and balanced scales corrected for acquiescence bias in a SEM. Panel A combines all aligned scales while panel B separates by scale direction. The gray line represents identical correlations regardless of method factor estimation. The blue line is a loess curve showing the average deviation away from identical correlations.

Figure 6B separates the aligned correlations into PW-PW and NW-NW pairs. More correlations increase among the NW-NW pairs (60%) than the PW-PW pairs (39%) when using the measurement model. We likewise find differences in the magnitude of the change in correlations. For the PW pairs, the average absolute difference in correlations is 0.07 (min=0.002, max=0.25, SD=0.06), with this doubling to 0.14 (min=0.004, max=0.36, SD=0.09) for the NW pairs.

But while one might conclude this demonstrates NW items contain more error and should be ignored for PW items, we believe this is misguided. First, information reported in the appendix (4 and 5) does not uniformly point to NW items and associated subscales as consistently poor quality. Second, simply correlating observed PW scales together provides misleading insights relative to pairing balanced scales and SEM.

Additional analyses (appendix 8) highlight two further points of caution for applied researchers. First, consistent with the meta-regression results, we find variation across constructs in these correlation differences. Researchers cannot assume their results necessarily underestimate true effects. For instance, the SEM approach produces stronger correlations for 63% of relationships with anti-democratic attitudes and 21% for those including populism. Second, the magnitude of change is not trivial. While the absolute difference in correlations is 0.07 points for the PW pairs, given one could easily either under- or overestimate the relationship between variables using only PW items the range of possible correlations is 0.14 points. Add to this sampling error and this range further increases.

Conclusion

While a longstanding concern, political scientists often overlook acquiescence bias. Established surveys regularly include unbalanced agree-disagree scales, such as in the World Values Survey (e.g., sexism), European Social Survey (e.g., authoritarianism), Comparative Study of Electoral Systems (e.g., populism), Cooperative Election Study (e.g., sexism), and International Social Survey programme (e.g., nationalism). Our systematic literature review shows that scholars have largely ignored acquiescence bias's consequences for assessing relationships between constructs, which we show can be substantial. Evidence from four original

studies and 160 tests indicates that correlations between constructs can change dramatically, both in substantive magnitude and sign, depending on scale direction. Ignoring acquiescence bias – as political scientists often do – may lead researchers to misleading substantive findings.

We also found that balanced scales may not sufficiently address acquiescence bias. Statistically correcting for it via SEM consistently altered estimated relationships between balanced scales, but sometimes in unexpected directions. Notably, this is not due to uniquely error-prone NW items. Correlations using only PW items also consistently differ from estimates derived from SEM, sometimes underestimating and sometimes overestimating the relationship. Together, these results suggest that it is important to correct statistically for acquiescence, even when using balanced scales (Leiton 2021).

These findings require reconsidering insights from the growing literature on anti-establishment orientations. We do not dispute this research agenda's importance or question which constructs underpin endorsement of political violence and anti-democratic tendencies. Instead, we urge scholars interested in studying relationships between these constructs to use measures designed to address acquiescence and, if necessary, pair them with appropriate statistical modeling techniques (e.g., Bartels 2020). Design and analysis choices matter.

Recommendations for Applied Researchers

Our findings point to recommendations at both the design and analysis stage, recognizing that researchers may be limited by the available data.

- When possible, adopt alternative question formats, such as forced choice or item-specific.
- When using an AD scale, use a balanced set of items, which may require the careful construction of new negatively-worded items.
- Do NOT remove negatively-worded items solely due to poor fit statistics.

- Control for methods effects in balanced scales using SEM (e.g., Savalei and Falk 2014).

While we recommend using alternative question formats, we also recognize they are not bias free. Indeed, our own analyses find similar levels of bias in criterion validity when correlating an agree-disagree scale with an item-specific scale as when correlating two agree-disagree scales. This suggests that scale direction alone (rather than *agreement* specifically) may play an important role in acquiescence bias. However, other work directly comparing agree-disagree scales to item-specific scales finds significantly lower method bias with item-specific scales (Kuru and Pasek 2016). Forced-choice formats may be particularly effective at removing shared response bias but they require balanced options and can have lower reliabilities than rating scales (Ray 19990; Kreitchmann et al 2019). Overall, the alternatives to agree-disagree scales seem superior but further work is needed to identify and control method biases in these scales.

Scholars using balanced scales face some additional obstacles. They may need to write new negatively worded items and, if so, we encourage pre-testing these and validating them with the existing items. The direction of items should be balanced with respect to the substantive content to avoid inadvertently introducing bias from alignment between direction and content. We emphasize validation should assess the whole scale, where question direction joins face validity as reasons to retain items even with low factor loadings (Brown 2015). At the same time, we note that violating SEM assumptions of multivariate normality and linearity can introduce bias (Kline 2023). But typically these can be addressed by selecting appropriate estimation strategies. SEM's specific advantage is in removing known systematic error, conditional on these assumptions.

Testing hypotheses requires not only clear causal identification strategies and proper sampling techniques, it also requires clearly capturing one's concepts of interest. If survey measures capture both trait and response set, then careful causal identification or sophisticated modeling strategies cannot provide appropriate answers about relationships between constructs. As should be clear, addressing acquiescence bias is both important and doable at both the design and analysis stages.

References

Akkerman, Agnes, Cas Mudde, and Andrej Zaslove. 2014. "How Populist Are the People? Measuring Populist Attitudes in Voters." *Comparative Political Studies* 47(9): 1324–53.

Archer, Allison M. N., and Scott Clifford. 2021. "Improving the Measurement of Hostile Sexism." *Public Opinion Quarterly*.

Armaly, Miles T., David T. Buckley, and Adam M. Enders. 2022. "Christian Nationalism and Political Violence: Victimhood, Racial Identity, Conspiracy, and Support for the Capitol Attacks." *Political Behavior* 44(2): 937–60.

Armaly, Miles T., and Adam M. Enders. 2022. "Who Supports Political Violence?" *Perspectives on Politics*: 1–18.

Aytac, S. Erdem, Ali Carkoglu, and Ezgi Elci. 2021. "Partisanship, Elite Messages, and Support for Populism in Power." *European Political Science Review* 13(1): 23–39.

Bakker, Bert N. and Ypthach Lelkes. 2018. "Selling Ourselves Short? How Abbreviated Measures of Personality Change the Way We Think About Personality and Politics." *Journal of Politics* 80(4): 1311-1325.

Baumgartner, Hans, and Jan-Benedict E.M. Steenkamp. 2001. "Response Styles in Marketing Research: A Cross-National Investigation." Journal of Marketing Research 38(2): 143–56.

Bartels, Larry M. 2020. "Ethnic Antagonism Erodes Republicans' Commitment to Democracy." *Proceedings of the National Academy of Sciences* 117(37): 22752–59.

Bartusevičius, Henrikas, Alexander Bor, Frederik Jørgensen, and Michael Bang Petersen. 2021. "The Psychological Burden of the COVID-19 Pandemic Is Associated With Antisystemic Attitudes and Political Violence." Psychological Science 32(9): 1391–1403.

Berinsky, Adam J, Michele F Margolis, and Michael W Sances. 2014. "Separating the Shirkers

from the Workers? Making Sure Respondents Pay Attention on Self-Administered Surveys." American Journal of Political Science 58(3): 739–53.

Berinsky, Adam J., Michele F. Margolis, Michael W. Sances, and Christopher Warshaw. 2019. "Using Screeners to Measure Respondent Attention on Self-Administered Surveys: Which Items and How Many?" *Political Science Research and Methods*.

Billiet, Jaak B., and Eldad Davidov. 2008. "Testing the Stability of an Acquiescence Style Factor Behind Two Interrelated Substantive Variables in a Panel Design." *Sociological Methods & Research* 36(4): 542–62.

Billiet, Jaak B., and McKee J. McClendon. 2000. "Modeling Acquiescence in Measurement Models for Two Balanced Sets of Items." Structural Equation Modeling: A Multidisciplinary Journal 7(4): 608–28.

Bolt, Daniel M., and Joseph R. Newton. 2011. "Multiscale Measurement of Extreme Response Style." *Educational and Psychological Measurement* 71(5): 814–33.

Bor, Alexander, Frederik Jørgensen, and Michael Bang Petersen. 2023. "The COVID-19 Pandemic Eroded System Support but Not Social Solidarity." PLOS ONE 18(8): e0288644.

Brotherton, Robert, Christopher C. French, and Alan D. Pickering. 2013. "Measuring Belief in Conspiracy Theories: The Generic Conspiracist Beliefs Scale." Frontiers in Psychology 4.

Brown, Timothy A. 2015. Confirmatory Factor Analysis for Applied Research. New York: Guilford Press.

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.

Campbell, Donald T, Carole R Siegman, and Matilda B Rees. 1967. "Direction-of-Wording

Effects in the Relationships Between Scales." Psychological Bulletin 68(5): 293–303.

Cassese, Erin C., and Mirya R. Holman. 2019. "Playing the Woman Card: Ambivalent Sexism in the 2016 US Presidential Race." *Political Psychology* 40(1): 55-74.

Cattell, Raymond B. 1944. "Psychological Measurement: Normative, Ipsative, Interactive." Psychological Review 51(5): 292–303.

Castanho Silva, Bruno et al. 2018. "Public Opinion Surveys: A New Scale." In *The Ideational Approach to Populism: Theory, Method & Analysis*, eds. Kirk Hawkins, Ryan E. Carlin, Levente Littvay, and Cristobal Rovira Kaltwasser. London: Routledge.

Castanho Silva, Bruno, Sebastian Jungkunz, Marc Helbling, and Levente Littvay. 2020. "An Empirical Comparison of Seven Populist Attitudes Scales." *Political Research Quarterly* 73(2): 409–24.

Chan, Vai. 2003. "Analyzing Ipsative Data in Psychological Research." Behaviormetrika 30(1): 99–121.

Clifford, Scott, Yongkwang Kim, and Brian Sullivan. 2020. "An Improved Question Format for Measuring Conspiracy Beliefs." *Public Opinion Quarterly*.

Cloud, Jonathan, and Graham M. Vaughan. 1970. "Using Balanced Scales to Control Acquiescence." *Sociometry* 33(2): 193–202.

Couper, Mick P., Roger Tourangeau, Frederick G. Conrad, and Chan Zhang. 2013. "The Design of Grids in Web Survey." *Social Science Computer Review* 31(3): 322–45.

Cunningham, William H., Isabella C. M. Cunningham, and Robert T. Green. 1977. "The Ipsative Process to Reduce Response Set Bias." The Public Opinion Quarterly 41(3): 379–84.

Elchardus, Mark, and Bram Spruyt. 2014. "Populism, Persistent Republicanism and Declinism: An Empirical Analysis of Populism as a Thin Ideology." *Government and Opposition*

51(1): 111–33.

Engelhardt, Andrew M, Stanley Feldman, and Marc J Hetherington. 2023. "Advancing the Measurement of Authoritarianism." Political Behavior 45(2): 537–60.

Farhart, Christina, Erin B. Fitz, Joanne M. Miller, and Kyle L. Saunders. 2023. "By Any Memes Necessary: Belief- and Chaos-Driven Motives for Sharing Conspiracy Theories on Social Media." *Research & Politics* 10(3): 20531680231193514.

Ferrando, Pere Joan, Fàbia Morales-Vives, and Urbano Lorenzo-Seva. 2016. "Assessing and Controlling Acquiescent Responding When Acquiescence and Content Are Related: A Comprehensive Factor-Analytic Approach." Structural Equation Modeling: A Multidisciplinary Journal 23(5): 713–25.

Gilbert, Daniel T., Romin W. Tafarodi, and Patrick S. Malone. 1993. "You Can't Not Believe Everything You Read." *Journal of Personality and Social Psychology* 65(2): 221–33.

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.

Golec de Zavala, Agnieszka, and Christopher M. Federico. 2018. "Collective Narcissism and the Growth of Conspiracy Thinking over the Course of the 2016 United States Presidential Election: A Longitudinal Analysis." *European Journal of Social Psychology* 48(7): 1011–18.

Green, Donald P., and Jack Citrin. 1994. "Measurement Error and the Structure of Attitudes: Are Positive and Negative Judgments Opposites?" *American Journal of Political Science* 38(1): 256–81.

Hibbing, Matthew V. et al. 2019. "The Relationship Between Personality and Response Patterns

on Public Opinion Surveys: The Big Five, Extreme Response Style, and Acquiescence Response Style." *International Journal of Public Opinion Research* 31(1): 161–77.

Hill, Seth J., and Margaret E. Roberts. 2023. "Acquiescence Bias Inflates Estimates of Conspiratorial Beliefs and Political Misperceptions." Political Analysis 31(4): 575–90.

Hussey, Ian, and Sean Hughes. 2020. "Hidden Invalidity Among 15 Commonly Used Measures in Social and Personality Psychology." Advances in Methods and Practices in Psychological Science 3(2): 166–84.

Joel, Samantha, John K. Sakaluk, James J. Kim, Devinder Khera, Helena Yuchen Qin, and Sarah C. E. Stanton. 2025. "Pseudo Effects: How Method Biases Can Produce Spurious Findings About Close Relationships." *Psychological Science* 36(10): 780–95.

Kalmoe, Nathan P. 2014. "Fueling the Fire: Violent Metaphors, Trait Aggression, and Support for Political Violence." *Political Communication* 31(4): 545–63.

Kalmoe, Nathan P., and Lilliana Mason. 2022. *Radical American Partisanship: Mapping Violent Hostility, Its Causes, and the Consequences for Democracy*. University of Chicago Press.

Kam, Cindy D, and Camille D Burge. 2018. "Uncovering Reactions to the Racial Resentment Scale across the Racial Divide." The Journal of Politics 80(1): 314–20.

Kertzer, Joshua D. 2022. "Re-assessing elite-public gaps in political behavior." *American Journal of Political Science* 66(3): 539-553.

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

Kreitchmann, Rodrigo Schames et al. 2019. "Controlling for Response Biases in Self-Report Scales: Forced-Choice vs. Psychometric Modeling of Likert Items." Frontiers in Psychology 10.

Krosnick, Jon A. 1991. “Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys.” *Applied Cognitive Psychology* 5(3): 213–36.

Krosnick, Jon, and Stanley Presser. 2010. “Question and Questionnaire Design.” In *Handbook of Survey Research*, eds. P.V. Marsden and J.D. Wright. Emerald, 263–314.

Kuru, Ozan, and Josh Pasek. 2016. “Improving Social Media Measurement in Surveys: Avoiding Acquiescence Bias in Facebook Research.” *Computers in Human Behavior* 57: 82–92.

Lechner, Clemens M., Melanie V. Partsch, Daniel Danner, and Beatrice Rammstedt. 2019. “Individual, Situational, and Cultural Correlates of Acquiescent Responding: Towards a Unified Conceptual Framework.” *British Journal of Mathematical and Statistical Psychology* 72(3): 426–46.

Leiton, Maria Fernanda Alvarado. 2021. “Using Balanced Scales to Address Acquiescent Response Style.” University of Michigan.

Lelkes, Yphtach, and Rebecca Weiss. 2015. “Much Ado about Acquiescence: The Relative Validity and Reliability of Construct-Specific and Agree–Disagree Questions.” *Research & Politics* 2(3): 205316801560417.

Liu, Mingnan, Sunghee Lee, and Frederick G. Conrad. 2015. “Comparing Extreme Response Styles between Agree-Disagree and Item-Specific Scales.” Public Opinion Quarterly 79(4): 952–75.

Lovakov, Andrey, and Elena R. Agadullina. 2021. “Empirically Derived Guidelines for Effect Size Interpretation in Social Psychology.” *European Journal of Social Psychology* 51(3): 485–504.

Malka, Ariel, Yphtach Lelkes, and Nissan Holzer. 2017. “Rethinking the Rigidity of the Right Model.” In The Politics of Social Psychology, eds. Jarret T Crawford and Lee Jussim. New

York: Psychology Press.

McClendon, McKee J. 1991. “Acquiescence and Recency Response-Order Effects in Interview Surveys.” Sociological Methods & Research 20(1): 60–103.

Messick, Samuel. 1991. “Psychology and Methodology of Response Styles.” In Improving Inquiry in Social Science, eds. Richard E. Snow and David E. Wiley. New York: Routledge.

Neuert, Cornelia E., Joss Roßmann, and Henning Silber. 2023. “Using Eye-Tracking Methodology to Study Grid Question Designs in Web Surveys.” Journal of Official Statistics 39(1): 79–101.

Oliver, J. Eric, and Thomas J. Wood. 2014. “Conspiracy Theories and the Paranoid Style(s) of Mass Opinion.” *American Journal of Political Science* 58(4): 952–66.

Pasek, Josh, and Jon A. Krosnick. 2010. “Optimizing Survey Questionnaire Design in Political Science: Insights from Psychology.” In *The Oxford Handbook of American Elections and Political Behavior*, ed. Jan E. Leighley. Oxford University Press, 27–49.

Paulhus, Delroy. 1991. “Measurement and Control of Response Bias.” In Measures of Personality and Social Psychological Attitudes, eds. John P Robinson, Phillip R Shaver, and Lawrence S Wrightsman. San Diego: Academic Press, 17–59.

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.

Piazza, James A. 2023. “Populism and Support for Political Violence in the United States: Assessing the Role of Grievances, Distrust of Political Institutions, Social Change Threat, and Political Illiberalism.” Political Research Quarterly.

Primi, Ricardo, Daniel Santos, Filip De Fruyt, and Oliver P. John. 2019. “Comparison of

Classical and Modern Methods for Measuring and Correcting for Acquiescence." *British Journal of Mathematical and Statistical Psychology* 72(3): 447–65.

Rammstedt, Beatrice, Christoph J. Kemper, and Ingwer Borg. 2013. "Correcting Big Five Personality Measurements for Acquiescence: An 18–Country Cross–Cultural Study." *European Journal of Personality* 27(1): 71–81.

Ray, John J. 1990. "Acquiescence and Problems with Forced-Choice Scales." *The Journal of Social Psychology* 130(3): 397–99.

Rollero, Chiara, Peter Glick, and Stefano Tartaglia. 2014. "Psychometric Properties of Short Versions of the Ambivalent Sexism Inventory and Ambivalence Toward Men Inventory." *Testing, Psychometrics, Methodology in Applied Psychology* 21(2): 1–11.

Saris, Willem E, Melanie Revilla, Jon A Krosnick, and Eric M Shaeffer. 2010. "Comparing Questions with Agree/Disagree Response Options to Questions with Item-Specific Response Options." *Survey Research Methods* 4(1): 61–79.

Savalei, Victoria, and Carl F. Falk. 2014. "Recovering Substantive Factor Loadings in the Presence of Acquiescence Bias: A Comparison of Three Approaches." *Multivariate Behavioral Research* 49(5): 407–24.

Schaffner, Brian F. 2021. "Optimizing the Measurement of Sexism in Political Surveys." *Political Analysis*.

Schulz, Anne et al. 2018. "Measuring Populist Attitudes on Three Dimensions." *International Journal of Public Opinion Research* 30(2): 316–26.

Schuman, Howard H., and Stanley Presser. 1981. *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context*. San Diego, CA: Academic Press.

Sprong, Stefanie et al. 2019. "'Our Country Needs a Strong Leader Right Now': Economic

Inequality Enhances the Wish for a Strong Leader." Psychological Science 30(11): 1625–37.

Spruyt, Bram, Gil Keppens, and Filip Van Droogenbroeck. 2016. "Who Supports Populism and What Attracts People to It?" *Political Research Quarterly* 69(2): 335–46.

Tarman, Christopher, and David O Sears. 2005. "The Conceptualization and Measurement of Symbolic Racism." The Journal of Politics 67(3): 731–61.

Uscinski, Joseph E., Casey Klofstad, and Matthew D. Atkinson. 2016. "What Drives Conspiratorial Beliefs? The Role of Informational Cues and Predispositions." *Political Research Quarterly* 69(1): 57–71.

Uscinski, Joseph E et al. 2021. "American Politics in Two Dimensions: Partisan and Ideological Identities versus Anti-Establishment Orientations." *American Journal of Political Science*.

Voelkel, Jan G. et al. 2023. "Interventions Reducing Affective Polarization Do Not Necessarily Improve Anti-Democratic Attitudes." Nature Human Behaviour 7(1): 55–64.

Weijters, Bert, and Hans Baumgartner. 2012. "Misresponse to Reversed and Negated Items in Surveys: A Review." *Journal of Marketing Research* 49: 737–47.

Weijters, Bert, Maggie Geuens, and Niels Schillewaert. 2010. "The Stability of Individual Response Styles." *Psychological Methods* 15(1): 96–110.

Westwood, Sean J, Justin Grimmer, Matthew Tyler, and Clayton Nall. 2022. "Current Research Overstates American Support for Political Violence." *Proceedings of the National Academy of Sciences*.

Wetzel, Eunike, Oliver Lüdtke, Ingo Zettler, and Jan R. Böhnke. 2016. "The Stability of Extreme Response Style and Acquiescence Over 8 Years." *Assessment* 23(3): 279–91.

Wuttke, Alexander, Christian Schimpf, and Harald Schoen. 2020. "When the Whole Is Greater

than the Sum of Its Parts: On the Conceptualization and Measurement of Populist Attitudes and Other Multidimensional Constructs." American Political Science Review 114(2): 356–74.

Young, Clifford, Chris Jackson, and Nicolas Boyon. 2023. "Defining and Validating the System Is Broken Index." Social Science Quarterly 104(1): 5–10.

Zhang, Xijuan, Ramsha Noor, and Victoria Savalei. 2016. "Examining the Effect of Reverse Worded Items on the Factor Structure of the Need for Cognition Scale." *PLOS ONE*.