When the Whole is Greater Than the Sum of its Parts: On The Concept and Measurement of Populist Attitudes and Other Multi-dimensional Constructs

Wuttke, Alexander; Schimpf, Christian H.; Schoen, Harald

Preprint / Preprint
Zeitschriftenartikel / journal article

When the Whole is Greater Than the Sum of its Parts:

On the Conceptualization and Measurement of Populist Attitudes and Other Multi-dimensional Constructs

Alexander Wuttke, Christian Schimpf, Harald Schoen

This is an unedited pre-print. Please cite the original study:
Wuttke, Alexander; Schimpf, Christian; Schoen, Harald (2020): '*When the Whole is Greater Than the Sum of its Parts: On the Conceptualization and Measurement of Populist Attitudes and Other Multi-dimensional Constructs*', American Political Science Review.

Replication Material: https://doi.org/10.7910/DVN/KPS1KY
Shiny Web Application: http://populism.alexander-wuttke.de
Supplementary Material: At the end of this document

Abstract

Multi-dimensional concepts are non-compensatory when higher values on one component cannot offset lower values on another. Thinking of the components of a multi-dimensional phenomenon as non-compensatory rather than substitutable can have wide-ranging implications, both conceptually and empirically. To demonstrate this point, we focus on populist attitudes that feature prominently in contemporary debates about liberal democracy. Given similar established public opinion constructs, the conceptual value of populist attitudes hinges on its unique specification as an attitudinal syndrome, which is characterized by the concurrent presence of its non-compensatory concept subdimensions. Yet this concept attribute is rarely considered in existing empirical research. We propose operationalization strategies that seek to take the distinct properties of non-compensatory multi-dimensional concepts seriously. Evidence on five populism scales in twelve countries reveals the presence and consequences of measurement-concept inconsistencies. Importantly, in some cases, using conceptually sound operationalization strategies upsets previous findings on the substantive role of populist attitudes.

Giovanni Sartori; Concept formation; Democratic backsliding; Reflective latent variables; composite indices

Introduction

Populism is an essential social science concept, but it is also an essentially contested concept (Mudde 2017). Despite ongoing disagreement about its content, origins, and consequences (e.g., Aslanidis 2016; Rooduijn 2014), populism is en vogue, and the term is widely used among public intellectuals, politicians and academic scholars. In the academic realm, a recent strand of social inquiry investigates populism at the individual level (Akkerman, Mudde, and Zaslove 2014; Hawkins, Riding, and Mudde 2012). Often referred to as the study of "populist attitudes," this research starts from the assumption that populist ideas must resonate with the public to be influential (Hawkins and Kaltwasser 2018; Spruyt, Keppens, and van Droogenbroeck 2016). Therefore, scholars examine how populist discourses, styles, and strategies among political elites correspond with the distribution of populist ideas among ordinary citizens (e.g., Castanho Silva et al. 2018; Hameleers, Bos, and Vreese 2017; Hawkins, Riding, and Mudde 2012; Hieda, Zenkyo, and Nishikawa 2019; Schulz et al. 2018; Spruyt, Keppens, and van Droogenbroeck 2016). The promise of this line of research is that understanding populism at the individual level may help understanding populism at the societal level, thus, promoting the comprehension of how the "Populist Zeitgeist" (Mudde 2004) affects health and outlook of pluralist democracies.

In this study, we take a step back and consider some unresolved problems regarding the concept structure of populism at the individual level. Responding to criticism that populist attitudes do not provide any theoretical import to the established public opinion literature, populism scholars have devotedly been discussing the conceptual core of populist attitudes, thereby achieving a notable convergence about the concept's essential characteristics (Rooduijn 2019). Most scholars now concur that populist attitudes denote a multi-dimensional construct, comprised of two or more concept components (Akkerman, Mudde, and Zaslove

2014; Castanho Silva et al. 2018; Mudde and Kaltwasser 2013a; Oliver and Rahn 2016; Schulz et al. 2018; Stanley 2011). Although unanimity has not yet been reached about the exact content of these components, there is widespread agreement on the idea that populist attitudes lie at the intersection of the concept's sub-dimensions (Castanho Silva et al. 2018; Hameleers, Bos, and Vreese 2017, 482; Hieda, Zenkyo, and Nishikawa 2019, 3; March 2017, 283; Mudde and Kaltwasser 2013a; 2013b, 149; Spruyt, Keppens, and van Droogenbroeck 2016, 336, but see Hameleers, Bos, and Vreese 2017, 482).[1] In other words, the supposed unique property of populist attitudes is the co-existence of its components. Therefore, what distinguishes populist attitudes from similar attitudinal concepts is its status as an attitudinal syndrome, which only considers citizens as populists if, for instance, they accept anti-elitist views *and* a Manichean outlook *and* believe in unrestricted popular sovereignty. However, existing studies on populism at the mass level rarely transfer this crucial concept feature into empirical practice. Because of the apparent mismatch between the theoretically derived concept specification and the concept's operational use, many populism studies do not measure what they are intended to measure and, therefore, the reported results do not necessarily speak about the concept under investigation.

Although we focus on the concept of populist attitudes, this study also highlights the general importance of aligning a measure's mathematical structure with the target concept's theoretical structure for valid inferences in social inquiry (cf. Goertz 2006, 125) and thus has implications for other non-compensatory multi-dimensional concepts beyond research on populism. Using inadequate operationalization strategies may yield dubious results. For instance, after thorough discussions on measurement-concept inconsistencies in studies on democratization (Collier and Mahon 1993; Goertz 2006; Munck and Verkuilen 2002),

[1] In this study we use the terms concept components, attributes, and subdimensions interchangeably.

scholars began to derive non-compensatory measures of democracy in which countries could not substitute low levels of electoral fairness with high levels of minority protection because both components constitute necessary elements of liberal democracies (Møller and Skaaning 2012, 135). Similar arguments play a role in other research areas where the presence of a concept depends on the simultaneous presence of all concept components such as deprivation (Alkire et al. 2015), human development (Greco et al. 2019), or democratic support (Schedler and Sarsfield 2007).

Using populist attitudes as an illustrative case, this study shows that different operationalization strategies of multi-dimensional concepts may ultimately yield diverging substantive inferences. For the study of populism on the individual level, deriving adequate operation strategies is particularly important as the conceptual value of populist attitudes hinges on the specification as a multi-dimensional concept with non-substitutable subdimensions. This article thus demonstrates how simple adjustments of empirical practices enable the blossoming literature on populist attitudes to avoid measures that misrepresent the object under investigation. Importantly, the presented analytical framework and the operational tools for aligning concept and measurement also apply to other manifestations of populism, such as party strategies or communication styles (e.g., March 2017) and to other multi-dimensional concepts beyond populism research that is constituted by non-compensatory concept components.

To help overcome the inconsistency between theory and research practice, this article proceeds in several steps. Because any attempt to derive valid measures requires a solid understanding of the target concept, we first review debates about the essence of populist attitudes. We show that the idea proposed by populism scholars to ensure the concept's field utility is largely neglected where the concept is put into empirical practice. In a broader vein,

we then argue that current empirical research relies on a paradigm for operationalizing latent constructs, which–despite its prevalence in public opinion research–is not suitable for multi-dimensional concepts with non-interchangeable concept components. Therefore, we draw on classical and modern approaches to concept formation to discuss strategies for the operationalization of non-compensatory concepts that enables us to take this distinct characteristic of populist attitudes seriously. Using data from twelve countries and five scales of populist attitudes, we show substantial disparities between individual populism scores derived from the existing methods compared to the new approach. Moreover, we demonstrate that these methods can lead to diverging conclusions about the relationship of populist attitudes with substantive variables of interest.

What are "Populist Attitudes"?

Arguably, no other question surrounding populism has achieved as much attention as the question of what populism is. This inquiry can be divided into two sub-issues. First, there are the theoretical propositions that scholars associate with populism as a concept. Among other things, populism has been defined as mass movement or distinct form of mobilization (Barr 2009; Jansen 2011; Kenny 2017), as a discourse (Laclau and Mouffe 1985), a thin-centered ideology (Mudde 2004), a frame (Aslanidis 2016), a style (Moffitt and Tormey 2014), a strategy or organization (Weyland 2001), a set of ideas (Hawkins 2009), as a form of political representation (Caramani 2017), or a conception of democracy (Pappas 2016). Accordingly, there is no consensus about the essence of the concept. Second, scholars must also define the set of attributes that constitute populism. Depending on how populism is conceived of, the concept is characterized by numerous combinations from a set of at least a dozen different attributes, ranging from anti-elitism over polarization to the centralization of leadership and

simplistic language (Rooduijn 2014, 578). Altogether, there is ongoing uncertainty about populism's conceptual core.

However, notable conceptual convergence is apparent in scholarship on populism at the mass level. Most studies of populist attitudes (e.g., Akkerman, Mudde, and Zaslove 2014; Mohrenberg, Huber, and Freyburg 2019; Schulz et al. 2018; Spierings and Zaslove 2017) follow the ideational approach (Hawkins et al. 2018) and start from the definition of populism as a so-called "thin-centered ideology" (Mudde 2004, 543; 2007, 23). Moreover, scholars mostly agree that populist attitudes consist of two or more essential components or subdimensions (Akkerman, Mudde, and Zaslove 2014; Castanho Silva et al. 2018; Müller 2017; Schulz et al. 2018; Stanley 2011). Anti-elitism is usually on the list of core dimensions (e.g., Castanho Silva et al. 2018; Mudde and Kaltwasser 2013a; Schulz et al. 2018, see supplement 1). Disagreement persists about the exact definitions of the remaining components (Quinlan and Tinney 2019), for which scholars have suggested, for instance, support for popular sovereignty (Schulz et al. 2018) or Manichean outlook on society (Castanho Silva et al. 2018). In any case, despite disagreement about the number and denomination of concept components, the vast majority of scholars perceive populist attitudes as a multi-dimensional construct comprised of anti-elitist attitudes and further orientations about the role of the people.

Yet this conception of populist attitudes faces questions about the concept's theoretical import. Even more than other branches of populism research, the study of populism at the individual level is embedded in a field of inferences where many similar concepts already exist (Geurkink et al. 2019). In the public opinion literature, scholars for long have examined anti-elitist orientations under the rubrics of political cynicism (Agger, Goldstein, and Pearl 1961), efficacy (Niemi, Craig, and Mattei 1991), trust (Miller 1974) or support (Easton 1975).

Similarly, orientations toward popular sovereignty and homogeneity have been examined, for instance, in the national identity (e.g., Mader et al. 2018) and ethnocentrism (e.g., Bonikowski and DiMaggio 2016) literatures or by studies on citizens' process preferences (Schedler and Sarsfield 2007) and orientations toward representational roles (e.g., Katz 1997). This list could be continued.

Importantly, the substantive overlap with existing public opinion constructs undermines the conceptual value of populist attitudes. Applying the criteria developed by Gerring (1999) to evaluate concepts in the social sciences, concept differentiation, and field utility suffer severely if populist attitudes are not more than the sum of established constructs. Concept differentiation refers to "the clarity of [concept] borders within a field of similar terms" (Gerring 1999, 376). Thus, concept differentiation is low when a concept cannot be clearly distinguished from related constructs. Furthermore, conceptual overlap undermines field utility because populist attitudes may "steal referents from neighboring terms, leaving these terms as empty" (Gerring 1999, 383). Hence, if populist attitudes are to be added to the list of valuable public opinion constructs, the concept must bring to the table a theoretical proposition that is distinct and new.

In response to these challenges, populism scholars uphold that the distinct conceptual characteristic of populism at the mass level is the concept's status as an attitudinal syndrome which is defined in terms of necessary and sufficient conditions, more specifically by the simultaneous presence of the concept's constituent components (Akkerman, Mudde, and Zaslove 2014, 1326; Castanho Silva et al. 2018; Elchardus and Spruyt 2016, 113, 120; Spierings and Zaslove 2017, 824; Spruyt, Keppens, and van Droogenbroeck 2016, 336, for a rare exception see Hameleers, Bos, and Vreese 2017, 482). In the words of Hawkins and Kaltwasser, the "peculiarity of the populist set of ideas lies precisely in the combination of

these elements" (Hawkins and Kaltwasser 2018, 6). By implication, when populist attitudes lie at the intersection of the concept components, individual populism scores cannot be high when anti-elitist orientations are low even when a person strongly supports the remaining components of populism. For instance, assuming a three-dimensional populism concept, understanding populist attitudes as an attitudinal syndrome suggests considering citizens as populists only if they exhibit anti-elitist orientations *and* a Manichean outlook *and* support popular sovereignty. Using a concept specification that treats the concept components as non-compensatory, i.e., as jointly necessary for the presence of populism at the individual level, the concept of populist attitudes indeed would represent more than the sum of its parts.

However, even though many scholars implicitly or explicitly specify populist attitudes as an attitudinal syndrome with non-compensatory concept components at the theoretical level, few studies consider this property when applying the construct empirically (for exceptions, see Mohrenberg, Huber, and Freyburg 2019; Vehrkamp and Wratil 2017). In many cases, the concept operationalization does not respect the necessary conditions. Instead, scholars obtain individual populism scores by computing weighted or unweighted averages across the concept dimensions. More specifically, scholars aggregate the concept attributes manually by computing mean values or using data-driven approaches such as factor analysis (e.g., Spierings and Zaslove 2017, 831; Spruyt, Keppens, and van Droogenbroeck 2016, 340). Regardless of the level of methodological sophistication, however, these aggregation methods have in common that higher values on one concept component can compensate for low values on other subdimensions. These operationalization strategies may, for instance, assign high populism scores to citizens who do not show slightest signs of anti-elitist sentiments. In contrast to the concept's theoretical specification, these approaches might identify citizens as "populists" who strongly oppose giving power to the people or who strongly favor elitist rule.

Hence, compensatory aggregation rules neglect the one theoretical proposition of the concept that is supposed to make it distinct from similar concepts, i.e., its status as an attitudinal syndrome at the intersection of the concept components.

In the following, we argue that this compensatory operationalization approach is rooted in a measurement paradigm that is often applied to latent constructs but which is incompatible with constructs that presuppose necessary conditions as essential concept properties.

On the relationship between concepts and concept components

Concepts with two or more components such as populist attitudes can be understood as multi-level constructs with the concept's essential core at the basic level, the concept subdimensions or concept components at the second, and the indicators at the third level (Goertz 2006). The psychometric literature on populist attitudes is usually concerned with the indicator level, investing considerable efforts into identifying adequate item batteries to capture the construct of interest (Akkerman, Mudde, and Zaslove 2014; Castanho Silva et al. 2018; Hawkins, Riding, and Mudde 2012; Schulz et al. 2018; van Hauwaert, Schimpf, and Azevedo 2019). As long as the existence of second-level concept components does not complicate the operationalization of an attitudinal construct, the public opinion literature often views the unobservable concept of interest as the common cause of the observed measures at the indicator level. The general idea underlying this perspective, often associated with Bollen and Lennox (1991), is to view entities at one level as the effects of entities at the other level. Even though causality may go both ways,[2] the causal approach usually perceives observed measures as reflections of the unobservable concept.

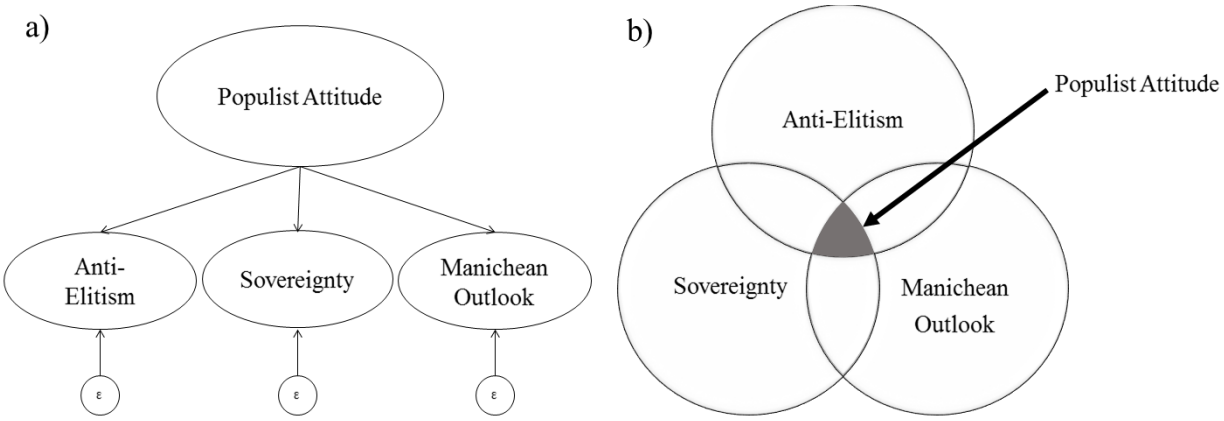
[2] The essentialist approach (Figure 1, panel b) partly overlaps with the concept of formative indicators. However, the fact that Bollen and Lennox (1991) call them causal indicators emphasizes the temporal and causal order that is assumed to underlie the relationship between concept and formative indicators. Because we do not see anti-

The perspective of reflective causality has empirical implications. Because the observed indicators are seen as effects of a common cause (the latent concept), the indicators are expected to correlate with each other (Bollen and Lennox 1991, 306; Edwards and Bagozzi 2000, 158). Practically, this perspective suggests the use of factor analyses to assess the extent of inter-item correlation and to exclude indicators with unsatisfactorily low correlations. Methodologically, this approach regards differences between constructs at the same level as measurement error. Measurement error stems from unrelated constructs that exert influences in addition to the common latent construct and would cancel out by averaging across all constructs at one level. Consequently, the causal perspective views disparities between lower-order concepts as subordinate statistical entities without substantive meaning.

The causal perspective is common in the literature on populist attitudes where it is applied to the indicator level but –implicitly or explicitly– also to describe the relationship between the concept and the concept components (e.g., Castanho Silva et al. 2019; Hieda, Zenkyo, and Nishikawa 2019; Schulz et al. 2018). What we call the Bollen approach to the operationalization of populism at the individual level considers populist attitudes as the common cause of the concept components, such as anti-elitist orientations and support for popular sovereignty (Figure 1, Panel A). Because the Bollen approach views the concept components as partly interchangeable, this perspective implies geometric or linear functions (summary scores, factor analyses) to aggregate concept components into composite indices (i.e., an individual score of populist attitudes). Empirically, the Bollen approach is the most common operationalization approach in the literature on populist attitudes (see Supplement 1 for an overview).

elitism as *causing* populist attitudes, we deem the essentialist approach a better conceptual fit than the causal approach including its formative variant.

Figure 1. **The causal and the essential perspectives on the relationship between concept and concept attributes**



Note: Similar illustrations were shown by Wong, Law, and Huang (2008) and Castanho Silva et al. (2018).

Even though the causal paradigm serves well for most latent constructs, we argue that it is inept for non-compensatory multi-dimensional constructs where each component constitutes a necessary condition for the presence of the concept. First, even though the causal perspective prescribes a correlation between the concept components, no theoretical argument implies such a correlation between the subdimensions in the case of populist attitudes. We do expect the components to overlap among some individuals as these are the individuals who hold populist attitudes (see Figure 1, panel B). Yet we also expect that there are non-populist individuals who agree with none or only some of the components of populist attitudes. Hence, the concept of populist attitudes as an attitudinal syndrome describes attitudinal configurations *among individuals,* but it is agnostic about correlations *between the concept attributes*. Second, disregarding large differences between the attributes merely as a measurement artifact is incompatible with the unique property of populist attitudes as an attitudinal syndrome for which low levels on one subdimension cannot be compensated for by higher levels on another one.

Altogether, these arguments suggest that the Bollen approach does not adequately characterize the relationship between the concept and its components in the case of multi-

dimensional concepts with non-substitutable concept components such as populist attitudes. Instead of assuming that populist attitudes *cause* anti-elitist orientations and other orientations (Figure 1, Panel A), a more intuitive understanding of the concept suggests that populist attitudes *are* anti-elitist orientations in combination with other orientations (Figure 1, Panel B). Put differently, the relationship between the concept and the concept component is not a causal one but one of essence and identity.

Aggregation rules for non-compensatory concepts such as populist attitudes

The identity relationship between the concept and the concept components paves the way for alternative operationalization strategies for non-compensatory concepts in which the aggregation rules are more closely aligned with the concept's theoretical propositions. Having established that interchangeable subdimensions are incompatible with the unique concept property of jointly necessary concept components (qualifier of concept structure), different concept structures are conceivable (see Table 1). In the following, we contrast two prototypical concept structures that differ in the sharpness of the membership boundaries, i.e., the quantifier of concept structure. The dichotomous quantifier presupposes clear boundaries and sets an all-or-nothing structure for the concept. The continuous quantifier, in contrast, presupposes grey space between the concept poles, indicating fine-grained differences between entities.[3] Hence, the quantifier defines whether the concept of populist attitudes separates populists from non-populists or whether we distinguish individuals by the degree to which they hold populist attitudes. If populist attitudes should be assessed with an either-or assertion or with a more-or-less assertion is not a question of right-or-wrong. Instead, the quantifier reflects the researcher's conception of the construct.

[3] For a similar discussion of "degreeism" concerning populism or populist attitudes, see Aslanidis (2016), van Kessel (2015), Pappas (2016), and van Hauwaert and van Kessel (2018).

Table 1. **Prototypical concept structures of populist attitudes**

		Quantifier of concept structure	
		Dichotomous	*Continuous*
Qualifier of Concept Structure	*Non-compensatory*	Sartori	Goertz
	Compensatory	Residual	Bollen

Note: In practice, quantifier and qualifier have more than two manifestations. The prototypical concept structures shown in Table 1 contrast no substitutability with medium substitutability (qualifier), yet other degrees of substitutability are also conceivable. Likewise, Table 1 contrasts two prototypical quantifiers, although ordinal quantifiers are also conceivable.

A dichotomous concept structure of populist attitudes, which takes the necessary conditions among the subdimensions into account amounts to a *Sartorian concept structure*. Even though rarely employed in the populist attitude literature (but see Vehrkamp and Wratil 2017), we can draw on classical approaches to concept formation for its operationalization (Sartori 1970; 1984). Specifically, the necessary conditions can be operationalized by setting thresholds among the attributes (see Supplement 2 for a comprehensive discussion of operationalization strategies). For example, one may specify that respondents need to agree or strongly agree on all anti-elitism- and popular-sovereignty-items to be counted as populists. If participants disagree with one or more items, they would be considered non-populists. The intuitive appeal of such thresholds is the transparency of the classification. This approach is particularly appealing if the cut-off points have informational value (e.g., above the mean or at the scale mid-point). Yet obvious cut-off points are often unavailable which leads to an increased risk of inducing arbitrary choices or misuse of analytical discretion (Wuttke 2019). More fundamentally, human evaluations of any given entity usually differ by degree. Therefore, based on the reasoning that continuous measures better resemble the mental representations of attitudinal constructs, cognitive sciences have generally departed dichotomous measures of attitudes (Murphy 2002). Against this backdrop, we conclude that the Sartorian concept

structure is a reasonable approach, but practical and epistemological reasons speak against dichotomous measures as a standard choice for research on populist attitudes.

The *Goertz concept structure* merges elements of both approaches to concept formation already discussed. Building on work by Gary Goertz (2006), this concept structure employs Fuzzy Logic to combine the conceptual rigorousness of necessary attributes with the operational flexibility of continuous outcomes. Hence, like the Bollen approach, the Goertzian concept structure views individuals as differing by the degree to which they hold populist attitudes. Yet unlike the Bollen construct, the Goertzian approach does not classify individuals as high on populism if they disagree with one core component (e.g., anti-elitism). Instead, the Goertz construct structure requires the acceptance of all essential concept attributes for high populism scores. Rarely employed in practice (but see Mohrenberg, Huber, and Freyburg 2019), several options are available for operationalizing Goertzian constructs (see Supplement 2), the easiest of which is to use the minimum value of the concept subdimensions.[4] The minimum represents the logical equivalent of the intersection in set theory. Using the minimum may appear statistically inefficient at first as this operationalization seems to disregard the information on all but the lowest attribute. However, as the mathematical representation of the minimum function shows (see further below), the computation of the minimum, in fact, does consider all attributes when identifying the attribute with the lowest value. More importantly, the minimum is the mathematical equivalent of the theoretical proposition that the lowest attribute determines the overall concept (cf. Goertz 2006, 138).

[4] Different aggregation strategies are conceivable, depending on the assumptions regarding the components' substitutability, see Møller and Skaaning (2012, 122ff); Goertz (2006, 111ff); Munck (2009, 48ff). The minimum is the adequate operationalization under the assumption of non-compensatory and non-interactive attributes. However, if the concept implies some interchangeability or interaction of the attributes, then the researchers may opt for more flexible solutions (e.g., weighted arithmetic mean, multiplication, geometric mean), which closely resemble the structure of necessary conditions by emphasizing the attributes with the lowest value but also giving some weight to all other observed attributes; see Supplement 2.

Consequently, the Goertzian concept structure accounts for necessary conditions among the concept components while distinguishing varying degrees of accepting populist attitudes.

Altogether, we distinguish three approaches to operationalizing multi-dimensional concepts such as populist attitudes. These concept structures map the construct on different scales (quantifier), and the operationalization strategies reflect different perspectives on the substitutability of the concept components (qualifier). The conceptual analysis demonstrated that the Bollen operationalization fails to capture the unique theoretical proposition of populist attitudes. We presented two operationalization strategies that account for the non-compensatory relationship of the subdimensions, thereby taking populism seriously as an attitudinal syndrome. Among those operationalization strategies, the Goertz concept structure has the advantage of reflecting the nuances of human attitudes. Accordingly, epistemological and methodological reasons suggest a preference for the Goertz concept structure when operationalizing populist attitudes.

For empirical research, it is crucial whether this result of our analysis makes a difference for substantive conclusions. As these approaches employ different aggregation rules for combining the concept components into individual construct scores, it is plausible to expect that they yield different results. However, this cannot be taken for granted. We, therefore, turn to empirical analysis to explore whether the operationalization strategies matter for substantive findings.

Data and operationalization

In our analysis of commonalities and disparities between the empirical representations of the Sartori, Bollen, and Goertz concept structures, we focus on three scales of populist attitudes: the populism scales by Schulz et al. (2018) and Castanho Silva et al. (2018) and Akkerman, Mudde, and Zaslove (2014). Due to limitations of space (see Supplement 5 on scale selection),

results on additional scales of populist attitudes (CSES scale, scale by Oliver and Rahn 2016) are reported in an interactive Shiny web application. Supplement 6 provides further practical information on using these scales to assess populist attitudes as an attitudinal syndrome.

All three scales analyzed in this study conceptualize populism at the individual level as multi-dimensional with three concept components and view anti-elitism as an essential element of populist attitudes. Yet, the scales differ slightly in the specification of the other two components (Schulz et al.: Sovereignty, Homogeneity; Castanho Silva et al.: People-centrism, Manichean Outlook, Akkerman et al.: Sovereignty, Manichean Outlook, see Supplement 4 for question wordings). To give an impression of the measurement of populist attitudes, Table 2 shows the indicators of an adaption of the Schulz et al. populism scale (see Supplement 3 for differences between original scales and their adapted versions used in this study).

Table 2 **Items to measure populist attitudes**

Subdimensions	*Item 1*	*Item 2*	*Item 3*
Anti-Elitism	Politicians talk too much and take too little action.	The differences between the people and the so-called elite are greater than within the people.	Politicians care about what ordinary people think [R]
Homogenous People	Ordinary people are of good and honest character.	Ordinary people all pull together.	Ordinary people share the same values and interests
People's Sovereignty	The people should have the final say on the most important political issues by voting on them directly in referendums.	The people, not the politicians, should make our most important policy decisions.	The politicians in Parliament need to follow the will of the people.

Note: [R] indicates a reverse-coded item. Question-wording from the German Longitudinal Election Study using a modification of the Schulz et al. populism scale, see Supplements 3,4 for details.

Importantly, while all considered scales conceptualize populist attitudes as multi-dimensional, they employ different strategies to account for the concept's non-compensatory qualifier. By and large, the Schulz et al. and Castanho Silva et al. scales measure each subdimension separately with distinct items for each dimension. The Akkerman et al. scale, on the other hand, often uses double-barreled items that tap into more than one subcomponent so that

individuals only score highly when agreeing with all aspects that are mentioned in a survey item (e.g., "*I would rather be represented by a citizen than by a specialized politician*" which taps into sovereignty orientations and anti-elitist orientations, see Table S6-1 for an overview of the scale items). The latter strategy has certain disadvantages such as the ambiguous meaning of low values or greater differential item functioning (see Supplement 14). However, it has the advantage of at least partly accounting for the concept's non-compensatory dimensions before operationalization already on the stage of measurement. Note, however, that the individual Akkerman et al. items do not tap into all subcomponents (see Supplement 6). Therefore, grouping and aggregating the items into composite scores is still necessary and may thus lead to varying results depending on the operationalization strategy.

For the analysis in the main text, we make use of three data sources. First, we employ the Campaign Panel 2017 of the German Longitudinal Election Study (Roßteutscher et al. 2018). Respondents for this multi-wave online survey were drawn from a heterogeneous online sample, using quotas that are representative of the German online population. The dataset offers a large sample of respondents (N=13,563) and a rich set of substantively relevant variables, including an adaptation of the populism scale by Schulz et al. (2018). We use populism measures from survey wave 5, which was conducted between August 17 and August 28, 2017, one month before the German federal election (see Supplement 7 for descriptive tables of sample characteristics). To assess the Akkerman et al. and Silva et al. populism scale, we use another dataset from the German election study (Roßteutscher et al. 2019) that was collected using CAPI in the fall of 2017 among 2112 respondents. In addition, we use data collected by Castanho Silva et al. (2018; 2019), which was kindly made available by the authors. The data was collected through MTurk and Crowdflower between November 2016 and March 2017 in nine countries: Brazil, France, Greece, Ireland, Italy, Mexico, Spain, the

United Kingdom, and the United States. The sample sizes vary between 505 (MTurk, USA) and 186 (Crowdflower, Ireland) respondents. Note that the respondents interviewed in the GLES CAPI survey were randomly drawn from local population registration, whereas the GLES campaign panel and the Castanho Silva et al. datasets were both collected online using non-probability samples. In Supplement 8, we discuss potential sample biases and compare results from probability and non-probability based samples, which do not indicate systematic disparities by sampling strategy.

At the core of the *Bollen* approach is the (weighted) summation of the concept subdimensions into aggregate populism scores where each subdimension also represents the (weighted) aggregation of multiple indicators.[5] Existing studies employ various computational strategies to aggregate the subdimensions into attitudinal populism scores, from simple means to exploratory or confirmatory factor analyses (see Supplement 1). In the main text, we opted to use the most transparent and straightforward aggregation technique: the average of the equally weighted concept subdimensions. More specifically, we first standardized all indicators to a mean of zero and a standard deviation of one. As discussed at greater length below, standardization was conducted for all operationalization approaches in order to ensure scale invariance of the concept subdimensions. The standardized indicators were then aggregated into subdimensions by computing the unweighted average score of the indicator. To attain Bollen populism scores, we computed the unweighted mean of the concept dimensions. Hence, according to this approach, populist attitudes form a continuous variable for which higher values on one subdimension may compensate for lower values on another

[5] In compliance with standard measurement practices, it is entirely adequate to represent the relationship between a subdimension of populist attitudes and its indicators as latent and reflective, thus, to use structural equation modelling for the operationalization of these subdimensions. The theoretical propositions of populist attitudes only require the specification of necessary conditions with regard to the relationship between the concept subdimensions and the resulting aggregated populism score.

subdimension and for which an increase on one subdimension always contributes to higher overall populism scores. More sophisticated aggregation methods yield similar substantive results (see Supplement 9).

$$Bollen := \sum_{i=1}^{n} Weight_i * Component_i$$

At the core of the *Sartori a*nd *Goertz* approach is the necessary condition that all subdimensions require high values to result in high populism scores. Various operationalization strategies are conceivable for various degrees of substitutability between concept subdimensions (see Supplement 2). However, the most straightforward operationalization strategy for a non-compensatory concept with a continuous scale (*Goertz* concept structure) is to use the minimum value of the concept subdimensions. Hence, higher values on one subdimension were not allowed to compensate for lower values on another subdimension. Moreover, individuals with low scores on one concept subdimension are not assigned high populism scores.

$$Goertz := \min |Component_{1,} \dots, Component_n|$$

Operationalizing the Sartori concept structure involves setting thresholds for each subdimension. If individuals surpass these thresholds on all subdimensions, they are categorized as populists. Otherwise, individuals are considered as non-populists. Although thresholds may differ across subdimensions, in the absence of meaningful theoretical reasons for differentiation, we set equal thresholds for all subdimensions. We opted for a threshold at the 75% percentile. Thus, populists are those individuals who embrace each subdimension of populist attitudes much more strongly than their fellow citizens. This threshold is, arguably, somewhat arbitrary and we report the results with other thresholds in Supplement 10.

$$Sartori := \begin{cases} 1 & if\ Component_{1,} \ldots, Component_n > Threshold \\ 0 & else \end{cases}$$

In order to increase the transparency of methodology and results, we provide additional information in the Supplementary Files. Moreover, the aforementioned Shiny Web Application allows users to analyze the data underlying this study with an easy-to-use interface. Users may compute correlations between Bollen and Goertz populism scores with substantive variables of interest using additional survey samples that are not reported in the main text. In addition, users may examine the sensitivity of different Sartori thresholds when computing the shares of populist citizens in various populations, see http://populism.alexander-wuttke.de.

Analysis

We examine the degree to which different strategies for the operationalization of populist attitudes lead to different empirical results in three steps. The first analysis examines the distribution of the different populism concept structures and the relationship between them. Having demonstrated substantial disparities between populism constructs, the next analytical step examines variations between countries and populism scales. In a third step, we examine whether different concept structures yield different conclusions about the relationship of populist attitudes with attitudinal variables of interest; institutional trust in particular.

We start the analysis by conducting descriptive analyses of the populism constructs in order to understand how the concept attributes give rise to composite scores of populist attitudes. For the dataset's large sample size, we use the German campaign panel survey for the first analysis. Figure 2 shows univariate distributions and bivariate correlations of the three subdimensions of the Schulz et al. populism scale and how they relate to the Goertzian, Sartorian, and Bollen operationalizations of populist attitudes.

We first examine the empirical properties of populism's concept subdimensions, inspecting the univariate distributions on the plot's diagonal. Anti-elitist orientations and support for popular sovereignty are strongly right-skewed, whereas the acceptance of homogeneity is more or less normally distributed. The sub-dimensions are weakly to highly correlated with the strongest association between anti-elitism and popular sovereignty. Apparently, disdain for the political elites and support for the transfer of power into the hands of ordinary people often go hand in hand, at least among German respondents in the surveyed period. The case is different for the perception of popular homogeneity.[6] Many respondents in the sample hold strong anti-elitist views but do not perceive the populace as homogenous. Consequently, the subdimensions of populist attitudes are statistically related but they represent distinct political orientations.

The unbalanced correlations pattern between the subdimensions has implications for the aggregation of these subdimensions into populism scores. Because the Bollen populism scores follow from averaging across all concept attributes, it is not surprising that the Bollen concept structure is strongly correlated with the anti-elitism and the sovereignty subdimensions which themselves are very similar. At the same time, it has less in common with the homogeneity subdimension of populist attitudes. In contrast, the Goertz concept structure correlates to a similar degree with all subdimensions.[7] Hence, even with equal weighting factors, the level to which one subdimension is reflected in the Bollen populism

[6] It is conceivable that the peculiar role of the homogeneity dimensions might result from discrepancies between the original Schulz et al. scales and the adapted version of the instrument that is employed in the GLES Campaign Panel. Yet, other studies (see Hieda, Zenkyo, and Nishikawa 2019) and the Castanho Silva et al. replication dataset in which these modifications to the Schulz et al. scale were not undertaken show similar patterns for the homogeneity dimension (see Supplement 3).

[7] This result is due to the standardization at the indicator-level that was conducted before aggregation. If no standardization is applied, then the subdimensions with the highest item difficulty will correlate most strongly with the Goertzian populism score.

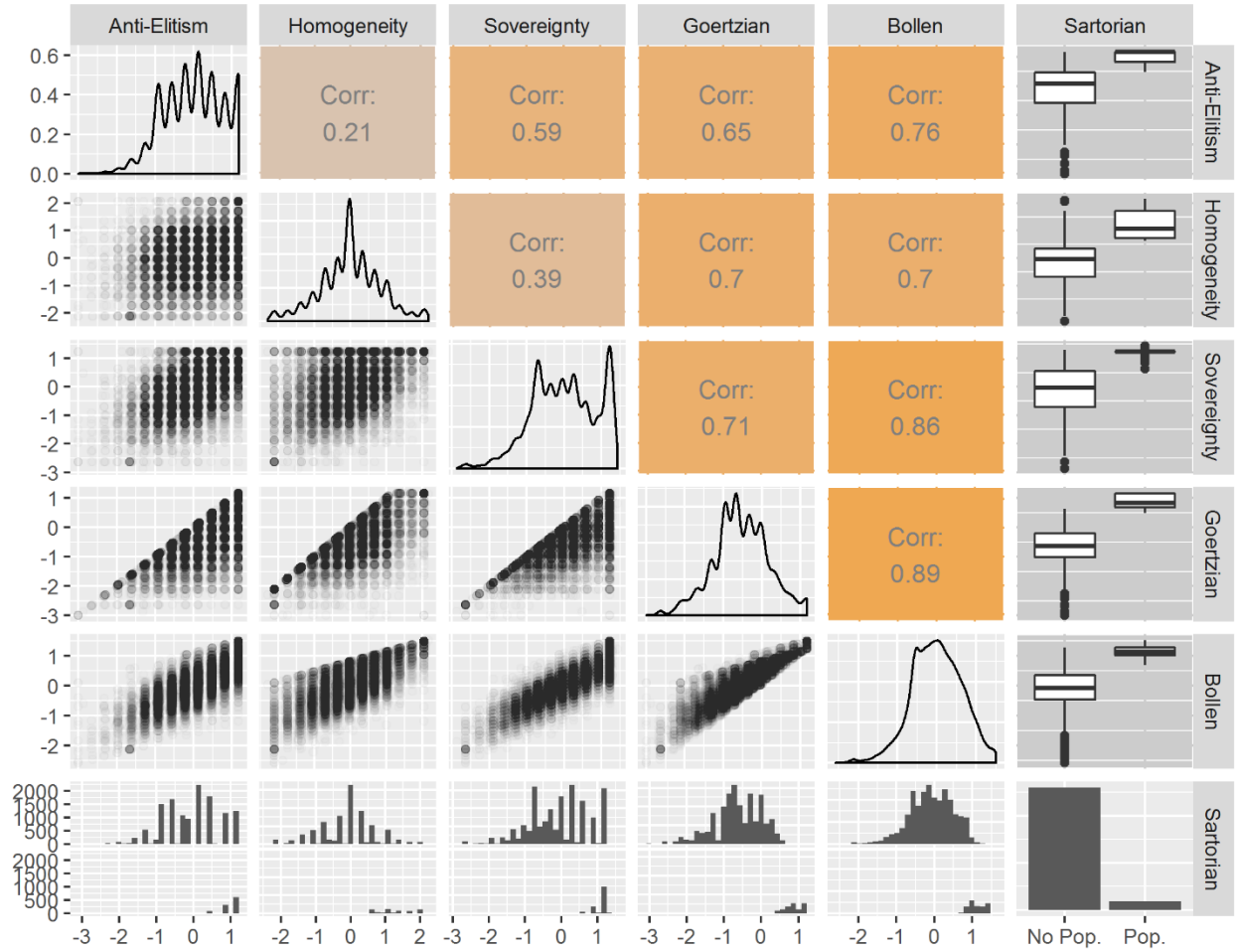
score varies as it depends on the relative closeness of that subdimension to the other components of populism.

Turning to the consistency of concept and measurement, the Bollen operationalization does not fare well. The Bollen operationalization is vulnerable to concept-measurement inconsistencies if individuals are low on one subdimension but very high on the remaining subdimensions. Due to the empirical distribution of the homogeneity component, this hypothetical scenario appears rather plausible in the German dataset. Indeed, the joint scatterplot for the Bollen scores and the homogeneity subdimension in the bottom left of Figure 2 provides evidence for such inconsistencies. To some of the respondents, the Bollen operationalization assigns relatively high populism scores despite homogeneity scores close to zero. Concept-measurement inconsistencies in the Bollen populism scores are most prevalent for the homogeneity subdimension, but they can be observed for all concept attributes. If we take the argument seriously that individuals must accept all concept attributes to achieve high populism scores, then observing individuals with low values on one subdimension but fairly high populism scores is evidence that the Bollen approach fails to operationalize populism's essential theoretical propositions.

This kind of concept-measurement inconsistency does not occur for the Sartorian and Goertzian concept structures. Examining the Sartori concept, the histograms in the plot's lowest row show that Sartori populists hold anti-elitist views *and* support popular sovereignty *and* view the populace as a homogeneous unit. About five percent of the sample surpasses the threshold on all subdimensions and were categorized as populists. Consequently, populists exist at the mass level and proper operationalization enables measuring the share of populists within a society. If populist attitudes are conceived of as continuous, we can employ the Goertz concept structure. Inspecting the joint distribution of the Goertz populism scores and

the concept components, one characteristic of the plot signals the desired properties of the Goertz concept structure. There is a clear-cut bisecting line in all of these scatterplots, which is the graphical equivalent of the rule that populism scores must not be high if an individual is low on one subdimension. Hence, the Sartori and Goertz operationalization strategies ensure that high populism scores are assigned only to individuals who exhibit high values on all concept components.

Figure 2. **Distributions of and correlations between concept structures of populist attitudes and concept attributes (Germany, Schulz et al. populism scale)**



Note: The scatterplots on the left show the joint distributions of the variables which are labeled at the top of each column and on the right of each row. The variable at the column top is plotted on the scatterplot's x-axis. Histograms show the distribution of the dichotomous Sartori measure. The plot's diagonal row shows the univariate distribution of the variables labeled at the column's top. The plot's upper right panels show Pearson's R correlation coefficients between continuous variables and boxplots for the Sartori measure. Data from the German panel survey. Plots were created using the ggally R-package.

Having established that not all concept structures respect the concept's theoretical propositions of populism at the mass level, we assess the empirical commonalities and differences between these concept structures directly. The boxplots in the right-hand column report the distribution of the Goertzian and Bollen scores among populists and non-populists according to the Sartori operationalization. Among Sartori-populists, the average Bollen populism score is lower than the average Goertz populism score. More importantly, among non-populist respondents, there are more individuals with high Bollen populism scores than

individuals with high Goertz populism scores. In other words, the Bollen operationalization assigns high populism scores to several respondents whom the Sartori approach classifies as non-populists.

Finally, we turn to compare the Bollen and Goertz concept structures. The German dataset shows a correlation of r=0.89 between the measures, which is usually considered a strong association. We discuss the strength of this relationship at greater length below, but here it suffices to say that the correlation is notably different from 1. Hence, comparing the Bollen and Goertz populism scores leads to the first important conclusion: the operationalization strategy that most current populism studies employ does not always match the populism scores of the Goertz concept structure, which also entails a continuous scale but which strictly adheres to the concept's theoretical propositions.

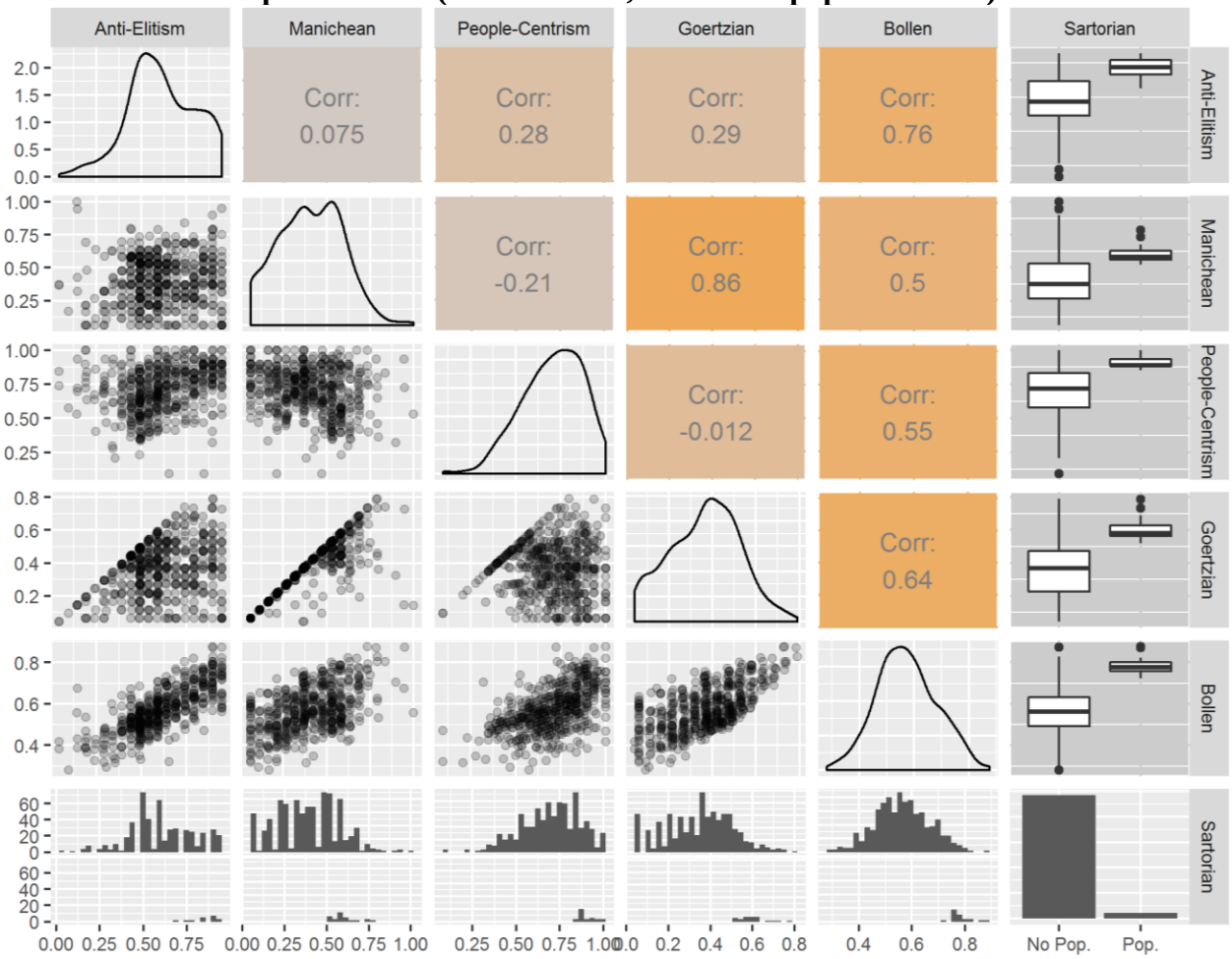
What is more, the discrepancies between the measures have systematic origins. As the scatterplot shows, the discrepancies mainly occur because the Bollen operationalization assigns high values to individuals with low Goertz scores. As we saw before, these cases refer to respondents for which the Bollen operationalization violates the concept's theoretical proposition not to assign high values if one concept subdimension is low. Consequently, when the Bollen populism scores exceed the Goertz scores, then only the latter capture the distinctive position of populist attitudes at the intersection of its concept components. In contrast, in these cases, the Bollen scores tap into a mix of attitudinal concepts that are related to populism at the mass level but do not reflect the concept itself. Put differently, in case of a mismatch, the Bollen operationalization does not measure what it intends to measure, and the Goertz scores should be preferred.

So far, these results only speak to the distribution of populism scores as derived from one populism scale in one country. To examine the generalizability of these findings, the

following analyses will broaden the perspective by turning to the Castanho Silva scale in the United States, the largest sample in the Castanho Silva et al. replication dataset. Figure 3 shows that the anti-elitist subdimension of the Castanho Silva et al. scale is right-skewed, whereas the distribution of Manichean outlook orientation is left-skewed. Notably, these concept components of the Castanho Silva et al. scale are barely correlated. That is, knowing a person's stance on political elites does not help predict her perception of good and evil in society. The fact that the subdimensions of populist attitudes are almost orthogonal in the US survey is noteworthy from a methodological perspective. The Bollen approach conceptualizes a person's latent populist inclination as the common cause of the concept subdimensions which implies that the subdimensions should be correlated. Therefore, a lack of correlations among them raises doubts about the psychometric quality of the measurement. In contrast, from an essentialist perspective whether an attitude is an essential component of populism at the individual level is not an empirical but an ontological question. Consequently, the approach which underlies the Sartori and Goertz operationalizations is fully compatible with low co-variances. From this perspective, the implication of weakly correlated concept components merely is that few individuals hold populist attitudes because in most cases the building blocks of this concept are not jointly present.[8]

[8] Indeed, using 75th percentiles on the US dataset, the Goertz concept structure yields far less populists on the Castanho Silva et al. scale with weakly correlated sub-dimensions (0.6 percent) than on the Schulz et al. scale (5.1%) with more strongly correlated sub-dimensions.

Figure 3. **Distributions of and correlations between concept structures of populist attitudes and concept attributes (United States, Silva et al. populism scale)**



Note: The scatterplots on the left show the joint distributions of the variables which are labeled at the top of each column and on the right of each row. The variable at the column top is plotted on the scatterplot's x-axis. Histograms are used to show the distribution of the dichotomous Sartori measure. The plot's diagonal row shows the univariate distribution of the variables labeled at the column's top. The plot's upper right panels show Pearson's R correlation coefficients between continuous variables and boxplots for the Sartori measure.

The magnitude of the covariance between the subdimensions affects the composite scores derived from compensatory or non-compensatory operationalization strategies.[9] The Goertz and Bollen measures are related because lower covariances imply a larger share of individuals who exhibit low values on one concept subdimension but high values on the remaining subdimensions. Because this configuration of attitudes leads to differences between the Goertz

[9] Due to limitations of space, in the discussion here and below, we focus on the Bollen measure (for its frequent usage in the literature) and the Goertz measure (for its practical and epistemological qualities), but disregard the Sartori measure.

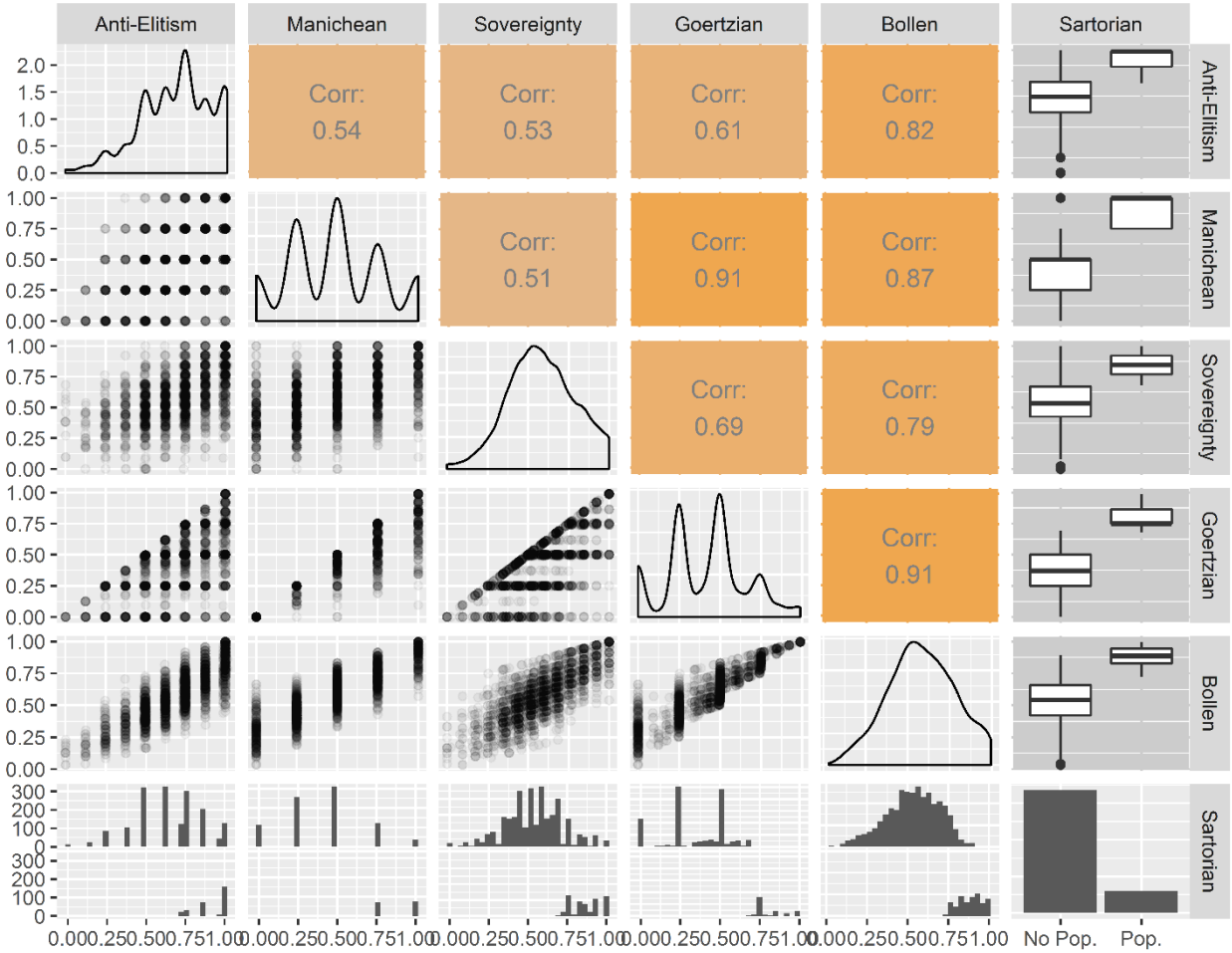
and Bollen populism scores, low component covariances tend to go along with higher disparities between composite populism scores (for simulations of this interdependency, see Supplement 11). Against this backdrop, it is little wonder that the correlation between the Bollen and Goertz scores amounts to r=.64 in the US dataset with the Castanho Silva et al. scale and is thus considerably lower than in the German dataset with the Schulz et al. scale (r=.89) considering that the covariance between components is much higher for the Schulz et al. scale. Hence, the weaker the correlation between the subdimensions, the larger the concept-measurement inconsistency of the Bollen operationalization will be.

This interdependency between the concept components and the composite scores has substantive ramifications for the extent of resulting disparities between operationalization strategies in research practice. Keeping in mind that not all populism scales specify populist attitudes in the same way, composite scores are most sensitive to operationalization strategies if scales conceptualize populist attitudes as constituted by distinct and hardly related attitudinal orientations. As scale properties also vary between samples, another implication is that composite scores of non-compensatory concepts are most sensitive to operationalization strategies in populations with lower covariances between the concept components.

That the extent of measurement concept-inconsistencies varies across scales and countries is further emphasized by the internal structure of the Akkerman et al. populism scale, as observed in the German probability survey (Figure 4). Presumably partly due to the scale's particular measurement strategy, the discrepancy between the Bollen and Goertz scores are the lowest of all the scales (r=.91). As the scale already considers, in part, the non-substitutability of the concept subdimensions at the measurement stage it makes less of a difference whether we choose an operationalization strategy that is adequate for non-compensatory concepts. Note, however, that operationalizing the Akkerman et al. scale with

either the Goertz or Bollen approach still does not yield identical results as their correlation coefficient is notably different from one.

Figure 4. **Distributions of and correlations between concept structures of populist attitudes and concept attributes (Germany, Akkerman et al. populism scale)**

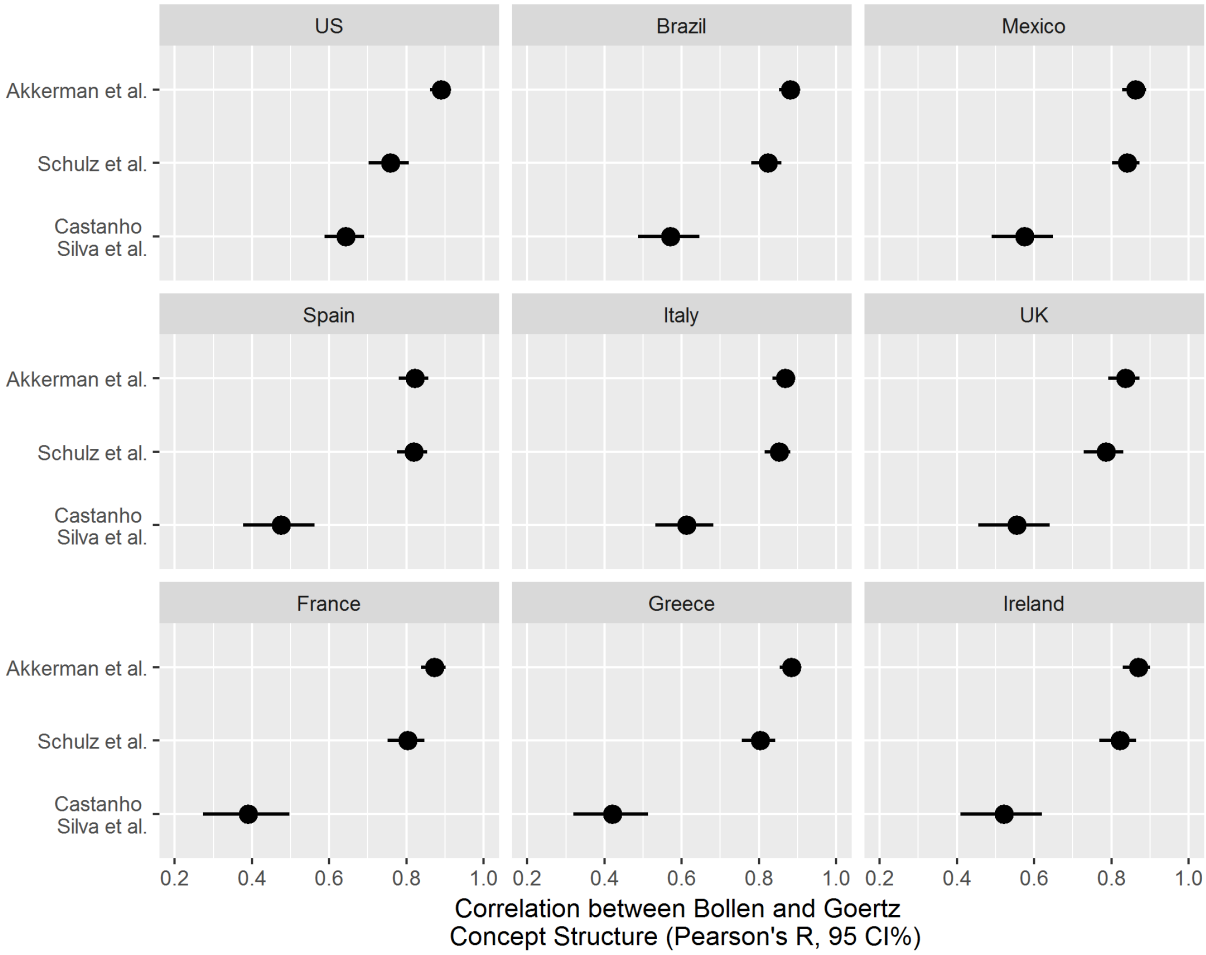


Note: The scatterplots on the left show the joint distributions of the variables which are labeled at the top of each column and on the right of each row. The variable at the column top is plotted on the scatterplot's x-axis. Histograms are used to show the distribution of the dichotomous Sartori measure. The plot's diagonal row shows the univariate distribution of the variables labeled at the column's top. The plot's upper right panels show Pearson's R correlation coefficients between continuous variables and boxplots for the Sartori measure. Data from the German cross-sectional election survey.

To broaden the view on the variation of discrepancies between operationalization strategies in different survey samples and populism scales, the next step of the analysis uses the Castanho Silva et al. replication dataset with survey data from nine countries. Figure 5 shows correlations (Pearson's R) between the Goertz and Bollen concept structures for the Akkerman

et al., Schulz et al., and Castanho Silva et al. populism scales. The figure provides two significant insights.

Figure 5. **Correlations between Goertz and Bollen concept structures of populist attitudes**



First, the figure underscores the modest but notable variation in the similarity of the populism constructs depending on which scale is used and where the survey was conducted. The Goertz and Bollen scores of populist attitudes usually correlate at r=.8 when measured with the Schulz et al. scale. The Akkerman et al. scale exhibits slightly higher correlations in most cases. In comparison, the correlations of the composite scores of the Castanho Silva et al. scale lies on a lower level, between r=.39 in France and r=.63 in the United States. Figure 5 also shows some country differences. For instance, the populism constructs are more strongly correlated

in the Italian sample than in other samples. Consequently, for future research on populist attitudes, it is important to note that the empirical implications of operationalization decisions differ between scales and samples.

The cross-national results show, second, that although both populism constructs are derived from the exact same indicators, the discrepancies between the Goertzian and the Bollen populism scores are often substantial. None of the discovered correlations comes close to a perfect correlation of r=1. In fact, all correlations in Figure 5 are smaller than r=.9, and many depicted correlation coefficients hover between r=.5 and r=.8. Correlations of this magnitude are usually conceived of as moderate to strong associations. However, when evaluating the correlation coefficients presented here, we have to keep in mind that these correlations do not pertain to different constructs. Rather, both of the correlated scores allegedly capture the same concept. Specifically, both populism scores were derived from the same data-generating process, and they only differ in the aggregation rule that was administered. Against this backdrop, even a far-from-perfect correlation of, e.g., r=.8 between the Goertz and Bollen scores is notable. A correlation of 0.8 means that knowing one variable only allows us to predict two thirds (64%) of the variance of the other populism score. Using a real-world analogy, the statistical association between populism constructs derived from the same indicators is about as strong as the correlation between arm-span and height (Reeves, Varakamin, and Henry 1996). In other words, the Bollen and the Goertz populism scores can be considered similar for many individuals particularly when the Akkerman et al. scale is used, but they are far from identical when the Schulz et al. scale is used, and the scores differ even more noticeably for the Castanho Silva et al. scale.

So far, the analysis has established that the Bollen and Goertz constructs do not always derive identical populism scores, and when they diverge, the Bollen operationalization does

not do justice to the essence of populism as an attitudinal syndrome with non-compensatory subdimensions. However, given the moderate to strong correlations, it is possible that both concept structures yield the same substantive results when analyzing the determinants and consequences of populist attitudes. If this was the case, the distinction between the concept structures would primarily be of theoretical value but less critical for applied empirical work, apart from the distribution of populist attitudes. Hence, we now turn to explore whether the choice between concept structures makes a difference in explaining other substantive variables.

While it is impossible to address all relevant questions tackled in prior research, we selected some that, in our view, are of interest to empirical scholars in populism. To reduce analytical discretion and to enhance transparency, we opted for the simplest model possible: bivariate correlations. Due to limitations of space, we only show results on the Castanho Silva et al. scale in the main text using the authors' cross-national dataset. Yet, the Shiny Web Application enables readers to investigate statistical associations between composite scores of populist attitudes and an extensive list of correlates using different datasets and scales. By and large and for most scales and samples, the analyses reveal weaker associations with Bollen populism scores compared to the Goertz construct of populist attitudes. While these disparities often do not affect the substantive conclusions, in a few extreme instances, the correlations even point in different directions. The Castanho Silva et al. scale seems most sensitive to differences in operationalization strategies. Yet instances in which the substantive conclusion suggested by the Bollen construct does not square with findings with Goertz populism scores can be observed for all scales.

In the following analysis, we employ the Castanho Silva et al. replication dataset. The dataset only contains two potential correlates (conspiratorial thinking, institutional trust) but

covers nine countries, thus demonstrating the robustness of findings in multiple samples. As reported in Supplement 12, correlations of Goertz and Bollen populism scores with conspiratorial thinking differ notably in size when they have the same sign, but often the correlations even point in opposite directions. In the following, we report findings for the second available variable, institutional trust. Institutional trust is measured by a summary score of trust in political parties, government, and parliament.[10]

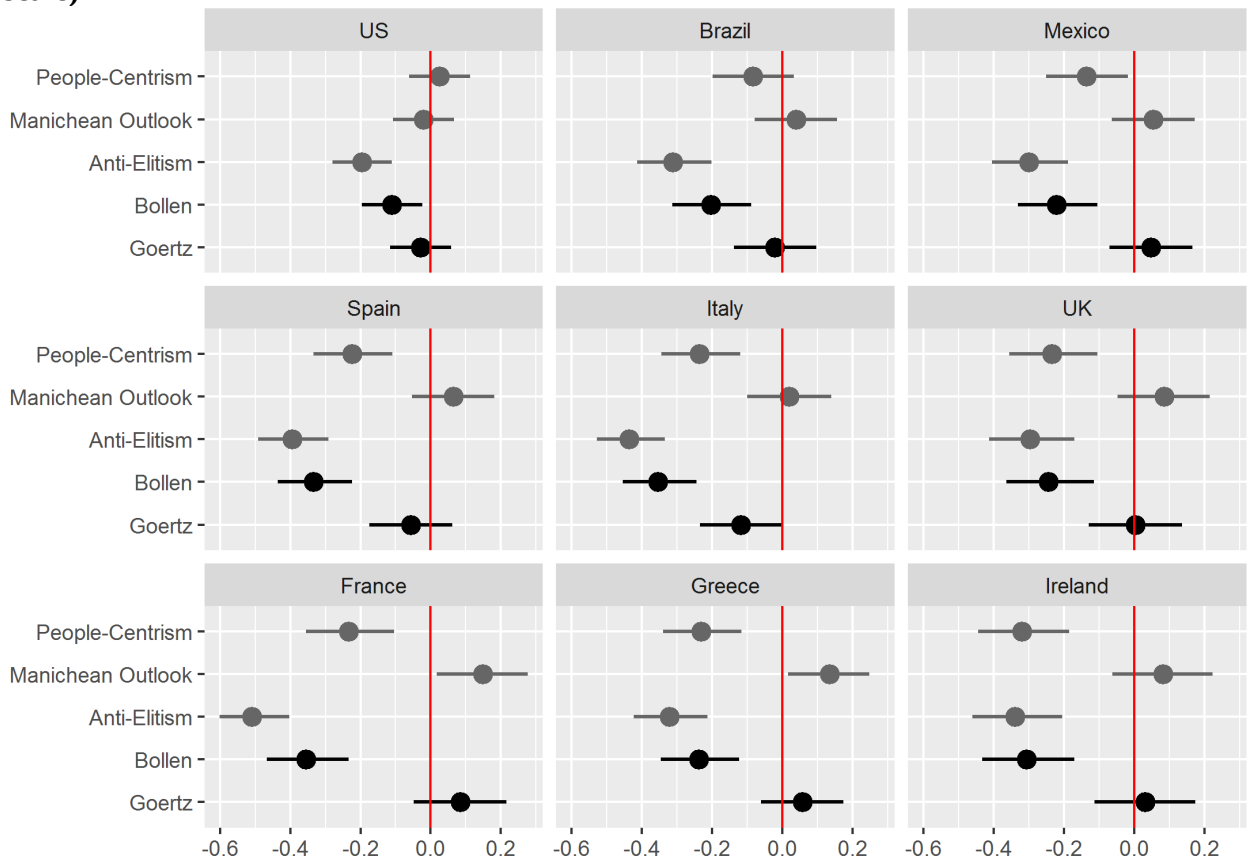
Institutional trust is a particularly interesting case because for this variable we might for theoretical reasons expect differences in the results for the Bollen and the Goertz concept structures. This expectation derives from the following reasoning. It is well established that stronger anti-elitist sentiments go hand in hand with lower levels of trust in political and societal institutions (Erber and Lau 1990). After all, it is political elites that usually run these institutions. Beyond anti-elitism, however, it is more difficult to reason a substantive link between institutional trust and populist attitudes. It is not self-evident why other attributes of populist attributes, such as belief in a homogeneous citizenry, might go along with a person's levels of institutional trust. Most importantly, it is not self-evident why the intersection of these attitudes should be associated with trust in institutions. As the Bollen construct reflects the average of all subdimensions, the respective populism scores react to changes on one subdimension regardless of the values on the other subdimensions. Bollen scores will thus reflect the correlation of institutional trust and anti-elitism in any case. The Goertz construct, by contrast, considers all concept components in combination and thus does not reflect changes of the subdimensions above the minimum. If anti-elitism happens to be not the subdimension with the minimum score, variation on it is irrelevant for Goertz populism scores.

[10] In Supplement 13 we examine the internal consistency of the summary score on institutional trust. Cronbach's alpha is around 0.85 in most samples but around 0.65 in the US and in the UK, suggesting that the structure of perceptions of these institutions differ between countries.

As the subdimension with the strongest relationship to institutional trust does not feed into the Goertz populism scores for all respondents, it is straightforward to expect that the Bollen scores exhibit stronger correlations with institutional trust than the Goertz scores do.

Figure 6 displays bivariate correlations between institutional trust and the Bollen and Goertz constructs of populist attitudes as well as with the concept's subdimensions. As expected, there is a consistent negative association between anti-elitist orientations and trust in political institutions. Institutional trust is not associated with a Manichean outlook and not consistently associated with people-centrist attitudes. Hence, one subdimension of populist attitudes exhibits the expected link with the concept of interest, whereas the other subdimensions do not.

Figure 6. **Bivariate correlations with institutional trust (Castanho Silva et al. populism scale)**



Note: In all samples, the difference between the correlation sizes of the Bollen and the Goertz concept structures is statistically significant (computed with the CoCor R-package)

Figure 6 also shows a statistically significant correlation between populist attitudes and institutional trust for the Bollen populism scores. Considering the mathematical formula underlying the Bollen concept, it is not surprising that the aggregated populism construct inherits partial correlations of the concept subdimensions. Using the Bollen operationalization blurs the distinction between the subdimensions of populist attitudes and the concept of populist attitudes itself as the conjunction of its concept components. Replicating previous findings in the public opinion literature on political cynicism (e.g., Erber and Lau 1990), it is not surprising to observe an attitudinal linkage between a person's alienation with political actors (anti-elitism) and the disdain for the organizations they represent (institutional trust). However, if the correlation of trust with the Bollen populism score is merely a remnant of the association with anti-elitism, then the reported correlation would represent the replication of old findings with a new name. The Bollen construct thus cannot distinguish whether a finding reflects the unique properties of populism or only that of one subdimension. Put differently, the Bollen concept structures leaves unclear whether a finding obtained with the construct is driven by the concept of interest, populist attitudes.

The Goertz concept structure captures the correlation between institutional trust and populist attitudes understood as an attitudinal syndrome rather than the shared variance of its sub-dimensions. Using this construct, different conclusions about populism's relationship with institutional trust emerge. In eight out of nine countries, Figure 6 shows no meaningful correlation between institutional trust and the Goertz populism construct. Hence, the correlation observed when using the Bollen construct disappears if populism is adequately operationalized. In substantive terms, as many non-populists are equally critical of elites the analysis demonstrates that populist attitudes are not related to institutional trust.

The finding that different operationalization strategies of populist attitudes may yield inconsistent results when applied in correlative analyses is the keystone in a longer series of assessments we have conducted in this article. The preceding analyses demonstrated that composite scores of populist attitudes can differ to a smaller or larger degree, depending on the properties of scales and samples. These differences were shown to occur in cases when the Bollen concept structure fails to capture the concept's core proposition but instead taps into related attitudinal constructs that are represented in populism's subdimensions. These concept-measurement inconsistencies of the Bollen approach will not always, but they can have consequences for substantive research since we demonstrated that they can lead to wrong judgments about the nature and correlates of populism at the mass level. Consequently, scholars are well-advised to employ the Goertz or Sartori concept structures that are capable of reflecting populism as an attitudinal syndrome in order to capture the essence of populist attitudes in their empirical analyses.

Conclusion

The notion of a "populist Zeitgeist" (Mudde 2004) has given rise to a quickly growing research field on the prevalence, causes, and consequences of populist attitudes (e.g., Akkerman, Mudde, and Zaslove 2014; Tsatsanis, Andreadis, and Teperoglou 2018; van Hauwaert, Schimpf, and Azevedo 2018). What makes populist attitudes valuable as a concept is the claim that populism at the mass level represents more than the sum of its parts. In its distinctive position at the intersection of the concept components, populist attitudes are not just another variant of concepts well-known to scholars of public opinion. This concept structure has to be taken into account when operationalizing populist attitudes in empirical research. Research thus far, however, often neglected the key characteristic of the concept, namely, that populist

attitudes are an attitudinal syndrome with non-substitutable subdimensions. As a consequence, the mathematical structure of measures of populism and the concept's theoretical structure fall apart (cf. Goertz 2006, 125). By implication, populism constructs often do not reflect what they are intended to measure. Prevailing practices thus put at risk the concept's field utility and the differentiation from established concepts in public opinion research. In effect, critics may ask whether research on populist attitudes is just old wine in new bottles and may call for abandoning research on populist attitudes (cf. Geurkink et al. 2019). We think this conclusion would be premature because populist attitudes are an original concept and empirical research on it may provide valuable insights, once the distinctive nature of the concept is taken seriously.

In order to overcome the current state of affairs and to protect research on populist attitudes against unjustified accusations, we proposed operationalization techniques that are capable of preserving the concept's unique properties in empirical applications. We identified two suitable approaches. Both the Goertz and the Sartori approach properly account for the necessary conditions of populist attitudes and are thus superior to the Bollen approach. They differ, however, in the assumptions about the continuous or dichotomous nature of attitudes. While the Sartori approach builds on a dichotomous conception, the Goertz approach presupposes a continuous conception that is more adequate in attitudinal research and, therefore the preferred strategy to operationalize populist attitudes. Considering that both approaches are not by necessity computationally more demanding than the widely used Bollen approach, this study suggests that for proper empirical analyses it is at least as important to align conceptual reasoning with computational practices as to demonstrate technical sophistication (e.g., Sartori 1970).

As applications of multi-dimensional concepts blossom in many subfields of political science, the discussion of concept structures in this article may help to carefully consider the theoretical properties of concepts with multiple concept components when creating or applying indicators in empirical research (see Supplement 2 for specific advice). This applies to populism research beyond the individual level, e.g., populism of parties (March 2017; Mudde and Kaltwasser 2013a) as well as to other fields such as research on social capital (e.g., Putnam 2000). Other complex and currently debated concepts such as support for democracy (e.g., Schedler and Sarsfield 2007), ideological orientations (Converse 2006), political sophistication (e.g., Luskin 1987), public attitudes toward globalization (e.g., Mader, Steiner, and Schoen 2019), and political extremism (e.g., Jungkunz 2019) are among the topics for which this suggestion may prove fruitful. In these and other fields of research, it seems worthwhile to consider whether central concepts involve non-compensatory subdimensions and whether prevailing operationalization strategies sufficiently account for these concept properties.

The evidence demonstrated that the Bollen approach's vulnerability to concept-measurement inconsistencies can make a considerable difference for substantive conclusions about the prevalence, causes, correlates, and consequences of populist attitudes. This finding raises issues concerning the validity of results in prior research. The good news is that the evidence in this paper demonstrated that the aggregation rule does not always make a substantial difference. In some cases, the Bollen and Goertzian concept structures are highly correlated and may thus lead to similar substantive conclusions; in other cases, this does not apply (see Shiny Web Application: http://populism.alexander-wuttke.de). Whether measurement-inconsistencies occur and induce biases in correlative analyses varies between populations and populism scales. It is not the aim of this study to evaluate existing populism

scales but to provide guidance for their proper operationalization. Nonetheless, the general pattern that non-compensatory operationalization strategies are most consequential for scales with low covariances between the subdimensions is noteworthy. In contrast, scales are less prone to concept-measurement inconsistencies when they have few subdimensions or already consider the non-compensatory properties at the stage of measurement even though the latter approach may result in a whole set of new challenges. In addition, measurement-inconsistencies of the Bollen concept structure are more likely to bias results of correlative analyses if one subdimension of populist attitudes drives the shared variance with a variable of interest. Whether the concept-indicator misfit in prior research matters for substantive findings thus depends on empirical distributions. Our analysis may thus be read as a call for systematic testing of the validity of prior results. This, in turn, may make research on populist attitudes even more vibrant and foster its comparative outlook.

Using indicators that properly reflect the concept structure of populist attitudes may affect the substantive conclusions about the empirical properties of populist attitudes as well as their implications for political processes. Given the nature of populist attitudes as lying at the intersection of different components, they resemble political ideologies that are at the core of political belief systems (Converse 2006; Gerring 1997; Kinder and Kalmoe 2017). As holding a coherent ideology is quite challenging, it is thus little wonder that we found populist attitudes to be not widely held by citizens. However, those citizens who hold populist attitudes in a strict sense subscribe to not only one but several ideas that have a strained relationship with pluralist and representative forms government (Caramani 2017; Müller 2017; Urbinati 2013). These appear to be people on which we cannot count when it comes to defending democracy against its foes. What is more, strongly interconnected nodes of political belief systems such as ideologies are more resistant to persuasive influences, more easily accessible,

and more consequential in affecting other attitudes and behaviors (Petty and Krosnick 1995). It is thus implausible to expect citizens who subscribe to populism sensu stricto to be very responsive to attempts at making them more supportive of core ideas of representative democracy. In a sense, a more adequate measure of populist attitudes may thus paint a paler and brighter picture of the societal foundations of democracy: populist attitudes may be less widespread but more resistant to change than suggested by previous research.

As this paper sought to investigate the relationship between conceptual, methodological, and substantive issues, we would like to highlight the implications of a subtle methodological decision. Before aggregating them into populism scores, all items were z-standardized. This decision is consequential because standardization changes the substantive meaning of the measures from absolute to relative quantities. For instance, after standardization, the Sartori concept structure classifies those individuals as populists who accept all subdimensions of populism much more strongly, compared to the average respondent. Using relative measures raises two conceptual and empirical issues. Conceptually, nominalist approaches to concept formation might hold that a concept's boundary conditions must always be defined in absolute instead of relative terms (Sartori 1970). Empirically, relativist measures hinder the comparison of populism across time and space. Considering these arguments against standardization, it should be noted that standardization is not mandatory. For the Sartori concept, setting absolute thresholds is feasible based on the indicators' substantive meaning and the researcher's specification of populist attitudes. Likewise, researchers can use the minimum of unstandardized concept subdimensions to set the Goertzian construct. Unstandardized approaches, however, rest on the assumption of comparable measurement scales across dimensions. This assumption is not necessarily met, however, because researchers may have selected items with different item difficulties which

affect the distributions. If differences in the distributions of subdimensions solely result from researcher choices during scale development and do not reflect substantive differences between the respondents towards the subcomponent's essence, then the aggregation of unstandardized indicators is problematic. Such scoring issues are well known to qualitative researchers under the rubric of calibration (Goertz 2006) and have also gained attention in quantitative studies where they are discussed under the rubric of equating and linking (Kolen and Brennan 2014). In particular, recent advance in Item Response Theory (DeMars 2016; Liu and Chalmers 2018) appears to show a promising way forward to operationalize non-compensatory multi-dimensional constructs without neglecting issues of scale incomparability. To conclude, standardization lowers the burden of assumptions about the data-generating process, but it has implications for the meaning of the derived measures. Researchers interested in comparing populist attitudes over time or in quantifying absolute levels might prefer absolute measures, but in turn, must pay close attention to the substantive meaning of the indicator and the scales that are used.

Altogether, populist attitudes are an important topic for political science. Given its substantive significance, the concept deserves careful analysis and adequate measurement in empirical applications in order to avoid biased estimates and invalid conclusions. Our analysis demonstrated that populist attitudes should be treated as an attitudinal syndrome that is more than the sum (or average) of its subdimensions. Otherwise, it would be difficult to justify populist attitudes as a specific concept in public opinion research. Obtaining an adequate score for this concept requires operationalization strategies that differ from those widely used in previous research. By providing these conceptual and methodological suggestions, this article may prove helpful in paving the way for fruitful empirical research on an important political phenomenon of this era.

References

Agger, Robert E., Marshall N. Goldstein, and Stanley A. Pearl. 1961. "Political Cynicism: Measurement and Meaning." *The Journal of Politics* 23 (3): 477–506.

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.

Alkire, Sabina, James E. Foster, Suman Seth, Maria E. Santos, José M. Roche, and Paola Ballón. 2015. *Multidimensional poverty measurement and analysis*. Oxford: Oxford University Press.

Aslanidis, Paris. 2016. "Is Populism an Ideology?: A Refutation and a New Perspective." *Political Studies* 64 (1S): 88–104.

Barr, Robert R. 2009. "Populists, Outsiders and Anti-Establishment Politics." *Party Politics* 15 (1): 29–48.

Bollen, Kenneth, and Richard Lennox. 1991. "Conventional wisdom on measurement: A structural equation perspective." *Psychological Bulletin* 110 (2): 305–14.

Bonikowski, Bart, and Paul DiMaggio. 2016. "Varieties of American Popular Nationalism." *American Sociological Review* 81 (5): 949–80.

Caramani, Daniele. 2017. "Will vs. Reason: The Populist and Technocratic Forms of Political Representation and Their Critique to Party Government." *American Political Science Review* 111 (01): 54–67.

Castanho Silva, Bruno, Ioannis Andreadis, Eva Anduiza, Nebojsa Blanusa, Yazmin M. Corti, Gisela Delfino, Guillem Rico, Saskia Ruth, Bram Spruyt, Marco R. Steenbergen, and Levente Littvay. 2018. "Public Opinion Surveys: A New Scale." In *The Ideational Approach to Populism: Concept, Theory, and Analysis*, eds. Kirk A. Hawkins, Ryan Carlin, Levente Littvay and Cristóbal R. Kaltwasser. London: Routledge.

Castanho Silva, Bruno, Jungkunz, Marc Helbling, and Levente Littvay. 2019. "An Empirical Investigation of Seven Populist Attitudes Scales." *Political Research Quarterly* First View: 1–16.

Collier, David, and James E. Mahon. 1993. "Conceptual "Stretching" Revisited: Adapting Categories in Comparative Analysis." *American Political Science Review* 87 (4): 845–55.

Converse, Philip E. 2006. "The nature of belief systems in mass publics (1964)." *Critical review* 18 (1-3): 1–74.

DeMars, Christine E. 2016. "Partially Compensatory Multidimensional Item Response Theory Models: Two Alternate Model Forms." [eng]. *Educational and Psychological Measurement* 76 (2): 231–57.

Easton, David. 1975. "A Re-assessment of the Concept of Political Support." *British Journal of Political Science* 5 (4): 435–57.

Edwards, Jeffrey R., and Richard P. Bagozzi. 2000. "On the nature and direction of relationships between constructs and measures." *Psychological Methods* 5 (2): 155–74.

Elchardus, Mark, and Bram Spruyt. 2016. "Populism, Persistent Republicanism and Declinism: An Empirical Analysis of Populism as a Thin Ideology." *Government and Opposition* 51 (01): 111–33.

Erber, Ralph, and Richard R. Lau. 1990. "Political Cynicism Revisited: An Information-Processing Reconciliation of Policy-Based and Incumbency-Based Interpretations of Changes in Trust in Government." *American Journal of Political Science* 34 (1): 236.

Gerring, John. 1997. "Ideology: A Definitional Analysis." *Political Research Quarterly* 50 (4): 957–94.

———. 1999. "What Makes a Concept Good?: A Criterial Framework for Understanding Concept Formation in the Social Sciences." *Polity* 31 (3): 357–93.

Geurkink, Bram, Andrej Zaslove, Roderick Sluiter, and Kristof Jacobs. 2019. "Populist Attitudes, Political Trust, and External Political Efficacy: Old Wine in New Bottles?" *Political Studies* 23 (4): 003232171984276.

Goertz, Gary. 2006. *Social science concepts: A user's guide*. Princeton, NJ: Princeton Univ. Press.

Greco, Salvatore, Alessio Ishizaka, Menelaos Tasiou, and Gianpiero Torrisi. 2019. "On the Methodological Framework of Composite Indices: A Review of the Issues of Weighting, Aggregation, and Robustness." *Social Indicators Research* 141 (1): 61–94.

Hameleers, Michael, Linda Bos, and Claes H. de Vreese. 2017. "The Appeal of Media Populism: The Media Preferences of Citizens with Populist Attitudes." *Mass Communication and Society* 20 (4): 481–504.

Hawkins, Kirk A. 2009. "Is Chavez Populist?: Measuring Populist Discourse in Comparative Perspective." *Comparative Political Studies* 42 (8): 1040–67.

Hawkins, Kirk A., Ryan Carlin, Levente Littvay, and Cristóbal R. Kaltwasser, eds. 2018. *The Ideational Approach to Populism: Concept, Theory, and Analysis*. London: Routledge.

Hawkins, Kirk A., and Cristóbal R. Kaltwasser. 2018. “Introduction - Concept, Theory, and Method.” In *The Ideational Approach to Populism: Concept, Theory, and Analysis*, eds. Kirk A. Hawkins, Ryan Carlin, Levente Littvay and Cristóbal R. Kaltwasser. London: Routledge.

Hawkins, Kirk A., Scott Riding, and Cas Mudde. 2012. “Measuring Populist Attitudes.” (55). IPSA Committee on Concepts and Methods Working Paper Series.

Hieda, Takeshi, Masahiro Zenkyo, and Masaru Nishikawa. 2019. “Do populists support populism? An examination through an online survey following the 2017 Tokyo Metropolitan Assembly election.” *Party Politics* 23 (4): 135406881984811.

Jansen, Robert S. 2011. “Populist Mobilization: A New Theoretical Approach to Populism.” *Sociological Theory* 29 (2): 75–96.

Jungkunz, Sebastian. 2019. “Towards a Measurement of Extreme Left-Wing Attitudes.” *German Politics* 28 (1): 101–22.

Katz, Richard S. 1997. “Representational roles.” *European Journal of Political Research* 32 (2): 211–26.

Kenny, Paul D. 2017. *Patronage and Populism: Explaining Populist Success in India, Asia, and Beyond*. Oxford: Oxford University Press.

Kinder, D. R., and N. P. Kalmoe. 2017. *Neither Liberal Nor Conservative: Ideological Innocence in the American Public*. University of Chicago Press.

Kolen, Michael J., and Robert L. Brennan. 2014. *Test equating, scaling, and linking: Methods and practices* [eng]. 3rd ed. *Statistics for social and behavioral sciences*. New York, NY: Springer.

Laclau, Ernesto, and Chantal Mouffe. 1985. *Hegemony and socialist strategy: Towards a radical democratic politics* [eng]. 2nd ed. London: Verso.

Liu, Chen-Wei, and R. P. Chalmers. 2018. “Fitting item response unfolding models to Likert-scale data using mirt in R.” *PLOS ONE* 13 (5): e0196292.

Mader, Matthias, Thomas J. Scotto, Jason Reifler, Peter H. Gries, Pierangelo Isernia, and Harald Schoen. 2018. “How political are national identities? A comparison of the United States, the United Kingdom, and Germany in the 2010s.” *Research & Politics* 5 (3): 205316801880146.

Mader, Matthias, Nils D. Steiner, and Harald Schoen. 2019. “The globalisation divide in the public mind: belief systems on globalisation and their electoral consequences.” *Journal of European Public Policy:* 1–20.

March, Luke. 2017. “Left and right populism compared: The British case.” *The British Journal of Politics and International Relations* 19 (2): 282–303.

Miller, A. H. 1974. “Political issues and trust in government: 1964-1970.” *Annual Review of Political Science* 68: 951–72.

Moffitt, Benjamin, and Simon Tormey. 2014. “Rethinking Populism: Politics, Mediatisation and Political Style.” *Political Studies* 62 (2): 381–97.

Mohrenberg, Steffen, Robert A. Huber, and Tina Freyburg. 2019. “Love at first sight? Populist attitudes and support for direct democracy.” *Party Politics* Online First.

Møller, J., and S. E. Skaaning. 2012. *Requisites of Democracy: Conceptualization, Measurement, and Explanation*. Taylor & Francis.

Mudde, Cas. 2004. “The Populist Zeitgeist.” *Government and Opposition* 39 (4): 542–63.

———. 2007. *Populist radical right parties in Europe*. Cambridge and UK and New York: Cambridge University Press.

———. 2017. “Populism: An Ideational Approach.” In *The Oxford Handbook of Populism*. *Oxford handbook of politics*, eds. Cristobal R. Kaltwasser, Paul Taggart, Paulina O. Espejo and Pierre Ostiguy. New York NY: Oxford University Press, 27–47.

Mudde, Cas, and Cristóbal R. Kaltwasser. 2013a. “Exclusionary vs. Inclusionary Populism: Comparing Contemporary Europe and Latin America.” *Government and Opposition* 48 (2): 147–74.

———. 2013b. “Populism.” In *The Oxford Handbook of Political Ideologies*. *Oxford handbooks in politics & international relations*, eds. Michael Freeden, Lyman T. Sargent and Marc Stears. Oxford: Oxford University Press, 493–512.

Müller, Jan-Werner. 2017. “Populism and Constitutionalism.” In *The Oxford Handbook of Populism*. *Oxford handbook of politics*, eds. Cristobal R. Kaltwasser, Paul Taggart, Paulina O. Espejo and Pierre Ostiguy. New York NY: Oxford University Press, 590–606.

Munck, G. L., and J. Verkuilen. 2002. “Conceptualizing and Measuring Democracy: Evaluating Alternative Indices.” *Comparative Political Studies* 35 (1): 5–34.

Munck, Gerardo L. 2009. *Measuring democracy : a bridge between scholarship and politics*. *Democratic transition and consolidation*. Baltimore, Md: Johns Hopkins Univ. Press.

Murphy, Gregory L. 2002. *The big book of concepts*. *Bradford Bks*. Cambridge, Mass: MIT Press.

Niemi, Richard G., Stephen C. Craig, and Franco Mattei. 1991. “Measuring Internal Political Efficacy in the 1988 National Election Study.” *The American Political Science Review* 85 (4): 1407.

Oliver, J. E., and Wendy M. Rahn. 2016. “Rise of the Trumpenvolk.” *The ANNALS of the American Academy of Political and Social Science* 667 (1): 189–206.

Pappas, Takis S. 2016. “Modern Populism: Research Advances, Conceptual and Methodological Pitfalls, and the Minimal Definition.” Oxford Research Encyclopedia of Politics.

Petty, Richard E., and Jon A. Krosnick. 1995. “Attitude Strength : Antecedents and Consequences.” Psychology Press.

Quinlan, Stephen, and Deirdre Tinney. 2019. “A Populist Wave or Metamorphosis of a Chameleon? Populist Attitudes and the Vote in 2016 in the United States and Ireland.” *The Economic and Social Review* 50 (2): 281–324.

Reeves, S. L., C. Varakamin, and C. J. Henry. 1996. “The relationship between arm-span measurement and height with special reference to gender and ethnicity.” *European journal of clinical nutrition* 50 (6): 398–400.

Rooduijn, Matthijs. 2014. “The Nucleus of Populism: In Search of the Lowest Common Denominator.” *Government and Opposition* 49 (04): 573–99.

———. 2019. “State of the field: How to study populism and adjacent topics? A plea for both more and less focus.” [en]. *European Journal of Political Research*. https://onlinelibrary.wiley.com/doi/full/10.1111/1475-6765.12314.

Roßteutscher, Sigrid, Rüdiger Schmitt-Beck, Harald Schoen, Bernhard Weßels, Christof Wolf, Ina Bieber, Lars-Christopher Stövsand, Melanie Dietz, Philipp Scherer, Aiko Wagner, Reinhold Melcher, and Heiko Giebler. 2019. “ZA6802: Pre- and Post-election Cross Section (Cumulation) (GLES 2017): version 3.0.1.” *GESIS Data Archive*.

Roßteutscher, Sigrid, Rüdiger Schmitt-Beck, Harald Schoen, Bernhard Weßels, Christof Wolf, Maria Preißinger, Agatha Kratz, Alexander Wuttke, and Lea Gärtner. 2018. *Short-term Campaign Panel 2017 (GLES)*. Trans. University of Mannheim, Chair for Political Psychology. GESIS Datenarchiv, Cologne: ZA6804.

Sartori, Giovanni. 1970. “Concept Misformation in Comparative Politics.” *The American Political Science Review* 64 (4): 1033–53.

———. 1984. “Guidelines for Concept Analysis.” In *Social science concepts: A systematic analysis*, ed. Giovanni Sartori. Beverly Hills Calif. u.a.: Sage Publ, 15–85.

Schedler, Andreas, and Rodolfo Sarsfield. 2007. “Democrats with adjectives: Linking direct and indirect measures of democratic support.” *European Journal of Political Research* 46 (5): 637–59.

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

Spierings, Niels, and Andrej Zaslove. 2017. “Gender, populist attitudes, and voting: Explaining the gender gap in voting for populist radical right and populist radical left parties.” *West European Politics* 40 (4): 821–47.

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.

Stanley, Ben. 2011. “Populism, nationalism, or national populism? An analysis of Slovak voting behaviour at the 2010 parliamentary election.” *Communist and Post-Communist Studies* 44 (4): 257–70.

Tsatsanis, Emmanouil, Ioannis Andreadis, and Eftichia Teperoglou. 2018. “Populism from Below: Socio-economic and Ideological Correlates of Mass Attitudes in Greece.” *South European Society and Politics* 47 (9): 1–22.

Urbinati, Nadia. 2013. “The Populist Phenomenon.” *Raisons politiques* 51 (3): 137.

van Hauwaert, Steven M., Christian H. Schimpf, and Flavio Azevedo. 2018. “Evaluating Existing Survey Measures of Demand-Side Populism using IRT.” In *The Ideational Approach to Populism: Concept, Theory,*

and Analysis, eds. Kirk A. Hawkins, Ryan Carlin, Levente Littvay and Cristóbal R. Kaltwasser. London: Routledge.

———. 2019. "The measurement of populist attitudes: Testing cross-national scales using item response theory." *Politics* Online first.

van Hauwaert, Steven M., and Stijn van Kessel. 2018. "Beyond protest and discontent: A cross-national analysis of the effect of populist attitudes and issue positions on populist party support." *European Journal of Political Research* 57 (1): 68–92.

van Kessel, Stijn. 2015. *Populist Parties in Europe: Agents of Discontent?* Basingstoke: Palgrave Macmillan.

Vehrkamp, Robert, and Christopher Wratil. 2017. *A Populist Moment?: Populist Attitudes of Voters and Non-Voters before the German Federal Election 2017*. Bertelsmann.

Weyland, Kurt. 2001. "Clarifying a Contested Concept: Populism in the Study of Latin American Politics." *Comparative Politics* 34 (1): 1–22.

Wong, Chi-Sum, Kenneth S. Law, and Guo-hua Huang. 2008. "On the Importance of Conducting Construct-Level Analysis for Multidimensional Constructs in Theory Development and Testing." *Journal of Management* 34 (4): 744–64.

Wuttke, Alexander. 2019. "Why Too Many Political Science Findings Cannot Be Trusted and What We Can Do About It: A Review of Meta-Scientific Research and a Call for Academic Reform." *Politische Vierteljahresschrift / German Political Science Quarterly* 60 (1): 1–19.

Supplementary Files: When the Whole is Greater Than the Sum of its Parts: On the Conceptualization and Measurement of Populist Attitudes and Other Multi-dimensional Constructs

Shiny Web Application

This study comes with an interactive web application. Providing a simple graphical user interface, the application allows readers to conduct analyses on the data that underlie this study. Specifically, the Shiny Web Application enables users to:

- Calculate the share of populists living in a country according to the Sartori approach, showing the sensitivity of the estimated share of populists to various thresholds
- Calculate the correlation of the Bollen and Goertz populism scores with substantive variables of interest (e.g., political interest, satisfaction with democracy) in multiple countries with multiple populism scales, showing the sensitivity of the estimated correlations to operationalization strategies.
- Show the internal structure of multiple populism scales (distribution and correlation of subdimensions and composite scores) in various countries.

The web application can be accessed at: http://populism.alexander-wuttke.de

Reproduction Material

All data that underlie this study and the analytical code with the data was analyzed can be accessed at: https://doi.org/10.7910/DVN/KPS1KY

Navigating the Supplementary Files

This study comes with extensive Supplementary Files (see Table of Content below).

- For practical guidance on specifying and operationalizing a multi-dimensional concept in any area of research, see *Supplement 2: Decisions to be made for the aggregation of multi-dimensional constructs*
- For a discussion on handling existing scales of populist attitudes, see *Supplement 6: Handling existing scales of populist attitudes*

Content

INTRODUCTION
WHAT ARE "POPULIST ATTITUDES"?
ON THE RELATIONSHIP BETWEEN CONCEPTS AND CONCEPT COMPONENTS
AGGREGATION RULES FOR NON-COMPENSATORY CONCEPTS SUCH AS POPULIST ATTITUDES
DATA AND OPERATIONALIZATION
ANALYSIS
CONCLUSION
REFERENCES

Shiny Web Application
Reproduction Material
Navigating the Supplementary Files
SUPPLEMENT 1: OVERVIEW OF OPERATIONALIZATION STRATEGIES IN THE LITERATURE ON POPULIST ATTITUDES
SUPPLEMENT 2: DECISIONS TO BE MADE FOR THE AGGREGATION OF MULTI-DIMENSIONAL CONSTRUCTS

#1 Defining the concept essence: Dimensionality of the concept
#2 Qualifier: Relationship of the concept components
#3 Quantifier: Scale of the target concept
#4 Typology of Concept Structures
#5 Aggregation Function
Bollen
Goertz
Sartori
Residual (Qualifier: Compensatory, Quantifier: Dichotomous)
Quantifier: In-between (Ordinal)
Qualifier: In-between (partly compensatory)
Example
#6 Normalization: Comparing raw or transformed concept subdimensions?
SUPPLEMENT 3: DISCREPANCIES BETWEEN ORIGINAL AND ADOPTED SCHULZ ET AL. POPULISM SCALE
SUPPLEMENT 4: QUESTIONNAIRES

Dataset: German Longitudinal Election Study, Campaign Panel 2017
Dataset: Castanho Silva et al. (2018) replication data set
Dataset: LISS
Dataset: National Election Survey with CSES module
SUPPLEMENT 5: CHOICE OF SCALES ON POPULIST ATTITUDES
SUPPLEMENT 6: HANDLING EXISTING SCALES OF POPULIST ATTITUDES

Akkerman et al. scale
CSES Wave 5 scale
Castanho Silva et al. scale
Schulz et al. scale
Oliver / Rahn scale
SUPPLEMENT 7: DESCRIPTIVE STATISTICS
SUPPLEMENT 8: NON-PROBABILITY AND PROBABILITY-BASED SURVEY DATA
SUPPLEMENT 9: HIGHER-ORDER FACTOR MODEL
SUPPLEMENT 10: VARIOUS OPERATIONALIZATIONS OF THE SARTORI CONCEPT STRUCTURE
SUPPLEMENT 11: SIMULATION OF THE RELATIONSHIP BETWEEN THE BOLLEN AND THE GOERTZ COMPOSITE SCORES
SUPPLEMENT 12: CORRELATIONS WITH CONSPIRATORIAL THINKING
SUPPLEMENT 13: CORRELATIONS WITH INSTITUTIONAL TRUST
SUPPLEMENT 14: ASSESSING NON-COMPENSATORY MULTI-DIMENSIONAL CONCEPTS ALREADY AT THE STAGE OF MEASUREMENT (AKKERMAN ET AL. SCALE)
SUPPLEMENT 15: OVERVIEW OF DATA SOURCES
REFERENCES

Supplement 1: Overview of operationalization strategies in the literature on populist attitudes

Table S1-1 provides a systematic overview of the concept structures of populist attitudes as employed in existing studies on the topic. We list the concept components according to how the concept components were reported in each study, adopting the authors' original labels. Quantifier, qualifier, and aggregation method represent our reading of the respective operationalization procedure. Note that the list is not necessarily an exhaustive list of all studies that have been published on populism at the mass level.

Three regularities stand out: First, most studies treat populist attitudes as a continuous concept. Second, implicitly or explicitly, most studies specify populism at the mass level as a set of ideas, that is: as an attitudinal syndrome with noncompensatory core components. Third, some kind of (exploratory or confirmatory) factor analysis is the most common aggregation method that scholars use to derive individual populism scores.

Table S1-1. **Overview of literature on populist attitudes**

Study	*Components*	*Qualifier (Concept Specification)*	*Quantifier*	*Aggregation method*
Akkerman, Mudde, and Zaslove (2014)	anti-elitism, people-centrism, sovereignty, Manichean division of good and evil	**necessity** ("set of ideas", p.1328), necessary and sufficient condition (p. 1326)	continuous	PCA
Andreadis, Stavrakakis, and Katsambekis (2016)	dealignment, anti-establishment, people-centrism, antagonism between people and elites	the "exact content of this (thin-centered) ideology depends on secondary elements that may be combined with such an (anti-establishment) appeal" (p.2)	continuous (p.4, 17)	Factor analysis
Elchardus and Spruyt (2016)	anti-elitism, antagonism ordinary people - politicians, people-centrism	**necessity** ("one component is always present..." "elements [...] that are always present", p. 113)	continuous	Factor analysis
Hameleers, Bos, and Vreese (2017)	anti-establishment, people-centrism, exclusionism	homogeneity as a necessary attribute (p. 483), but also: different emphases, indicating substitutability (p. 482)	continuous, but not aggregated	Factor analysis
Hawkins and Riding (2010)	Manichean outlook, people-centrism, anti-elitism	**necessity (**"set of ideas", p.1, 2)	continuous (p.12)	

Hawkins, Riding, and Mudde (2012)	Manichean outlook, reified popular will, (stealth attitudes)	**necessity** ("set of ideas", p. 3)	continuous (p. 14f)	Factor analysis
Hobolt and Tilley (2016)	anti-establishment, people-centrism, dissatisfaction of traditional parties, rejection of pluralism	"glorification of the people and anti-elitism (...) heart of populism" (p. 5) - implies necessity		
Mohrenberg, Huber, and Freyburg (2019)	anti-elitism, people-centrism	**necessity** "As Hawkins et al. (forthcoming) argue, so to classify an individual as having populist attitudes, both the dimensions 'anti-elitism' and 'people- centrism' need to be shared, albeit possibly to different degrees (cf. Castanho Silva et al. forthcoming). Only if an individual rejects the political elite as corrupt and self-serving (anti-elitism) and also perceives the people as homogeneous and morally superior (people-centrism), she can be called populist (cf. Rooduijn, de Lange, and Van Der Brug 2014, 567)."	continuous (p. 16)	EFA, geometric mean (p. 16)
Oliver and Rahn (2016)	anti-elitism, national affiliation, mistrust experts	not specified	no populism score derived	no populism score derived
Rico, Guinjoan, and Anduiza (2017)	two homogenous groups, praise of people, anti-elitism, sovereignty	**necessity** ("four distinct but interrelated constitutive elements ";p.4)	continuous (p.7)	Summary score
Schulz et al. (2018)	Anti- Elitism attitudes, people's sovereignty, belief in a homogenous and virtuous people	**necessity** ("set of ideas", p.2)	continuous (p. 5)	Factor analysis
Castanho Silva et al. (2018)	Homogeneity, Anti-Elitism, Manichean Outlook	"we suggest that populism sits at the intersection of these three broader kinds of discourse"		Factor analysis
Spierings and Zaslove (2017)	people-centrism, anti-elitism, manichean antagonism between people and establishment, notion of general will	"core characteristics of populism" (p. 824f) "this notion of populism is what unifies populist radical left and populist radical right parties ..." (p.825)	continuous (p. 831, 832)	Factor analysis
Spruyt, Keppens, and van Droogenbroeck (2016)	Manichean Outlook, Anti-Elitism, people-centrism, sovereignty	**necessity** ("combination"; "set of ideas" p.336; "it is their combination that constitutes the specific populist logic", p. 340)	continuous (p. 340)	Factor analysis
Stanley (2011)	anti-elitism, people-centrism, people's sovereignty (nationalism)	**necessity** "four core concepts the combination of which is characteristic of all manifestations of populism" (p.258)	continuous (p.263)	Factor analysis
Steiner and Landwehr (2018)	Anti- Elitism, illiberal democracy, anti-pluralist skepticism,	**Necessary condition** of anti-elitism as the core of populism "Anti-Elitismus [...] in Verbindung mit" (p. 5);	continuous (p. 13, (p.16)	Factor analysis

	majoritarianism, trustee model of political representation			
van Hauwaert and van Kessel (2018)	dissatisfaction of people, anti-elitism, Manichean distinction between virtuous people and evil elites	implicitly necessary conditions of anti-elitism and people-centrism but different levels etc. (p.70) "These characteristics do not constitute elements of populism per se, but can be linked back to the ideational properties of populism (...) antagonistic position vis-a-vis elites and its appeal to the common people."	continuous (p.76)	Factor analysis
Vehrkamp and Wratil (2017)	Anti-Establishment, Anti-Pluralism, Popular Sovereignty	**necessity**	dichotomous (and ordinal with three categories, p. 16)	Thresholds

Supplement 2: Decisions to be made for the aggregation of multi-dimensional constructs

The analysis in the main text focuses on operationalization techniques in research on populist attitudes. Yet the study's general argument is applicable to various social science concepts, namely those positioned at the intersection of the concept components, thus requiring operationalization techniques that account for this concept property. As mentioned in the main text, democracy is one illustrative political science example that also features necessary conditions on the level of the concept attributes: high levels of minority protection do not compensate for low levels of electoral fairness because both components constitute necessary elements of liberal democracies (Møller and Skaaning 2012, 135). Cognitive competencies may be considered as another example from a different scientific discipline (Chalmers and Flora 2014, 341): Imagine a task such as math word tests that require both arithmetic knowledge and text comprehension skills. Any measure used to reliably predict a person's success at that task warrants an operationalization strategy that takes into account that the presence of both competence components constitutes a necessary condition for solving math word problems.

Considering the prevalence of multi-dimensional concepts with necessary conditions in the social sciences, this supplement aims at providing guidance for researchers from diverse disciplinary backgrounds on the appropriate operationalization of multi-dimensional concepts. In an attempt to make the description accessible, we distinguish several steps in a sequence of decisions that may help scholars to arrive at deliberate and well-justified choices when operationalizing multi-dimensional constructs (for further reading, see Alkire and Foster 2011b; Goertz 2006).[11] In the following, we do not consider strategies for developing new measures for multi-dimensional concepts. Instead, we discuss the proper operationalization of multi-dimensional concepts using existing measures.

#1 Defining the concept essence: Dimensionality of the concept

The dimensionality of a concept is not always obvious, neither concerning the number nor the specific content of the subdimensions. The following arguments only apply to 1) multi-dimensional concepts and require 2) that it is possible to identify clearly specified concept subdimensions.

Questions to answer to choose the appropriate operationalization strategy:

- ✓ Has the concept two or more concept subdimensions?

[11] Note that the following discussion assumes equal relevance of the concept dimensions. For the operationalization of concepts for which this assumption is violated, see, for instance, Alkire and Foster (2011a).

- ✓ What are the concept subdimensions?

#2 Qualifier: Relationship of the concept components

The relationship between concept components can be characterized by varying degrees of interchangeability between the concept components, ranging from non-compensatory to fully interchangeable (qualifier). Which qualifier corresponds to a given concept can only be derived from the theoretical propositions associated with the essence of a concept. When determining the numerical value of a target concept, the question is whether high values on one concept component (partly) compensate low values on another concept component. When measuring democracy, for instance, whether we consider the rule of law an indispensable element of democracy (or whether its absence can be compensated for by, for instance, extensive participatory rights for the public in political or judicial decision-making processes) reflects the researcher's reasoning of what democracy is in its essence.

If the concept's subdimensions are considered non-compensatory, then the presence of all subdimensions constitute necessary conditions. If the lack of one concept component can be fully compensated by another concept component, then each concept component is fully interchangeable by another. In addition to these extreme cases, there are degrees of interchangeability. For example, very high values on one dimension cannot fully compensate for very low values on another subdimension, but partly make up for those very low values.

Determining the degree of interchangeability among the concept components is an integral element of devising a proper operationalization procedure. Practically, the specification of a concept qualifier establishes the aggregation function that is to be used for aggregating multiple concept components into a composite score (see *#5* aggregation function).

Questions to answer to choose the appropriate operationalization strategy:

- ✓ Interchangeability: Do the theoretical propositions toward the essence of the concept consider each concept component as necessary, or can values on one concept component compensate values on another concept component?
- ✓ If concept components are not necessary, to which degree (e.g., partly or fully) are the concept components interchangeable?

#3 Quantifier: Scale of the target concept

A concept may be dichotomous, continuous, or something in between (quantifier). Again, which quantifier corresponds to a given concept is to be derived primarily from the theoretical

propositions associated with the essence of a concept. The question is whether the concept has clear membership boundaries (and, if so, how many) or whether the concept entails grey space between the poles. For instance, whether we consider 'gender' a dichotomous, polytomous or continuous concept reflects our reasoning of what gender is in its essence.

In some instances, however, practical considerations may also play a role when determining the scale of the target concept. Precisely, the concept essence may imply a continuous quantifier ('temperature'), but researchers may still revert to a dichotomous quantifier ('warm') for simplicity. Concepts with a dichotomous quantifier may simplify the analysis and the reporting of results. However, these advantages come at the expense of overlooking potentially meaningful variation that the concept exhibits empirically or theoretically. Hence, opting for a dichotomous quantifier in light of an originally continuous concept entails a trade-off between simplicity and precision. In any case, specifying a concept quantifier must flow from theoretical reasoning about which quantifier is suitable for the concept essence.

Questions to answer to choose the appropriate operationalization strategy:

- ✓ Interchangeability: Do the theoretical propositions about the essence of the concept consider clear membership boundaries (and, if so, how many), or does the concept entail grey space between the poles?

#4 Typology of Concept Structures

Having identified the dimensionality of a concept (step *#1*), the decisions at steps *#2* and *#3* establish the concept structure for a given multi-dimensional concept.

Copying a table from the main text, Table S2-1 provides an overview of prototypical concept structures. Table S2-1 may be helpful to situate one's concept in the universe of potential concept structures. Having a clear understanding of the concept structure is vital because each concept structure has different practical implications for the decisions in the following steps (step *#5*: aggregation function; step *#6* standardization).

Note that Table S2-1 reports ideal types. As stated above, both the qualifier and the quantifier offer more than two options, whereas, for the sake of simplicity, the table only shows four combinations of quantifiers and qualifiers.

Questions to answer to choose the appropriate operationalization strategy:

- ✓ Where does the concept fall on the continuums of concept qualifier and concept quantifier?

- ✓ Does the combination of quantifier and qualifier of a given concept correspond to any of the prototypical concept structures?

Table S2-1. **Prototypical concept structures**

		Quantifier of concept structure	
		Dichotomous	*Continuous*
Qualifier of Concept Structure	*Non-compensatory*	Sartori	Goertz
	Compensatory	Residual	Bollen

Note: Table S2-1 maps prototypical concept structures, but in practice, quantifier and qualifier have more than two manifestations. Here, the qualifier contrasts no substitutability (necessity conditions) with medium substitutability, but higher and lower degrees of substitutability are also conceivable. Likewise, this table compares two prototypical quantifiers, although ordinal quantifiers are also conceivable.

#5 Aggregation Function

To ensure that the mathematical structure of a measure and the theoretical structure of the concept do not fall apart (cf. Goertz 2006, 125), the aggregation function for computing the target measure needs to correspond to the theoretically derived concept structure. Because social science concepts consist of a basic level (the concept essence), a second level (the concept components or subdimensions) and a third level (the indicators at the measurement level), we need to distinguish operationalization strategies for aggregating indicators at the concept's third level and for aggregating concept components at the concept's second level. The following discussion does not concern the aggregation of multiple indicators of a concept component into a summary score.[12] Instead, it deals with the aggregation of the concept components to derive an aggregate score for the target concept.

First, we discuss aggregation functions for the prototypical concept structures, as depicted in Table S2-1. Then, we discuss aggregation functions for other concept structures. At the end of the section, we provide a simple mathematical example to compare aggregate scores as computed with each of the aggregation functions presented below.

Bollen

The Bollen structure is a widely employed operationalization strategy both in the literature on populist attitudes and beyond. Therefore, techniques for computing a Bollen concept structure are well known in the social sciences (e.g., Bollen and Lennox 1991; Bollen and Pearl 2013; Kaplan

[12] In principle, the general line of reasoning would apply. In practice, in most cases the Bollen concept structure may be most appropriate to aggregate indicators at the concept's third level.

2008). These techniques include any operational procedure that sums the values of the concept components, including weighted or unweighted summary indexes, predicted scores using the factor scores derived from exploratory factor analysis or structural equation modeling.

$$Bollen := \sum_{i=1}^{n} Weight_i * Component_i$$

Goertz

According to the Goertz concept structure, the resulting target concept is located at the intersection of the concept components. Hence, the operationalization of the Goertz concept structure requires an aggregation function that treats the concept components as non-compensatory. The mathematical equivalent to the theoretical tenet that 'a chain is only as strong as its weakest link' is the minimum function. Using the minimum function ensures that two individuals with identical values on the lowest concept components do not differ in the derived aggregate score, regardless of the values on the remaining concept components. Other aggregation functions are conceivable if no strict non-substitutability is assumed; see *Qualifier: In-between (partly compensatory).*

$$Goertz := \min |Component_{1,} \dots, Component_n|$$

Sartori

The Sartori concept structure entails crisp membership boundaries and considers entities as members of the concept if and only if all concept components are jointly present. The necessary conditions are met if all concept components surpass a pre-specified threshold.

$$Sartori := \begin{cases} 1 & if\ Component_{1,} \dots, Component_n > Threshold \\ 0 & else \end{cases}$$

The challenge in operationalizing the Sartori concept structure is to specify a substantively meaningful, non-arbitrary threshold (Alkire and Foster 2011a, 482).[13] One option for setting a threshold is letting the data decide. Such data-driven methods can be carried out manually or automatically. Manually, one could set thresholds at a certain percentile of the distribution (e.g., upper half, top 10 percent). A manual data-driven approach entails a membership classification procedure that depends on the sample distribution. Because such manual data-driven procedures

[13] One response to the challenge of threshold selection may also involve robustness tests using specification curves or other means of transparent reporting various analytical options, see Wuttke 2019, 11.

do not assign membership by absolute properties of the concept essence, individuals may surpass the classification threshold in one sample but not in a different sample where the component is distributed differently (see *#6 Normalization: Comparing raw or transformed concept subdimensions?*).

An automated data-driven approach, in contrast, would rely on algorithmic classification, and examples are methods such as latent class analysis or latent profile analysis (Collins and Lanza 2010). Automated data-driven techniques for class selection are somewhat more common in the literature than manual techniques (e.g., Bonikowski and DiMaggio 2016). However, with regard to any data-driven approach, one might argue that "*such purely data-driven calibration strategies are fundamentally flawed […]. Measures like the mean or median are properties of the data at hand and, as such, void of any substantive meaning vis-à-vis the concept that one aims to capture with a set*" (Schneider and Wagemann 2012, 33).

In order to set thresholds based on substantive arguments, one might consider the scales of the measures that underlie the concept components. Given dichotomous concept components, the choice of a cut-off point is obvious.[14] When the concept components are measured on a scale with more than two categories, the choice of a cut-off point is often not straightforward. In these cases, setting a cut-off point is most intuitive when the concept components are measured on a meaningful scale as such scales may offer thresholds with informational value. For instance, surpassing the mid-point of a scale may denote a meaningful step that could serve to discriminate membership boundaries. Concept components that were measured on Likert-scales could also provide opportunities for meaningful cut-off points. In the Shiny Web Applet ('How many populists?'), we provide examples of meaningful cut-off points using various Likert-scale survey items on populist attitudes.

Residual (Qualifier: Compensatory, Quantifier: Dichotomous)

Operationalizing a concept with a dichotomous scale with interchangeable concept components draws on both the operationalization techniques used for the Sartori and the Bollen concept structures. First, the concept scores are derived using any weighted or unweighted summation technique for operationalizing Bollen concept structures. Akin to the technique for the operationalization of Sartori concept structures, a meaningful threshold is then chosen to delineate

[14] Note that research on the measurement of poverty has suggested more complex, two-stage thresholds, see Alkire and Foster (2011a; 2011b).

crisp membership boundaries. In practical terms, this approach differs from the Sartori operationalization as dichotomization is conducted at the level of the derived target concept and not on the level of the concept components (Alkire and Foster 2011a, 478). Substantively, therefore, the resulting concept is dichotomous, but unlike the Sartori concept structure, it does not represent the conjunction of the concept components.

Quantifier: In-between (Ordinal)

In case the concept at hand does not fall under any of the prototypical concept structures discussed above but entails an ordinal scale, the operationalization strategies for dichotomous quantifiers may be applied separately for each category of the concept.

Qualifier: In-between (partly compensatory)

The Goertz and Sartori concept structures entail strict non-substitutability on the level of the concept components. However, it is conceivable that concept components are partly interchangeable. Various aggregation strategies are conceivable that differ in technical details. Which aggregation function fits best a given concept, therefore, depends on the assumptions about the degree of substitutability of the concept components. In the following, we present a short description of selected aggregation functions for concepts with partial interchangeability. For further information readers may resort to details provided in Alkire et al. 2015; Alkire and Foster 2011a; Atkinson 2003; Goertz 2006, 111ff; Greco et al. 2019; Mazziotta and Pareto 2016; Møller and Skaaning 2012, 122ff; Munck 2009, 48ff:

- *Multiplication:* Multiplying all concept subdimensions implements partial interchangeability at the advantage of being easy to use and to comprehend. How multiplication implements non-substitutability becomes apparent when one concept component equals zero. In that case, the function is fully non-compensatory because the aggregate score equals zero regardless of the values of the other concept components. Note that multiplication introduces a non-linear function between the values of the concept components and the aggregate score, exacerbating differences between individuals.

- *Weighted arithmetic mean:* In contrast to the arithmetic mean (used in the main text to operationalize the Bollen approach), which weighs all components equally, the weighted arithmetic mean allows unequal weights. To implement a certain degree of non-interchangeability, one might place a higher weight on the term with the lowest value. In this vein, the minimum can be understood as a special case of the weighted arithmetic mean which puts all the weight on the term with the lowest value.

- *Geometric mean:* The geometric mean is the *n*th root of the product of *n* concept components. The geometric mean resembles the arithmetic mean, but it multiplies the concept components instead of adding them. The geometric mean is occasionally used in existing research to account for partly non-compensatory relationships among concept components. For instance, in 2010, the United Nations Human Development Index introduced the geometric mean because "*low achievement in one dimension is not linearly compensated for […] by high achievement in another dimension. The geometric mean reduces the level of substitutability between dimensions and at the same time ensures that a 1 per cent decline in the index of, say, life expectancy has the same impact on the HDI as a 1 per cent decline in the education or income index*" (United Nations 2019).

- *Counting approach and two-level thresholds*: Various other sophisticated methods in operationalizing varying degrees of substitutability for multi-dimensional concepts with a dichotomous quantifier were recently proposed in poverty research. These proposals include counting the number of dimensions on which an observational unit did (not) surpass a threshold and the use of two thresholds, see Alkire et al. 2015; Alkire and Foster 2011a; 2011b; Atkinson 2003; Mazziotta and Pareto 2016.

The discussion above considers operationalization strategies for concepts which are constituted by varying degrees of necessary conditions on the level of the concept components. However, at the other end of the substitutability-continuum (qualifier), sufficient conditions are conceivable as well (Goertz 2006). Scholars in socio-economic topics, for instance, discuss sufficient conditions using set-theoretic terms as *union* (as opposed to an *intersection*, representing necessary conditions). In this vein, for classifying a person as deprived, it is sufficient to fall below the threshold on only one of the multiple achievement dimensions. For further reading and operationalization strategies, see Alkire et al. 2015; Alkire and Foster 2011a; 2011b; Atkinson 2003; Greco et al. 2019; Mazziotta and Pareto 2016.

Example

For a hypothetical distribution of data among three observational units (e.g., respondents, countries), Table S2-I shows the resulting aggregate scores for a hypothetical concept with three concept components when different aggregation functions are employed. Similar to the calculations conducted in the Shiny Web Application, Table S2-I shows the sensitivity of resulting aggregate scores to the employed aggregation function. In addition to the aggregate score, Table S2-I also reports how each observational unit ranks compared to the other observational units with respect to these scores. Even though the values of the concept components remain unchanged, the rank

order of observational units is not consistent across aggregation functions. This observation underscores the importance of carefully choosing the aggregation function that corresponds to the concept's essence.

Table S2-I. **Illustrative computation of aggregate scores using different aggregation functions**

	Observational unit 1			Observational unit 2			Observational unit 3		
	Compo-nent 1	Compo-nent 2	Compo-nent 3	Compo-nent 1	Compo-nent 2	Compo-nent 3	Compo-nent 1	Compo-nent 2	Compo-nent 3
	0.2	0.2	0.3	0.2	0.5	1	0.3	0.3	0.6
	Score		*Rank*	*Score*		*Rank*	*Score*		*Rank*
Goertz [Minimum]	0.2		2	0.2		2	0.3		1
Bollen [Unweighted Mean]	0.23		3	0.5		1	0.4		2
Sartori [Threshold at 0.75]	0		1	0		1	0		1
Multiplication Weighted Average Mean [Low 0.5 Mid: 0.3; High: 0.2]	0.01		3	0.1		1	0.05		2
	0.07		3	0.45		1	0.27		2
Geometric mean	0.23		2	0.23		2	0.38		1

#6 Normalization: Comparing raw or transformed concept subdimensions?

Regardless of which concept structures a given concept amounts to when combining multiple dimensions into an aggregate score, all dimensions must be measured on a common or on a comparable scale (Edwards 2009, 519ff). Implicitly, all aggregation functions discussed in the previous section assume scale comparability. The minimum function, for instance, rests on the assumption that the value 0.5 has the same meaning on different concept components. In an ideal case, scale comparability was already considered during scale development. When using existing scales, however, one must judge their comparability even if they were not developed for the purpose of aggregation.

If scales are not comparable on their raw metric, one option is to use standardized transformations of the concept components (e.g., using z-score transformations). However, this procedure has the drawback of substantively changing the meaning of the scales on which the concept components are measured. Specifically, high or low levels on one concept dimension were then to be interpreted only in reference to the distribution within a given population. Consequently, scores in one population cannot necessarily be compared with scores on the same concept in a different population (or a different point in time).

Hence, the issue of scale comparability is cumbersome and often does not offer satisfactory options. Qualitative researchers have a long tradition of dealing with ‘calibration’ (Goertz 2006; Schneider and Wagemann 2012, 32ff). In the quantitative social science research culture,

‘equating’ and ‘scaling’ are now being discussed particularly in the realm of Item Response Theory (Edwards 2009, 518ff; Kolen and Brennan 2004, 198ff). Recent methodological developments in Item Response Theory also consider noncompensatory multi-dimensional constructs (Babcock 2011; Bolt and Lall 2003; Chalmers 2012; 2018; Chalmers and Flora 2014; DeMars 2016; Liu and Chalmers 2018). Even though the proposed methods are complex, computationally demanding, and require a large number of measurements for each concept dimension (Babcock 2011; Bolt and Lall 2003; Liu and Chalmers 2018), further methodological developments may prove helpful for solving scaling issues concerning both compensatory and noncompensatory concept structures.

- ✓ Are the scales of the subdimensions comparable, and, if not, what techniques will be employed to ensure comparability?

Supplement 3: Discrepancies between original and adopted Schulz et al. populism scale

In the main text, we report evidence on the Akkerman et al., Schulz et al., and the Castanho Silva et al. populism scale. Yet the analysis of the Schulz et al. populism scale that is reported in the main text uses data from the GLES campaign panel, which did not employ the original Schulz et al. populism scale but a modified version of the scale. The GLES version of the scale only contains nine items whereas the original Schulz et al. scale consists of 12 items. In addition, differences occur with respect to the wording of one item of the anti-elitism subdimensions and a systematic disparity in translating the homogeneity items from English to German (see Table S3-1).

The discrepancy in the anti-elitism subdimension concerns the item "*Politicians are not really interested in what people like me think*", which was originally included in the Schulz et al. scale's 15-item long version but the item was dropped during scale validation and was not included in the final list of twelve items. Nonetheless, the principal investigators of the German Longitudinal Election Study's (GLES) Campaign Panel included a modified version of that item ("*Politicians care about what ordinary people think*") because the long-running GLES survey program already contained a similar item. As an additional discrepancy, the respective item is positively worded in the adapted version of the GLES Campaign Panel (high agreement with the item indicates lower agreement with the tenets of populism) whereas the item in the original Schulz et al. scales is negatively worded (in the same direction as all other scale items). Because goodness of fit was a reason for the original scale authors to drop the respective item (Schulz et al. 2018, 322) and because the adapted version of the item is in reverse compared to the other items of the anti-elitism subdimension, it is possible that the anti-elitism subdimension as measured in the GLES Campaign Survey exhibits a lower degree of internal consistency compared to survey data that uses the original items on the anti-elitism subdimension.

Another discrepancy concerns the homogeneity dimension, namely the translation of 'ordinary people' in the German question-wording. Originally, the term was translated into German as 'einfache Leute' which could be back-translated as 'simple people'. In contrast, the adapted scale as used in the GLES Campaign Panel refers to 'normale Bürger' which could be back-translated as 'normal citizens'. Altogether, the wording in the GLES Campaign Panel may be understood as a more positive description of the respective group, potentially eliciting more people to identify with the respective group. Consequently, due to these wording discrepancies between

the original and the adapted scale, the level of agreement with the homogeneity items may be higher for the adapted scale compared to the original version.[15]

As argued above, it is conceivable that these issues affect the empirical properties of the Schulz et al. populism scale as it is measured in the GLES Campaign Panel. Specifically, it could be the case that the internal consistency of the anti-elitism subdimension and the level of agreement with the homogeneity subdimensions would differ if the original instrument was employed. While these issues should be kept in mind for anyone who intends to use the GLES data, it is not clear that these issues necessarily affect the estimands of this study. If we understand the primary estimand of the main text as the difference between the Bollen and Goertz populism scores, then the issues discussed above do not have apparent ramifications for the reported findings. Different levels of agreement with the homogeneity indicators, for instance, would affect the comparison of Bollen and Goertz scores only if the discussed issues regarding item translation affected a person's homogeneity scores differentially, depending on the scores on other subdimensions.

Table S3-1. **Question wordings in the original Schulz et al. scale and the adapted scale in GLES Campaign Panel**

Adapted scale	*Original version*	*Difference*
Anti-elitism		
Politicians talk too much and take too little action.	Politicians talk too much and take too little action.	Identical
The differences between the people and the so-called elite are greater than within the people.	The differences between people and the ruling elite are much greater than the differences between ordinary people.	Identical
[Politicians care about what ordinary people think.]	Politicians are not really interested in what people like me think.	Reverse 'Ordinary people' vs 'people like me'
Homogeneity		
Ordinary people are of good and honest character.	Ordinary people are of good and honest character	Different translation of 'ordinary people'
Ordinary people all pull together.	Ordinary people all pull together	Different translation of 'ordinary people'
Ordinary people share the same values and interests.	Ordinary people share the same values and interests.	Different translation of 'ordinary people'
Sovereignty		
The people should have the final say on the most important political issues by voting on them directly in referendums.	The people should have the final say on the most important political issues by voting on them directly in referendums.	Identical

[15] Note that Hieda, Zenkyo, and Nishikawa (2019) also found low correlations between the homogeneity subdimensions and the other subdimensions in a Japanese sample of respondents using the Schulz et al. (2018) scale.

The people, not the politicians, should make our most important policy decisions.	The people, not the politicians, should make our most important policy decisions.	Identical
The politicians in Parliament need to follow the will of the people.	The politicians in Parliament need to follow the will of the people.	Identical

Figures S3-1 through FS3-10 provide suggestive empirical evidence on whether the discrepancies between the original Schulz et al. scale and the adapted version might have affected the results that are reported in the main text. Figure S3-1 (below) was also plotted in the main text and shows the empirical properties of the adapted Schulz et al. populism scale in the GLES Campaign Panel. We can compare the results in Figure S3-1 with evidence using the original scale items as they were measured in various countries in the Castanho Silva et al. dataset (Figure S3-2 through Figure S3-10).[16] When comparing these results, however, we should keep in mind that the reported findings in all of these plots do not only differ in whether the adapted or the original scale was used but also differ in the employed samples (survey firm, country). Nonetheless, despite variation in the general strength of the association in each sample, we see similar patterns when comparing findings with the adapted and the original scale: Above all, the homogeneity subdimension behaves distinctly even when the original scale is used, exhibiting lower correlations than the other two subdimensions.

[16] Note that Castanho Silva et al. employed the Schulz et al. populism scale in a reduced version as well.

Figure S3-1. **Distribution of and correlations between concept structures of populist attitudes and concept attributes (Germany, adapted Schulz et al. populism scale)**

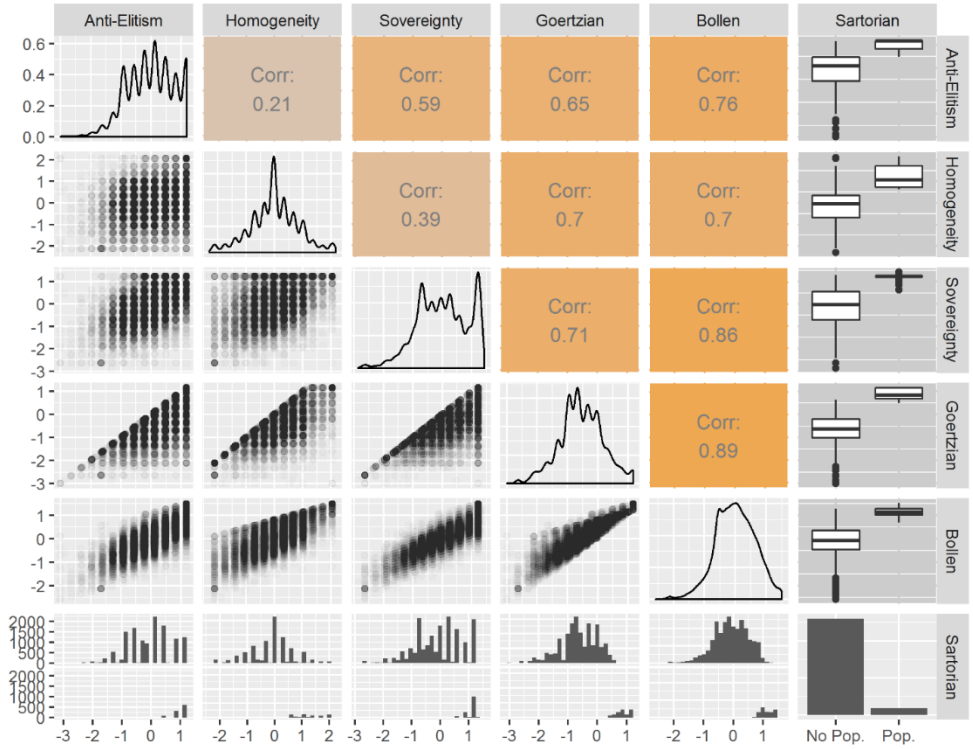


Figure S3-2. **Distribution of and correlations between concept structures of populist attitudes and concept attributes (US, original Schulz et al. populism scale)**

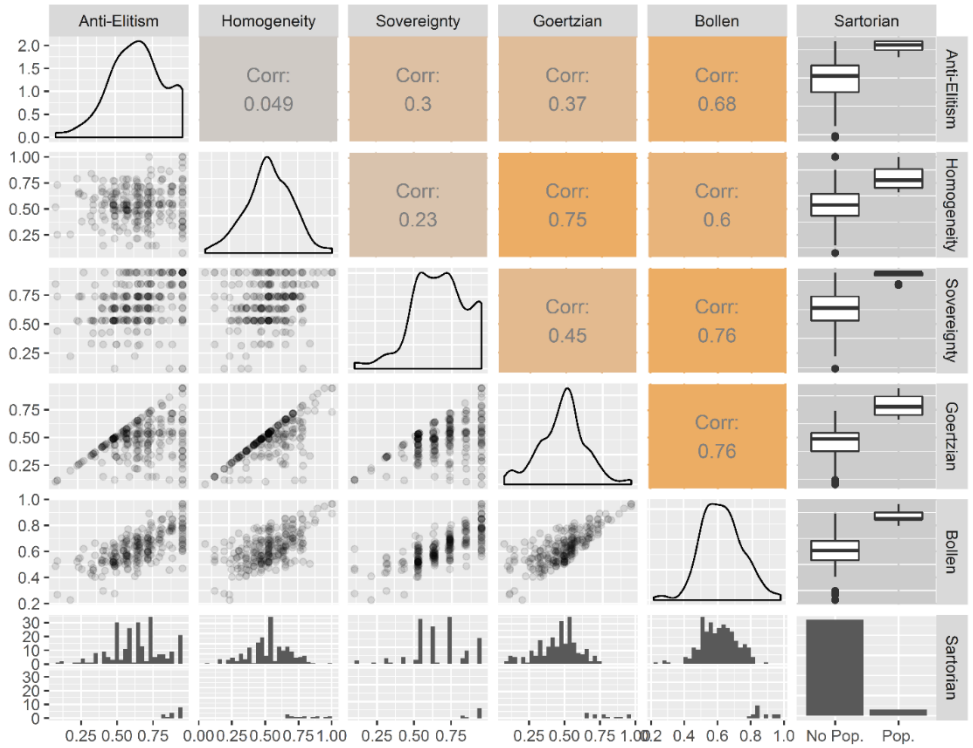


Figure S3-3. **Distribution of and correlations between concept structures of populist attitudes and concept attributes (Ireland, original Schulz et al. populism scale)**

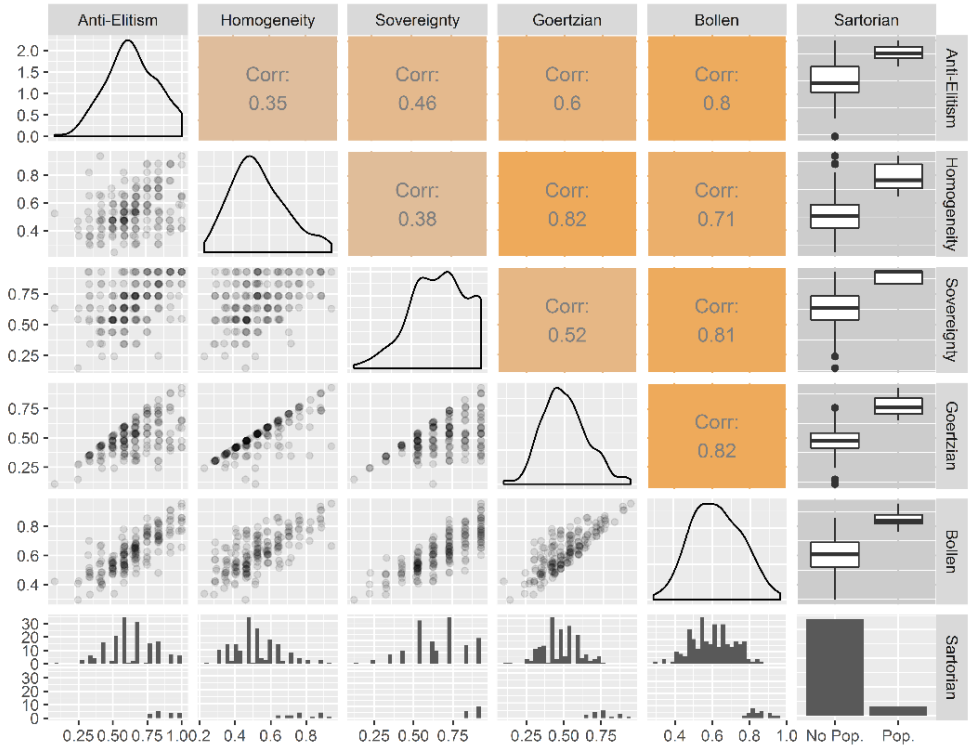


Figure S3-4. **Distribution of and correlations between concept structures of populist attitudes and concept attributes (Italy, original Schulz et al. populism scale)**

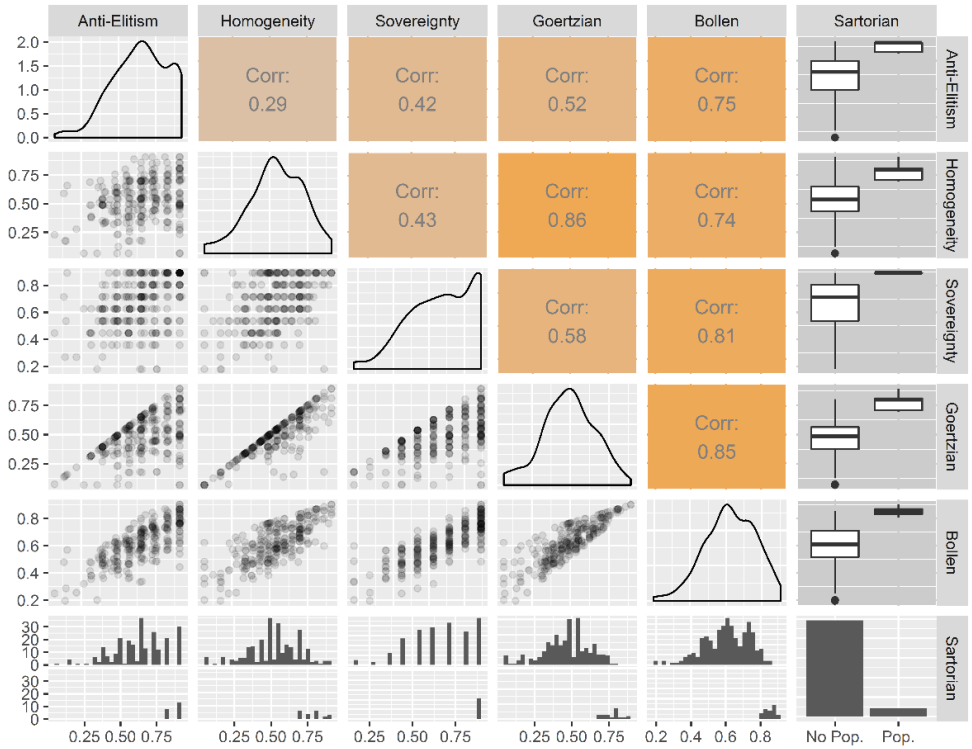


Figure S3-5. **Distribution of and correlations between concept structures of populist attitudes and concept attributes (Mexico, original Schulz et al. populism scale)**

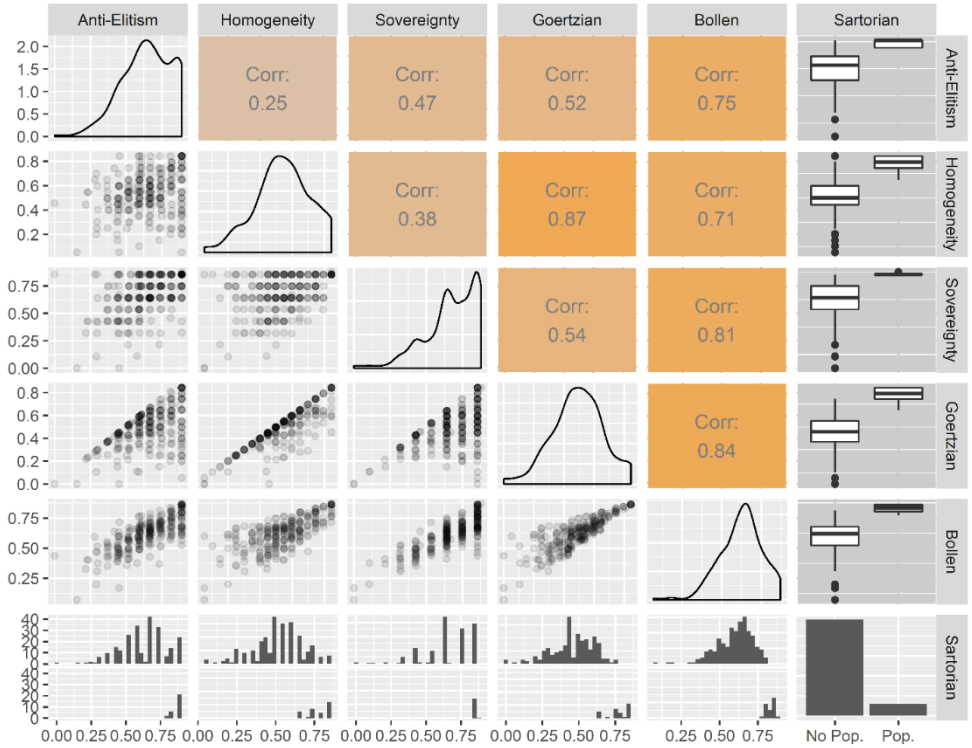


Figure S3-6. **Distribution of and correlations between concept structures of populist attitudes and concept attributes (Greece, original Schulz et al. populism scale)**

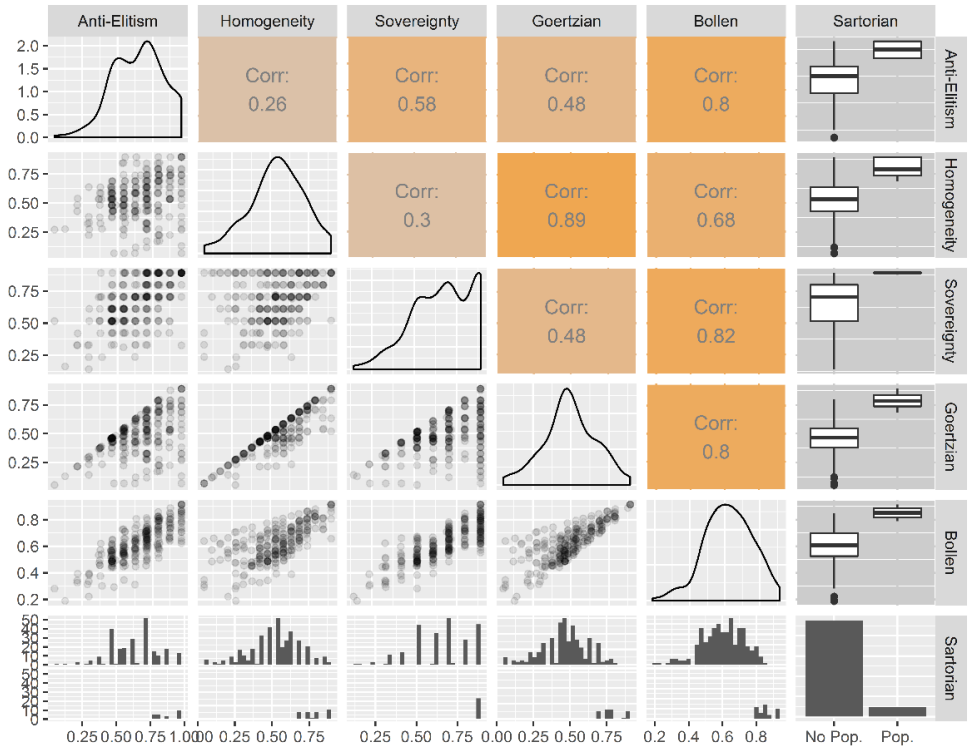


Figure S3-7. **Distribution of and correlations between concept structures of populist attitudes and concept attributes (Spain, original Schulz et al. populism scale)**

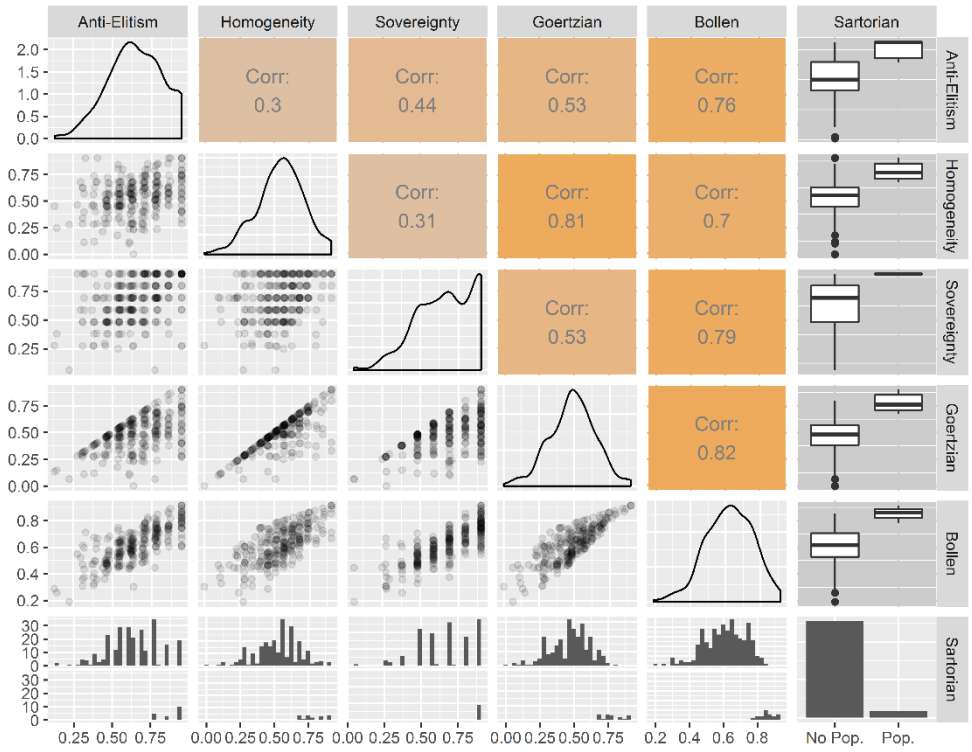


Figure S3-8. **Distribution of and correlations between concept structures of populist attitudes and concept attributes (UK, original Schulz et al. populism scale)**

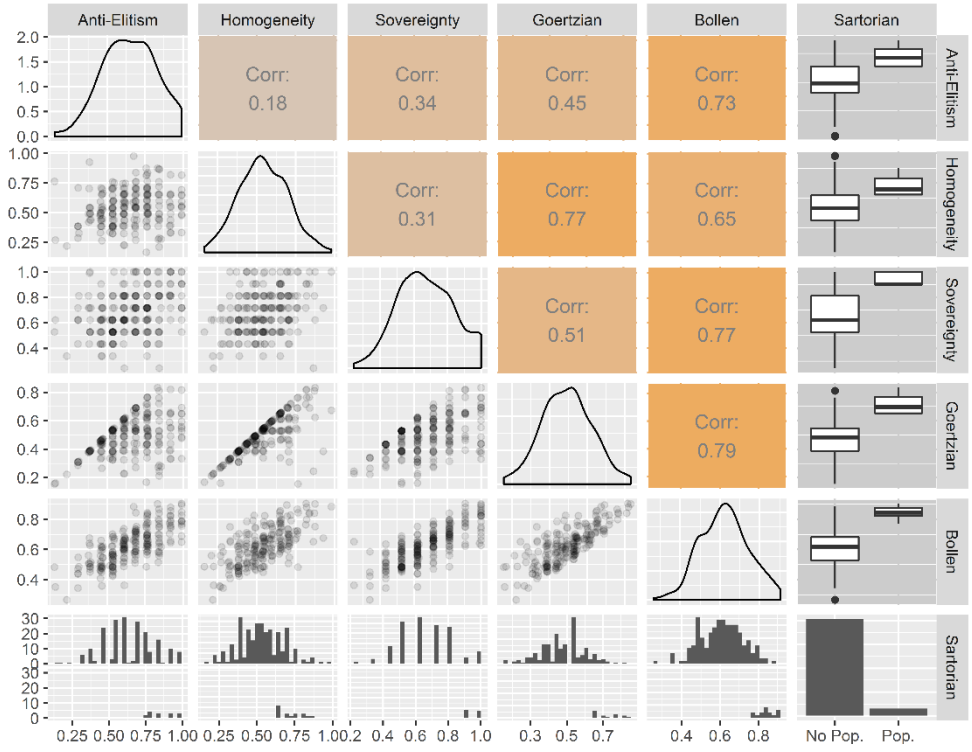


Figure S3-9. **Distribution of and correlations between concept structures of populist attitudes and concept attributes (France, original Schulz et al. populism scale)**

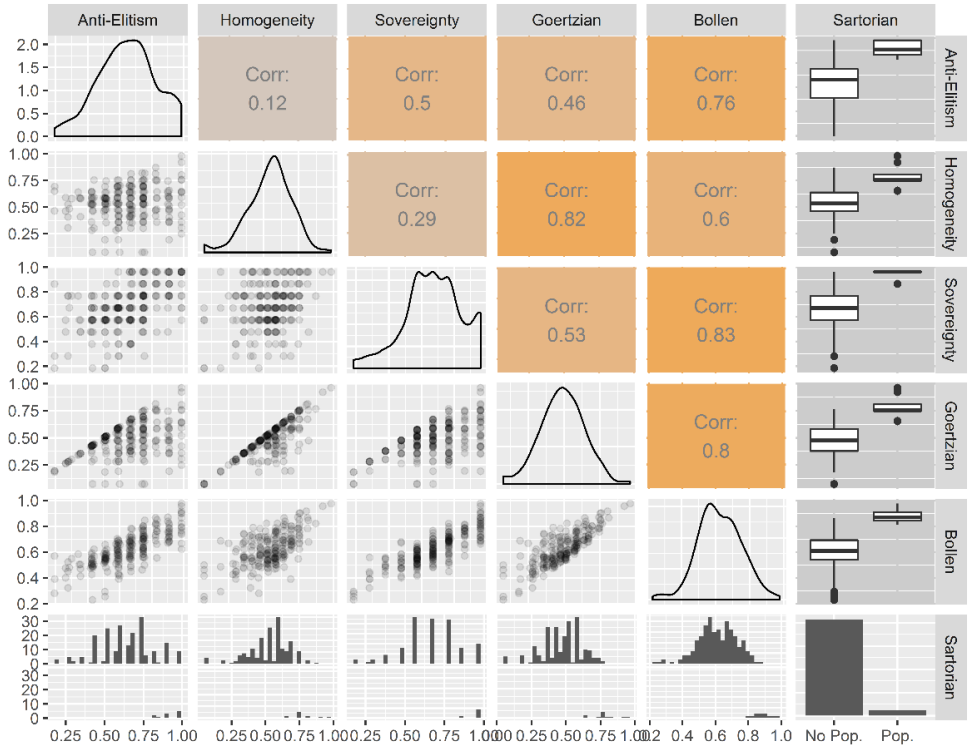
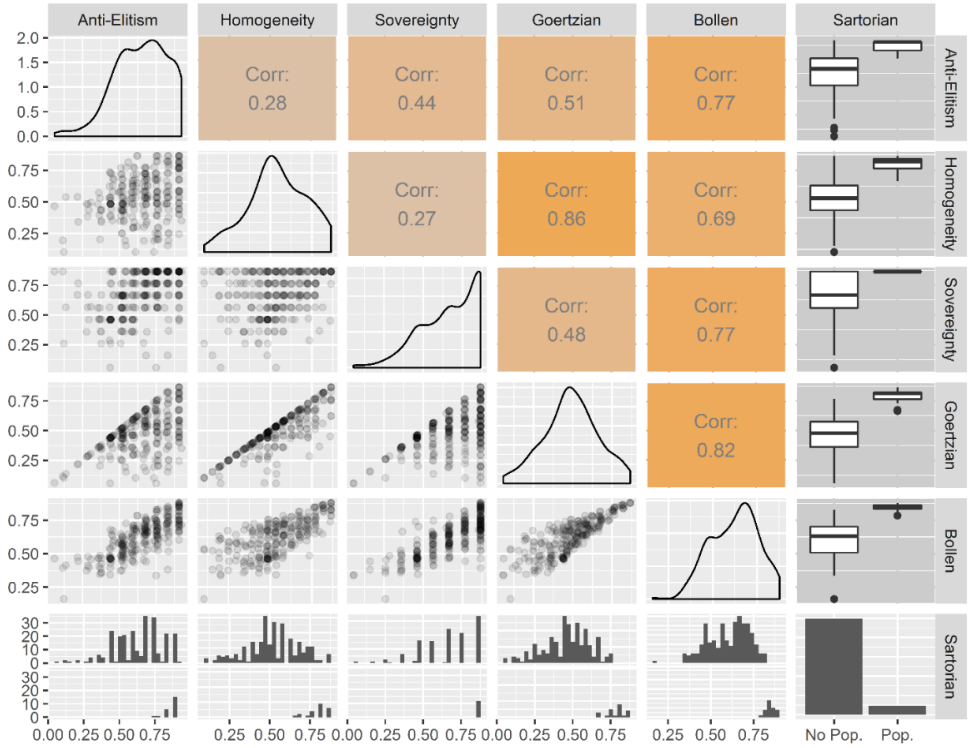


Figure S3-10. **Distribution of and correlations between concept structures of populist attitudes and concept attributes (Brazil, original Schulz et al. populism scale)**



Supplement 4: Questionnaires

This supplement reports the wordings of the questions of the items for measuring populist attitudes. See *Supplement 6: Handling existing scales of populist attitudes* for a discussion on operationalizing existing scales of populist attitudes.

Intro = introductory sentences
QT = question text
RO = response options

Dataset: German Longitudinal Election Study, Campaign Panel 2017

Note that the Campaign Panel contains an adapted version of the original Schulz et al. populism scale.

Schulz et al. (2018)

Intro*:* *Hier sind weitere Meinungen über Politik und Gesellschaft, denen manche Menschen zustimmen und andere nicht.*
QT: *Geben Sie bitte an, ob Sie diesen Meinungen zustimmen oder nicht.*
(A) Politiker reden zu viel und handeln zu wenig. [Anti-Elitism]
(B) Die normalen Bürger verbindet ein guter und ehrlicher Charakter. [Homogeneity]
(C) Das Volk sollte bei wichtigen politischen Sachfragen mittels Volksabstimmung das letzte Wort haben. [Sovereignty]
(D) Die normalen Bürger ziehen an einem Strang. [Homogeneity]
(E) Die Unterschiede zwischen dem Volk und der sogenannten Elite sind viel größer als die Unterschiede innerhalb des Volkes. [Anti-Elitism]
(F) Das Volk und nicht die Politiker sollten die wichtigsten politischen Entscheidungen treffen. [Sovereignty]
(G) Die Politiker im Parlament müssen dem Willen des Volkes folgen. [Sovereignty]
(H) Die normalen Bürger teilen die gleichen Werte und Interessen. [Homogeneity]
RO: *(1) stimme überhaupt nicht zu; (2) stimme eher nicht zu; (3) teils/teils; (4) stimme eher zu; (5) stimme voll und ganz zu.*

Intro: *In the following is a series of opinions on politics and society that people agree or disagree with.*
QT: *Please indicate how much you agree or disagree with the following statements:*
(A) Politicians talk too much and take too little action. [Anti-Elitism]
(B) Ordinary people are of good and honest character. [Homogeneity]
(C) The people should have the final say on the most important political issues by voting on them directly in referendums. [Sovereignty]
(D) Ordinary people all pull together. [Homogeneity]
(E) The differences between the people and the so-called elite are greater than within the people. [Anti-Elitism]
(F) The people, not the politicians, should make our most important policy decisions. [Sovereignty]
(G) The politicians in Parliament need to follow the will of the people. [Sovereignty]
(H) Ordinary people share the same values and interests. [Homogeneity]
RO: *(1) strongly disagree; (2) tend to disagree, (3) neither agree nor disagree; (4) tend to agree; (5) strongly agree*.

EFFICACY [USED AS THIRD ANTI-ELITISM ITEM]
Intro*:* *Hier ist eine Reihe von häufig gehörten Meinungen über Politik und Gesellschaft.*
QT: *Geben Sie bitte an, ob sie diesen Aussagen zustimmen oder nicht:*

(A) Die Politiker kümmern sich darum, was einfache Leute denken.
[remaining items of this battery are not used]

RO: *(1) stimme überhaupt nicht zu; (2) stimme eher nicht zu; (3) teils/teils; (4) stimme eher zu; (5) stimme voll und ganz zu.*

Intro: *In the following is a series of frequently expressed opinions on politics and society.*
QT: *Please indicate how much you agree or disagree with the following statements:*
(A) Politicians care about what ordinary people think.
[remaning items of this battery are not used]
RO: *(1) strongly disagree; (2) tend to disagree; (3) neither agree nor disagree; (4) tend to agree; (5) strongly agree.*

Dataset: Castanho Silva et al. (2018) replication data set

Castanho Silva et al. (2016)

ANTI-ELITISM
QT:
- *The government is pretty much run by a few big interests looking out for themselves.*
- *Government officials use their power to try to improve people's lives.*
- *Quite a few of the people running the government are crooked.*

RO: *(1) strongly disagree; (2), (3); (4) neither agree nor disagree; (5); (6); (7) strongly agree.*

PEOPLE CENTRISM
QT:
- *Politicians should always listen closely to the problems of the people.*
- *Politicians don't have to spend time among ordinary people to do a good job.*
- *The will of the people should be the highest principle in this country's politics.*

RO: *(1) strongly disagree; (2), (3); (4) neither agree nor disagree; (5); (6); (7) strongly agree.*

MANICHEAN OUTLOOK
QT:
- *You can tell if a person is good or bad if you know their politics.*
- *The people I disagree with politically are not evil.*
- *The people I disagree with politically are just misinformed.*

RO: *(1) strongly disagree; (2), (3); (4) neither agree nor disagree; (5); (6); (7) strongly agree.*

Oliver/Rahn (2016)

ANTI-ELITISM
QT:
- *The government is pretty much run by a few big interests looking out for themselves.*
- *Government officials use their power to try to improve people's lives.*
- *Quite a few of the people running the government are crooked.*
- *It doesn't really matter who you vote for because the rich control both political parties.*

RO: *(1) strongly disagree; (2); (3); neither agree nor disagree; (4); (5) strongly agree.*

QT: *People at the top usually get there ...*
RO: *(0) because the have more talent and work harder; (1) from some unfair advantage*

MISTRUST IN EXPERTS
QT:
- *I'd rather put my trust in the wisdom of ordinary people than the opinions of experts and intellectuals*
- *When it comes to really important questions, scientific facts don't help very much.*
- *Ordinary people can really use the help of experts to understand complicated things like science and health.*

RO: *(1) strongly disagree; (2); (3) neither agree nor disagree; (4); (5) strongly agree.*

NATIONAL AFFILIATION (HOMOGENEITY)
QT: *Trust in citizens on complex political issues*
RO: (*0) It would be unwise to trust the judgments of the citizens for todays complicated issues.; (1) I generally trust the collective judgments of the citizens, even for complex political issues.*

QT: *Different or alike compared to most [nationality]*
RO: *(0) different than most [nationality]; (1) like most other [nationality]*

QT: *How important is being [nationality] to you?*
RO: *(1) not important at all; (2); (3); (4) don't know; (5); (6); (7) very important*

Akkerman et al. (2014)

ANTI-ELITISM
QT: *Please indicate how much you disagree or agree with each statement.*
- *Political differences are larger between the elite and the people than they are among the people.*
- *Elected officials talk too much and take too little action.*

RO: *(1) strongly disagree; (2); (3); (4); (5) strongly agree*

SOVEREIGNTY
QT*: Please indicate how much you disagree or agree with each statement.*
- *The politicians in the [NATIONAL] Parliament need to follow the will of the people*
- *The people and not politicians should make our most important policy decisions.*
- *I would rather be represented by a citizen than by a specialized politician.*

RO: *(1) strongly disagree; (2); (3); (4); (5) strongly agree*

MANICHEAN OUTLOOK
QT: *Please indicate how much you disagree or agree with each statement.*
- *What people call "compromise" in politics is really just selling out on one's principles.*

RO: *(1) strongly disagree; (2); (3); (4); (5) strongly agree*

Schulz et al. (2018)

ANTI-ELITISM
QT:
- *MPs in Parliament very quickly lose touch with ordinary people.*
- *The differences between ordinary people and the ruling elite are much greater than the differences between ordinary people.*
- *People like me have no influence on what the government does.*

RO: *(1) strongly disagree; (2); (3) neither agree nor disagree; (4); (5) strongly agree.*

SOVEREIGNTY
QT:

- *The people should have the final say on the most important political issues by voting on them directly in referendums.*
- *The people should be asked whenever important decisions are taken.*

RO: *(1) strongly disagree; (2); (3) neither agree nor disagree; (4); (5) strongly agree.*

HOMOGENEITY

QT:

- *Ordinary people are of good and honest character.*
- *Ordinary people all pull together.*
- *Although the British are very different from each other, when it comes down to it they all think the same.*
- *Ordinary people share the same values and interests.*

RO: *(1) strongly disagree; (2); (3) neither agree nor disagree; (4); (5) strongly agree.*

Dataset: LISS

Akkerman et al. (2014)

We make use of data of the LISS (Longitudinal Internet Studies for the Social sciences) panel administered by CentERdata (Tilburg University, The Netherlands). Survey wave: Election Survey Ukraine referendum Measurement 3, datafile mj16a.

Please note that these are the English translations of the Dutch questionnaire provided by LISS panel. These translations differ slightly from the original English questionnaire of the Akkerman et al. scale, see *Supplement 6: Handling existing scales of populist attitudes*.[17]

QT: *Now we have some questions about your opinion on various political and societal themes. Please indicate to what extent you agree with the following statements.*

(E) Politicians in the House of Representatives should heed the opinion of the people.
[Politici in de Tweede Kamer moeten zich laten leiden door de mening van het volk.]
[Sovereignty]

(F) The most important political decisions should be made by the people and not by politicians.
[De belangrijkste politieke beslissingen zouden gemaakt moeten worden door het volk en niet door politici.]
[Sovereignty]

(G) I would rather be represented by an everyday citizen than by a professional politician.
[Ik word liever vertegenwoordigd door een gewone burger dan door een beroepspoliticus.]
[Sovereignty]

(E) The political divisions are greater between the elite and everyday citizens than between citizens.
[De politieke tegenstellingen zijn groter tussen de elite en gewone burgers dan tussen burgers onderling.]
[Anti-Elitism]

(F) Politicians talk too much and do too little.
[Politici praten te veel en doen te weinig.]
[Anti-Elitism]

(G) *In politics, reaching a compromise is often another way of describing a betrayal of principles.*
[In de politiek is het sluiten van compromissen vaak een ander woord voor het verraden van je principes.]
[Manichean Outlook]

[17] For a different translation of the Dutch items into English, see Jacobs, Akkerman, and Zaslove 2018.

RO: *(1) disagree completely; (2) disagree; (3) neither agree nor disagree; (4) agree; (5) agree completely; (6) no opinion*

Dataset: National Election Survey with CSES module

Because the questionnaire is supposed to be identical across national survey, we paste the survey instrument as described in "CSES Planning Committee Module 5 Final Report" (Hobolt et al. 2016).

QT: *Do you strongly agree, somewhat agree, neither agree nor disagree, somewhat disagree, strongly disagree with the following statements?*

(A) What people call compromise in politics is really just selling out one's principles.
[Challenges to representative democracy]
(B) Most politicians do not care about the people.
[Anti-Elitism]
(C) Most politicians are trustworthy
[Anti-Elitism]
(D) Politicians are the main problem in [COUNTRY]
[Anti-Elitism]
(E) Having a strong leader in government is good for [COUNTRY] even if the leader bends the rules to get things done.
[Anti-Elitism]
(F) The people, and not politicians, should make our most important policy decisions.
[Challenges to representative democracy]
(G) Most politicians care only about the interests of the rich and powerful.
[Anti-Elitism]

QT: *Now thinking about minorities. Do you strongly agree, somewhat agree, neither agree nor disagree, somewhat disagree, strongly disagree with the following statements?*

(B) The will of the majority should always prevail even over the rights of minorities..
[Challenges to representative democracy]

QT: *How widespread do you think corruption, such as bribe taking, is amongst politicians in [COUNTRY]:*

very widespread, quite widespread, not very widespread, it hardly happens at all?
[Anti-Elitism]

Supplement 5: Choice of scales on populist attitudes

In the main text, this study focuses on three scales of populist attitudes by Akkerman, Mudde, and Zaslove (2014), Castanho Silva et al. (2019), and Schulz et al. (2018). We selected the Akkerman et al. scale due to its prominence in the literature, and we selected the Castanho Silva et al. scale and the Schulz et al. scale because the respective studies conceptualize populist attitudes as multi-dimensional, they clearly specify the dimensional structure of the scale and its items, and both scales underwent psychometric validation during scale development.

However, there are various other scales on populist attitudes available that are also worth examining. We report additional evidence on two other scales in the Shiny Web Application: the populism scales by Oliver and Rahn (2016) as well as the scale which is included in the Comparative Study of Electoral Systems (CSES) Wave 5 Core Questionnaire (Hobolt et al. 2016). We selected the Oliver/Rahn scale for its multi-dimensional structure as it complements the evidence on the two scales in the main text, which are also constituted by three discrete dimensions. In addition, we selected the CSES scale for its (anticipated) relevance in the field of public opinion research. *Supplement 6: Handling existing scales of populist attitudes* provides specific information on the operationalization of each scale.

Supplement 6: Handling existing scales of populist attitudes

This supplement discusses operationalizing existing scales of populist attitudes as multi-dimensional concepts with non-compensatory subdimensions. The following reflections may provide guidance when operationalizing populist attitudes as an attitudinal syndrome. However, at times, the scale documentation provided in the original studies was not sufficient for unambiguous data processing. Above all, the logical structure of populist attitudes and the delimitation of the concept's subdimensions was not always evident (Castanho Silva et al. 2019, 3). While we did our best to carefully weigh the various options to process the available survey items and to transparently outline our reasoning behind the decisions, the reported approaches represent our subjective interpretations. Others may arrive at different conclusions.

In this discussion, we focus on the dimensionality and properties of the following scales to measure populist attitudes:

- *Akkerman et al. scale*
- *CSES Wave 5 scale*
-
-
- *Castanho Silva et* al. scale
- *Schulz et al. scale*
- *Oliver / Rahn scale*

Akkerman et al. scale

At the moment of writing, the instrument developed by Akkerman, Mudde, and Zaslove (2014) to measure populist attitudes was the most frequently used scale of populism on the individual level. In line with the argument of the present study, Akkerman et al. explicitly specify populist attitudes as a "*set of ideas*" (Akkerman, Mudde, and Zaslove 2014, 1328), comprised of "*necessary and sufficient conditions*" (Akkerman, Mudde, and Zaslove 2014, 1326). Specifically, the authors state that "*the focus of the questions [of the scale] is on the three core features of populism: sovereignty of the people, opposition to the elite, and the Manichean division between "good" and "evil."*" (Akkerman, Mudde, and Zaslove 2014, 1330f).[18] In other words, the authors conceptualize populist

[18] See the following similar quote on the dimensionality of the Akkerman et al. scale: "*The survey questions are designed to capture the full ideology of populism and its conception of democracy, in particular the will of the people (their sovereignty) and the distinction between the people and the elite. The Manichean nature of the distinction*

attitudes as a multi-dimensional concept, constituted by three non-compensatory concept components: Sovereignty of the people, anti-elitism, Manichean outlook.

Although the conceptual underpinnings of the Akkerman et al. scale thus mirror the present study's conceptualization of individual-level populism as an attitudinal syndrome, operationalizing the Akkerman et al. scale along the strategies discussed in the present study is not straightforward. These difficulties arise because the scale dimensionality is clear on the conceptual level, but it is less clear how to implement the multidimensionality operationally.

In the original study, Akkerman et al. „*perform[ed] a principal component analysis (PCA) to investigate whether it is possible to identify a populist dimension* [underscoring not in the original text]" (Akkerman, Mudde, and Zaslove 2014, 1326). In this vein, the authors then proceed to extract *one* dimension of populist attitudes from a longer battery of attitudinal items. It was in response to the scale's seeming uni-dimensionality that other scholars have developed new measures of populist attitudes with an explicitly multi-dimensional structure (e.g., Schulz et al. 2018, 318). As discussed in greater detail below, the concurrence of the multi-dimensional conceptual specification that underlies the scale and its usage as a one-dimensional measure (e.g., van Hauwaert and van Kessel 2018) is due to the fact that some items of the Akkerman et al. scale tap into multiple concept components at once. However, the logical structure of the scale is not always unambiguous (Castanho Silva et al. 2019, 3). Therefore, more than one approach to operationalize the scale is conceivable. Because each approach has distinct advantages and disadvantages, we discuss two strategies to operationalize the Akkerman et al. scale as an attitudinal syndrome with multiple concept components.

Scales may already account for the non-compensatory relationship between concept components at the stage of measurement by combining all concept components in each measure. More than is the case for the remaining scales examined in this study, the indicators of the Akkerman et al. scale can be understood as intended to measure not only one concept subdimension but as incorporating multiple concept components into each survey items (Akkerman, Mudde, and Zaslove 2014, 1332; Castanho Silva et al. 2019, 2). In this vein, some items of the Akkerman et al. scale are worded in a way so that respondents only agree with a statement if respondents concurrently agree with the tenets of multiple subdimensions (e.g., "*The people, and not*

between the people and the elites is also a feature of our survey questions" (Akkerman, Mudde, and Zaslove 2014, 1331).

politicians, should make our most important policy decisions" which taps into sovereignty orientations and anti-elitist orientations, see Table S6-1 for an overview of the scale items). Indeed, if all items were measuring the intersection of all concept components, then it would be unnecessary to account for the dimensions' non-compensatory relationship in the aggregation procedure. Under these circumstances, aggregating the items with simple averages - as if they reflected one dimension - would be feasible; the composite score would still assess populist attitudes as an attitudinal syndrome because the necessary conditions were already accounted for at the stage of measurement (see *Supplement* 14: Assessing non-compensatory multi-dimensional concepts already at the stage of measurement (Akkerman et al. scale) for a discussion of other vices and virtues of this approach).

In the case of the Akkerman et al. scale, however, these conditions are not met. Some indicators tap into two components simultaneously, but it is unclear whether all items are inherently multi-dimensional or whether they are linked more or less clearly to one subdimension (e.g., "*What people call "compromise" in politics is really just selling out on one's principles*" may reflect Manichean Outlook only). More importantly, not all items appear to measure all concept components simultaneously (e.g., "*Elected officials talk too much and take too little action*" seems not to reflect orientations towards popular sovereignty). Hence, because one cannot take for granted that the multi-dimensionality of the entire concept as already incorporated in each survey item, we need to examine the dimensionality of the scale more closely before combining the scale subdimensions according to non-compensatory operationalization strategies.

Table S6-1. **Items and subdimensions of Akkerman et al. populism scale**

Code	*Item*	*Strategy 1 Subdimension*	*Strategy 2 Subdimension*
POP 1	The politicians in the [NATIONAL] Parliament need to follow the will of the people	Anti-elitism and sovereignty	Sovereignty
POP 2	The people, and not politicians, should make our most important policy decisions	Anti-elitism and sovereignty	Sovereignty
POP 3	The political differences between the elite and the people are larger than the differences among the people	Anti-elitism and sovereignty	Anti-elitism
POP 4	I would rather be represented by a citizen than by a specialized politician	Anti-elitism and sovereignty	Sovereignty
POP 5	Elected officials talk too much and take too little action	Anti-elitism and Manichean outlook	Anti-elitism

POP 7	What people call "compromise" in politics is really just selling out on one's principles	Anti-elitism and Manichean outlook	Manichean outlook

Note: To ease comparison with the original study, the numbering of the items follows that of the original authors (Akkerman, Mudde, and Zaslove 2014, 1333). Strategy 2 was used in this study for the empirical analysis of the scale.

Strategy 1 generally acknowledges the inherent multidimensionality of the scale items and associates individual items with multiple concept components (see Table S6-1). Strategy 1 thus requires to examine how many and which subdimensions an item captures. Strategy 2, on the other hand, approximates the more classical approach to multi-dimensional concepts by understanding one survey item as reflective of one concept subdimension. Strategy 2 thus requires to identify the concept subdimension that is dominant in an item even when the item may tap into multiple subdimensions.

1) Strategy 1: Multi-dimensional items

According to the authors, the first four items [POP1-POP4] "reflect ideas about representative government, reflecting the ideas that there is a division between the people and the politicians (the elite) and that politicians do not represent the true will of the people" (Akkerman, Mudde, and Zaslove 2014, 1332). Because it is not always clear that those items also tap into the Manichean dimension (e.g., "*The politicians in the [NATIONAL] Parliament need to follow the will of the people*"), these items are understood as tapping into both the sovereignty and the anti-elitist dimensions. More specifically, one may interpret this battery of items as capturing the conjunction of anti-elitist and sovereignty-supporting attitudes.

According to the original authors, "*the Manichean dimension, that is, the tension between "good" and "evil," is captured in the questions POP5 through POP7*" (Akkerman, Mudde, and Zaslove 2014, 1334). While this sentence implies that POP5 ("*Elected officials talk too much and take too little action*") and POP7 ("*What people call "compromise" in politics is really just selling out on one's principles*", POP6 is not included in the final scale) measure the Manichean subdimension, a Manichean outlook in the context of populism is inherently linked with anti-elitist orientations because the evil is impersonated by the corrupted elite in the populist worldview (see also Akkerman, Mudde, and Zaslove 2014, 1331). Reflecting this link between anti-elitism and a Manichean outlook, the authors posit that "*statements POP5, POP6* [not included in the final scale], *and POP7 are intended to emphasize that the distinction between the people and the elite is a battle between good and evil*" (Akkerman, Mudde, and Zaslove 2014, 1331). Hence, the items

POP5 and POP7 may be understood as tapping into two subdimensions: anti-elitism and Manicheanism.

Having discussed all indicators of the Akkerman et al. populism scale, Table S6-1 documents our interpretation based on the authors' original description of the subdimensions that each indicator taps into. Based on this overview, it is possible to apply the Goertz or Sartori operationalization strategy to derive a composite populism score using the items of the Akkerman et al. scale. Specifically, using weighted or unweighted summary scores, one would first aggregate POP1 through POP4 into a composite index to reflect the conjunct acceptance of anti-elitist and sovereignty-supporting views. The second index of POP5 and POP7 would reflect a person's conjunct acceptance of anti-elitist and of Manichean orientations. Both indexes would then be combined using the Goertz or the Sartori operationalization strategy to derive a composite index of populist attitudes.

2) Strategy 2: One-dimensional items

While some statements of the original authors can be interpreted as suggesting that all scale items are inherently multi-dimensional, a different interpretation of the original study and, thus, a different strategy for operationalization is also conceivable. Moreover, inherently multi-dimensional items are not widespread in public opinion research where each item is usually intended to measure one substantive orientation (see Supplement 14: Assessing non-compensatory multi-dimensional concepts already at the stage of measurement (Akkerman et al. scale) for a critical discussion). Another option to operationalize the Akkerman et al. scale would thus be to follow the standard practices in public opinion research by grouping items by the subdimensions that each item is supposed to reflect, assuming that each indicator is reflective of one subdimension. Applying this approach to the Akkerman et al. scale thus requires to identify the concept subdimension that is dominant in an item even when the item may tap into multiple subdimensions.

As POP1, POP2, POP4 all relate to the distribution of political power –more specifically to the tenet of transferring power from the elites to the people– we understand these items as measuring one's support for popular sovereignty as the items' most salient facet and thus grouped these items to the respective sovereignty dimension. POP3 ("*The political differences between the elite and the people are larger than the differences among the people*"), in contrast, assesses differences between the elite and the people and refers to popular sovereignty only indirectly, and we thus grouped this item in the anti-elitism dimension. We also grouped POP5 ("Elected officials talk too much and take too little action") in the anti-elitist dimension in line with the Schulz et al.

populism scale which also uses the item in indicator of anti-elitist orientation. Finally, as POP7 ("*What people call "compromise" in politics is really just selling out on one's principles*") refers to elites only indirectly but assesses one's general outlook on political issues, we interpreted it as mainly reflecting the Manichean dimension.

Based on this strategy, one can aggregate the items of the Akkerman et al. scale into three distinctive subdimensions (for instance, by calculating average scores). Then, these sub-dimensions may be aggregated into individual-level populism scores according to the Goertz or Sartori operationalization strategy (see main text and *Supplement* 2: Decisions to be made for the aggregation of multi-dimensional constructs).

3) Conclusion

Both of the strategies for operationalizing the Akkerman et al. scale presented above are feasible options for using the scale items in a way that ensures to assign high populism to an individual only if the individual agrees with all subdimensions of the concept. In our empirical analysis, we opted to use Strategy 2 as it is more in line with the operationalization procedures of the other populism scales discussed in this article.

Notably, both of these strategies differ from the operationalization procedure performed by the original authors, which, mathematically, would amount to the Bollen procedure. However, even though we have argued that the Akkerman et al. scale does not fully incorporate all of the concepts' necessary conditions already on the measurement stage, various scale items can be understood as tapping into the conjunction of at least two concept components. Because the scale accounts for the non-compensatory relationship between concept components at least to a certain extent already at the measurement stage, whether a non-compensatory or a compensatory operationalization procedure is employed is less consequential in this case. In other words, whether the Bollen or the Goertz operationalization strategy is performed should make less of a difference compared to scales in which all items are clearly one-dimensional (see Shiny Web Application to compare scales in various samples). Indeed, our findings demonstrate that the Akkerman et al. scale is fairly robust to choices in operationalization strategies.

CSES Wave 5 scale

Operationalizing the CSES populism scale is not straightforward. As part of the module "People, Politicians and the Politics of Populism", CSES wave 5 provides various survey items that are more or less directly associated with populism at the mass level (Hobolt et al. 2016).

According to the documentation, the CSES Wave 5 Module measures three "core themes" (Hobolt et al. 2016, 5): attitudes toward political elites, attitudes toward representative democracy and majority rule, attitudes toward out-groups. When intending to use the CSES Module 5 for research on populist attitudes, the main question at the conceptual level is whether all of these "core themes" constitute essential elements of populist attitudes. At the indicator level then the question is which of the survey questions in CSES Module 5 tap into the concept of populism and whether some indicators measure related but different concepts.

The CSES documentation suggests that populism at the mass level is to be understood as a multi-dimensional concept (Hobolt et al. 2016, 4). However, it does not explicitly specify which items and concepts denote essential concept characteristics of populist attitudes and which denote collateral concepts. Therefore, we engaged in a close reading of the Planning Committee's Report on Module 5 to derive a coherent operationalization of populist attitudes that is consistent with the specification of the concept, as suggested by the original authors in the report (Hobolt et al. 2016). As a general rule, we decided only to include items and concepts which unequivocally constitute essential elements of populism at the mass level. Consequently, according to our interpretation, the CSES Module 5 item battery is comprised of two types of measures: those that tap into essential core elements of populist attitudes and those that indicate certain varieties of populism which do not necessarily capture core elements of the concept (e.g. authoritarian orientations). Based on these general considerations, we discuss our subjective interpretation of how CSES Module 5 data may be used to operationalize populist attitudes as a multi-dimensional concept.

According to the CSES documentation, "*the belief that political elites and the people have contrasting and incompatible interests is at the heart of populism*" (Hobolt et al. 2016, 4). Later, the authors re-iterate that „*the core aspect of populism is the notion of a clear distinction between the (good) people and the (evil) elite (Pappas 2012; Woods 2014). The antagonism between elites and the people is at the heart of populism (Mudde and Kaltwasser 2014)*" (Hobolt et al. 2016, 5). Therefore, the concept specification suggests considering anti-elitist orientations as the essential component of populist attitudes that the authors intend to measure with the CSES scale. Consequently, we included all items that were suggested by the authors to measure anti-elitist orientations: Q4 (b, c, d) and Q7. (The battery contains yet another anti-elitism item –Q4G – which

we have not included in the scale as it only taps into a specific kind of populism according to the official documentation.[19])

According to the CSES documentation, "*[p]opulism […] encompasses a rejection of pluralism and opposition to the protection of minorities*" (Hobolt et al. 2016, 4). In other words, the CSES planning committee conceptualizes attitudes toward principles and procedures of governing as an essential element of populist orientations. This proposition is shared by many, albeit not all scholars of populism (e.g., Pappas 2016; Plattner 2010; Urbinati 2014). Thus, the concept specification suggests that attitudes toward democracy and pluralism can be considered as an essential component of populist attitudes that is to be measured with the CSES scale. Yet, it is less clear which items to include for measuring this concept component.

The CSES report does not specify the dimensionality of the concept dimensions, nor does it clearly state the intended usage of the various items that are listed in the respective section on democracy-related attitudes. Items Q4e and Q4f are introduced as "*frequently voiced alternatives to the 'corrupted' representative system*" (Hobolt et al. 2016, 7). Considering that they are 'frequently voiced' implies that they may often go along with populism but do not necessarily constitute essential elements of populist attitudes. It is important to remember that populist attitudes are a form of thin ideology which can be linked to various host ideologies (Mudde 2017). One of these host ideologies can be but does not have to be authoritarianism. Therefore, we decided not to consider Q4e ("*Having a strong leader in government is good for [COUNTRY] even if the leader bends the rules to get things done*") as measuring an essential aspect of populism on the mass level as Q4e apparently taps into support for authoritarian regime preferences. Q4f ("*The people, and not politicians, should make our most important policy decisions*"), in contrast, taps into support for popular sovereignty, which authors of other populism scales also recognize as an essential component of populism at the mass level. Hence, we retained that item. We also kept Q4a ("*What people call compromise in politics is really just selling out one's principles*") which is an established item for measuring populism at the mass level and has been adopted from the Akkerman et al. scale. Q5b ("*The will of the majority should always prevail even over the rights of minorities*") was also retained as it is intended to measure attitudes towards the democratic process with respect

[19] According to the authors the item Q4G ("Most politicians care only about the interests of the rich and powerful.") would measure left-wing populism: "Also, leftwing populists in particular often portray the elite as representative of the rich, the economically advantaged, and large financial corporations in opposition to the ordinary economically and socially disadvantaged people (captured in items Q4g)." Hobolt et al. (2016, 7).

to minority rights which the authors have established as a constituent element of populism (see above and: "*populism tends to give priority to majority rule as a means to reach decisions, therefore leaving minority rights in a secondary place. This is captured in question Q5b*" (Hobolt et al. 2016, 7).

Altogether, inspecting the available measures and their respective description in the CSES documentation leaves us with three indicators to measure the component on "Challenges to representative democracy." Arguably, this component is internally less consistent than the anti-elitist subdimension as the items for that component tap into notions of anti-pluralism and sovereignty. Anti-pluralism and support for popular sovereignty are substantively linked but still represent distinct aspects of democracy-related attitudes. Nonetheless, due to their theoretical nexus and because the CSES documentation subsumes these items under the common label of "Challenges to representative democracy", we also combine these items in one concept component for the purpose of this paper.

Finally, the CSES Module 5 offers additional survey items related to out-groups. While these survey questions help determine attitudes to which host ideology populist orientations are linked in specific populations, these orientations represent potential kinds but not essential elements of populism. Therefore, we did not include any of these here.

Table S6-2 provides an overview of the scale items and the concept subdimensions that they correspond to according to the option discussed above. After aggregating the items within each subdimension (for instance, by calculating average scores), all concept components may be aggregated into individual-level populism scores according to the Goertz or Sartori operationalization strategy (see main text and *Supplement* 2: Decisions to be made for the aggregation of multi-dimensional constructs).

Table S6-2. **Items and subdimensions, CSES populism scale**

Code	*Item*	*Subdimension*
Q4b	Most politicians do not care about the people	Anti-elitism
Q4c	Most politicians are trustworthy*	Anti-elitism
Q4d	Politicians are the main problem in [COUNTRY]	Anti-elitism
Q7	How widespread do you think corruption, such as bribe-taking, is amongst politicians in [COUNTRY]: very widespread, quite widespread, not very widespread, it hardly happens at all?	Anti-elitism
Q4a	What people call compromise in politics is really just selling out one's principles.	Challenges to representative democracy

Q4f	The people, and not politicians, should make our most important policy decisions.	Challenges to representative democracy
Q5b	The will of the majority should always prevail even over the rights of minorities.	Challenges to representative democracy
CSES Module 5 Survey items which are not included in the populism measure		
Q4g	Most politicians care only about the interests of the rich and powerful.	Left-wing Anti-elitism
Q4e	Having a strong leader in government is good for [COUNTRY] even if the leader bends the rules to get things done.	Authoritarian Regime Preferences

Note: Asterisks indicate reverse coded items.

Castanho Silva et al. scale

Operationalizing the scale of populist attitudes by Castanho Silva et al. is straightforward. The original study (Castanho Silva et al. 2018) clearly specifies scale dimensionality and the association between item and subdimensions (Table S6-3). After aggregating the items within each subdimension (for instance, by calculating average scores), all concept components may be aggregated into individual-level populism scores according to the Goertz or Sartori operationalization strategy (see main text and *Supplement* 2: Decisions to be made for the aggregation of multi-dimensional constructs).

Table S6-3: **Items and subdimensions, Silva et al. populism scale**

Code	*Item*	*Subdimension*
Ppl 1	Politicians should always listen closely to the problems of the people.	People-centrism
Ppl 2	Politicians don't have to spend time among ordinary people to do a good job.*	People-centrism
Ppl 3	The will of the people should be the highest principle in this country's politics	People-centrism
Ant 1	The government is pretty much run by a few big interests looking out for themselves.	Anti-elitism
Ant 2	Government officials use their power to try to improve people's lives.*	Anti-elitism
Ant 3	Quite a few of the people running the government are crooked.	Anti-elitism
Man1	You can tell if a person is good or bad if you know their politics.	Manichean outlook
Man 2	The people I disagree with politically are not evil.*	Manichean outlook
Man 3	The people I disagree with politically are just misinformed.	Manichean outlook

Note: Asterisks indicate reverse coded items.

Schulz et al. scale

Operationalizing the scale of populist attitudes by Schulz et al. is straightforward. The original study (Schulz et al. 2018) clearly specifies scale dimensionality and the association between item and subdimensions (Table S6-4). After aggregating the items within each subdimension (for instance, by calculating average scores), all concept components may be aggregated into individual-level populism scores according to the Goertz or Sartori

operationalization strategy (see main text and *Supplement* 2: Decisions to be made for the aggregation of multi-dimensional constructs).

Table S6-4: **Items and subdimensions, Schulz et al. populism scale**

Code	*Item*	*Subdimension*
Ant 1	MPs in Parliament very quickly lose touch with ordinary people.	Anti-elitism
Ant 2	The differences between ordinary people and the ruling elite are much greater than the differences between ordinary people.	Anti-elitism
Ant 3	People like me have no influence on what the government does.	Anti-elitism
Ant 4	Politicians talk too much and take too little action.	Anti-elitism
Sov 1	The people should have the final say on the most important political issues by voting on them directly in referendums.	Sovereignty
Sov 2	The people should be asked whenever important decisions are taken.	Sovereignty
Sov 3	The people, not the politicians, should make our most important policy decisions.	Sovereignty
Sov 4	The politicians in Parliament need to follow the will of the people.	Sovereignty
Hom 1	Ordinary people all pull together.	Homogeneity
Hom 2	Ordinary people are of good and honest character.	Homogeneity
Hom 3	Ordinary people share the same values and interests.	Homogeneity
Hom 4	Although the Swiss are very different from each other, when it comes down to it they all think the same.	Homogeneity

Note: This table shows the original scale items. The analysis in the main text uses an adapted scale with slightly different wording. See Supplement 10: Various operationalizations of the Sartori concept structure.

Oliver / Rahn scale

Operationalizing the scale of populist attitudes by Oliver/Rahn is straightforward. The original study (Oliver and Rahn 2016) clearly specifies scale dimensionality and the association between item and subdimensions (Table S6-5).[20] After aggregating the items within each subdimension (for instance, by calculating average scores), all concept components may be aggregated into individual-level populism scores according to the Goertz or Sartori operationalization strategy (see main text and *Supplement* 2: Decisions to be made for the aggregation of multi-dimensional constructs).

Table S6-5. **Items and subdimensions, Oliver/Rahn populism scale**

Code	*Item*	*Subdimension*
Ppl 1	Trust in citizens on complex political issues (dichotomous).	National Affiliation
Ppl 2	Similarity with other fellow citizens (dichotomous)	National Affiliation
Ppl 3	How important is being [nationality] to you?	National Affiliation
Ant 1	The government is pretty much run by a few big interests looking out for themselves	Anti-elitism

[20] The Oliver/Rahn scale contains a discrete subdimension on national identity whereas we did not include items related to national identity as essential items of the CSES scale. We follow a different strategy due to the primacy of the original authors' concept specification over our interpretation of a scale. The original author' concept specification is clearly documented in the case of the Oliver/Rahn scale but required our interpretation in the case of the CSES scale.

Ant 2	Government officials use their power to try to improve people's lives.	Anti-elitism
Ant 3	Quite a few of the people running the government are crooked.	Anti-elitism
Ant 4	It doesn't really matter who you vote for because the rich control both political parties	Anti-elitism
Mis 1	I'd rather put my trust in the wisdom of ordinary people than the opinions of experts and intellectuals	Mistrust in Experts
Mis 2	I'd rather put my trust in the wisdom of ordinary people than the opinions of experts and intellectuals	Mistrust in Experts
Mis 3	Ordinary people can really use the help of experts to understand complicated things like science and health.	Mistrust in Experts

Supplement 7: Descriptive statistics

Table S7-1 shows the distribution of socio-demographic variables in the GLES Campaign Panel, as it was used in the main analysis (excluding respondents with missing values on the populism variables from survey wave 5). Tables S7-2 through Table S7-5 shows the distribution of socio-demographic variables in the Castanho Silva et al. replication data set, separated by countries.

Table S7-1. **Sample composition in German Longitudinal Election Study, Campaign Panel 2017**

	Total (N = 13998)
Age	
Mean (SD)	48.73 (14.7)
Gender	
Male	6844 (48.9%)
Female	7154 (51.1%)
Education	
School student ("Schüler")	60 (0.4%)
Left school without a certificate ("ohne Abschluss")	96 (0.7%)
Certificate of Secondary Education ("Hauptschulabschluss")	2362 (16.9%)
Intermediate school-leaving certificate ("Mittlere Reife")	3905 (27.9%)
Advanced technical college certificate ("Fachholschulreife")	1009 (7.2%)
Advanced school-leaving certificate("Abitur")	3186 (22.8%)

Notes: **Education**: current/highest completed school degree. Surveyed at wave 5.

Table S7-2. **Sample composition in Castanho Silva et al. (2018) replication data set**

	U.S. (N = 505)	*Brazil* (N = 281)	*France* (N = 274)	*Greece* (N = 275)	*Ireland* (N = 269)	*Italy* (N = 219)	*Mexico* (N = 221)	*Spain* (N = 280)	*UK* (N = 186)
Age									
Mean (SD)	35.0 (10.6)	29.5 (8.5)	29.7 (8.4)	34.0 (8.7)	37.8 (11.3)	35.4 (11.0)	33.1 (9.3)	32.2 (9.1)	34.0 (10.6)
Education									
Mean (SD)	4.1 (1.3)	14.4 (4.2)	16.3 (4.3)	17.0 (5.1)	15.0 (3.9)	14.3 (5.2)	14.1 (4.9)	14.5 (4.9)	14.4 (5.2)
Income									
Mean (SD)	5.8 (3.1)	4.0 (1.5)	6.3 (2.9)	5.7 (2.6)	4.8 (2.4)	4.6 (2.5)	4.6 (2.7)	3.6 (2.5)	4.4 (2.5)
Ideology									
Mean (SD)	4.2 (2.3)	5.5 (2.1)	4.9 (1.8)	4.4 (2.2)	5.1 (2.3)	5.1 (1.9)	5.5 (1.9)	5.2 (1.9)	5.4 (1.8)
Gender									
Male	223 (44.7%)	219 (81.1%)	206 (76.6%)	195 (73.6%)	142 (53.2%)	125 (59.5%)	152 (72.4%)	185 (73.7%)	106 (59.2%)
Female	276 (55.3%)	51 (18.9%)	63 (23.4%)	70 (26.4%)	125 (46.8%)	85 (40.5%)	58 (27.6%)	66 (26.3%)	73 (40.8%)

Notes: **Age**: mean age; **Education**: for American sample, mean of the highest degree achieved. For the others, mean number of years completed of formal education; **Income**: mean income decile; **Ideology**: mean left-right self-placement on a 1-9 scale, where 1 is left.

Table S7-3. **Schulz et al. populism scale in German Longitudinal Election Study, Campaign Panel 2017**

	Total (N = 13998)
Populism (Goertz)	
Mean (SD)	-0.54 (0.73)
Populism (Bollen)	
Mean (SD)	-0.01 (0.65)
Anti-Elitism	
Mean (SD)	-0.01 (0.79)
Manichean Outlook	
Mean (SD)	0.00 (0.84)
People-Centrism	
Mean (SD)	0.00 (0.85)
Populism (Sartori)	
0	12966 (93.0%)
1	969 (7.0%)

Table S7-4. **Populism scales (Castanho Silva et al. 2016) in Castanho Silva et al. (2018) replication data set**

	U.S. (N = 505)	*Brazil* (N = 281)	*France* (N = 274)	*Greece* (N = 275)	*Ireland* (N = 269)	*Italy* (N = 219)	*Mexico* (N = 221)	*Spain* (N = 280)	*UK* (N = 186)
Populism (Goertz)									
Mean (SD)	0.36 (0.16)	0.37 (0.14)	0.37 (0.14)	0.35 (0.15)	0.36 (0.16)	0.36 (0.15)	0.35 (0.16)	0.37 (0.15)	0.37 (0.15)
Populism (Bollen)									
Mean (SD)	0.53 (0.12)	0.53 (0.10)	0.53 (0.10)	0.53 (0.10)	0.53 (0.11)	0.53 (0.11)	0.53 (0.10)	0.53 (0.10)	0.53 (0.11)
Anti-Elitism									
Mean (SD)	0.61 (0.19)	0.61 (0.17)	0.61 (0.17)	0.61 (0.18)	0.61 (0.19)	0.61 (0.17)	0.61 (0.18)	0.61 (0.16)	0.61 (0.16)
Manichean Outlook									
Mean (SD)	0.40 (0.19)	0.40 (0.16)	0.40 (0.15)	0.40 (0.19)	0.40 (0.17)	0.40 (0.18)	0.40 (0.19)	0.40 (0.17)	0.40 (0.17)
People-Centrism									
Mean (SD)	0.59 (0.17)	0.59 (0.15)	0.59 (0.15)	0.59 (0.15)	0.59 (0.15)	0.59 (0.16)	0.59 (0.15)	0.59 (0.15)	0.59 (0.17)
Populism (Sartori)									
0	471 (93.3%)	271 (96.4%)	264 (96.4%)	266 (96.7%)	245 (91.1%)	207 (94.5%)	216 (97.7%)	270 (96.4%)	181 (97.3%)
1	34 (6.7%)	10 (3.6%)	10 (3.6%)	9 (3.3%)	24 (8.9%)	12 (5.5%)	5 (2.3%)	10 (3.6%)	5 (2.7%)

Table S7-5. **Populism scales (Oliver/Rahn 2016) in Castanho Silva et al. (2018) replication data set**

	U.S. (N = 505)	*Brazil* (N = 281)	*France* (N = 274)	*Greece* (N = 275)	*Ireland* (N = 269)	*Italy* (N = 219)	*Mexico* (N = 221)	*Spain* (N = 280)	*UK* (N = 186)
Goertz (Oliver/Rahn)									
Mean (SD)	0.41 (0.15)	0.43 (0.14)	0.42 (0.14)	0.42 (0.14)	0.43 (0.15)	0.42 (0.14)	0.42 (0.14)	0.42 (0.15)	0.42 (0.15)
Bollen (Oliver/Rahn)									
Mean (SD)	0.58 (0.11)	0.58 (0.11)	0.58 (0.11)	0.58 (0.10)	0.58 (0.12)	0.58 (0.11)	0.58 (0.11)	0.58 (0.11)	0.58 (0.11)
Anti-Elitism									
Mean (SD)	0.57 (0.20)	0.58 (0.17)	0.57 (0.18)	0.57 (0.18)	0.57 (0.20)	0.57 (0.19)	0.57 (0.19)	0.57 (0.19)	0.57 (0.18)
National Affiliation (Homogeneity)									
Mean (SD)	0.48 (0.18)	0.48 (0.16)	0.48 (0.16)	0.48 (0.15)	0.48 (0.15)	0.48 (0.17)	0.48 (0.16)	0.48 (0.16)	0.48 (0.17)
Mistrust in Experts									
Mean (SD)	0.69 (0.15)	0.69 (0.18)	0.69 (0.17)	0.69 (0.17)	0.69 (0.17)	0.69 (0.17)	0.69 (0.16)	0.69 (0.16)	0.69 (0.16)
Sartori (Oliver/Rahn)									
0	250 (98.0%)	264 (96.4%)	258 (95.6%)	268 (98.5%)	258 (96.6%)	211 (96.8%)	211 (97.2%)	258 (98.5%)	179 (96.8%)
1	5 (2.0%)	10 (3.6%)	12 (4.4%)	4 (1.5%)	9 (3.4%)	7 (3.2%)	6 (2.8%)	4 (1.5%)	6 (3.2%)

Table S7-6. **Populism scales (Schulz 2018) in Castanho Silva et al. (2018) replication data set**

	U.S. (N = 505)	*Brazil* (N = 281)	*France* (N = 274)	*Greece* (N = 275)	*Ireland* (N = 269)	*Italy* (N = 219)	*Mexico* (N = 221)	*Spain* (N = 280)	*UK* (N = 186)
Goertz (Schulz)									
Mean (SD)	0.48 (0.16)	0.49 (0.17)	0.49 (0.17)	0.49 (0.16)	0.49 (0.17)	0.49 (0.13)	0.49 (0.16)	0.49 (0.17)	0.50 (0.15)
Bollen (Schulz)									
Mean (SD)	0.62 (0.12)	0.62 (0.13)	0.62 (0.13)	0.62 (0.13)	0.62 (0.14)	0.62 (0.12)	0.62 (0.13)	0.62 (0.14)	0.62 (0.13)
Anti-Elitism									
Mean (SD)	0.65 (0.19)	0.65 (0.18)	0.65 (0.17)	0.65 (0.17)	0.65 (0.19)	0.65 (0.18)	0.65 (0.18)	0.65 (0.18)	0.65 (0.18)
Homogeneity									
Mean (SD)	0.68 (0.18)	0.68 (0.18)	0.68 (0.18)	0.68 (0.19)	0.68 (0.18)	0.68 (0.17)	0.68 (0.18)	0.68 (0.18)	0.68 (0.18)
Sovereignty									
Mean (SD)	0.54 (0.17)	0.54 (0.18)	0.54 (0.17)	0.54 (0.17)	0.54 (0.18)	0.54 (0.15)	0.54 (0.16)	0.54 (0.18)	0.54 (0.15)
Sartori (Schulz)									
0	239 (93.7%)	249 (91.5%)	240 (89.2%)	254 (93.4%)	246 (92.1%)	203 (93.1%)	207 (95.0%)	238 (91.2%)	168 (91.3%)
1	16 (6.3%)	23 (8.5%)	29 (10.8%)	18 (6.6%)	21 (7.9%)	15 (6.9%)	11 (5.0%)	23 (8.8%)	16 (8.7%)

Supplement 8: Non-probability and probability-based survey data

The analyses reported in the main text do not make use of probability-based samples of the general population. Instead, the analysis relies on non-probability samples, all of which were collected online. Specifically, the study uses the Campaign Panel of the German Longitudinal Election Study (GLES), which is a heterogeneous sample of respondents who were recruited and surveyed online. Even though GLES employs socio-demographic quotas to create a sample that closely resembles the German voting age population, the sample cannot be considered representative of the German population. Above all, the sample is restricted to individuals with Internet access. In addition, the sample exhibits moderate biases with respect to social characteristics such as formal education and political characteristics such as voting behavior (Mader and Schoen 2019).

The second data source underlying the analyses in the main text, the Castanho Silva et al. replication dataset, is also subject to sampling biases. The samples contained in the Castanho Silva et al. dataset were drawn from two crowd worker services, Crowdflower and MTurk. It is well known that the composition of respondents on these platforms do not resemble the general population and that the political attitudes measured on the platforms do not mirror the attitudes of the general population (Berinsky, Huber, and Lenz 2012).

Against the backdrop of these sampling errors that may affect the distribution of political attitudes, we are cautious not to draw inferences from these samples about the distribution of populist attitudes in the general population. Instead, for inferences about the distribution of single attitudes, we make use of probability-based samples in which the measured distribution of attitudes usually resembles that of the general population more closely (Berinsky, Huber, and Lenz 2012). In particular, in the Shiny Web Application where we use the Sartori operationalization strategy to report the share of populist citizens in various populations we only calculate the estimates using probability-based samples.

However, even though survey-methodological research shows that the distribution of attitudes in non-probability-based samples may differ compared with probability-samples or the general population, previous research has also shown that treatment effects (e.g., Mullinix et al. 2015) and the functional relationship between variables (e.g., Bieber and Bytzek 2012) are often not biased or not strongly biased in non-probability samples. Hence, if we understand the primary

estimand of the main text as the difference between the Bollen and Goertz populism scores, then this study's main interest also concerns functional relationships between variables, providing some confidence regarding the credibility of the estimates.

Even though there is thus more reason to be confident about the generalizability of the main estimand, we have conducted additional robustness tests to assess this claim. Specifically, we conducted analyses to compare the correlation between Bollen and Goertz populism scores between probability-based and non-probability-based samples. Unfortunately, the only scale for which we have access to both probability-based and non-probability-based samples is the Akkerman et al. populism scale. This is unfortunate because the Akkerman et al. scale exhibits strong robustness to operationalization strategies, that is a strong association between Goertz and Bollen composite scores. These strong correlations can give rise to floor effects, which may lead to an underestimation of the sensitivity of the scale to sample characteristics such as non-probability and probability-based samples. With this caveat in mind, there are no meaningful differences in the internal structure of the Akkerman et al. scale between probability-based samples (Figure S8-1, Figure S8-2) and non-probability samples (Figure S8-3, Figure S8-4). While not definitive evidence, these findings can be read as supporting the proposition that disparities between probability-based and non-probability samples may occur under very specific conditions that are not likely to be met in the context of this research question.

Figure S8-1. **Internal Structure of Akkerman et al. populism scale (Germany, GLES Face-to-face Cross-Section Survey)**

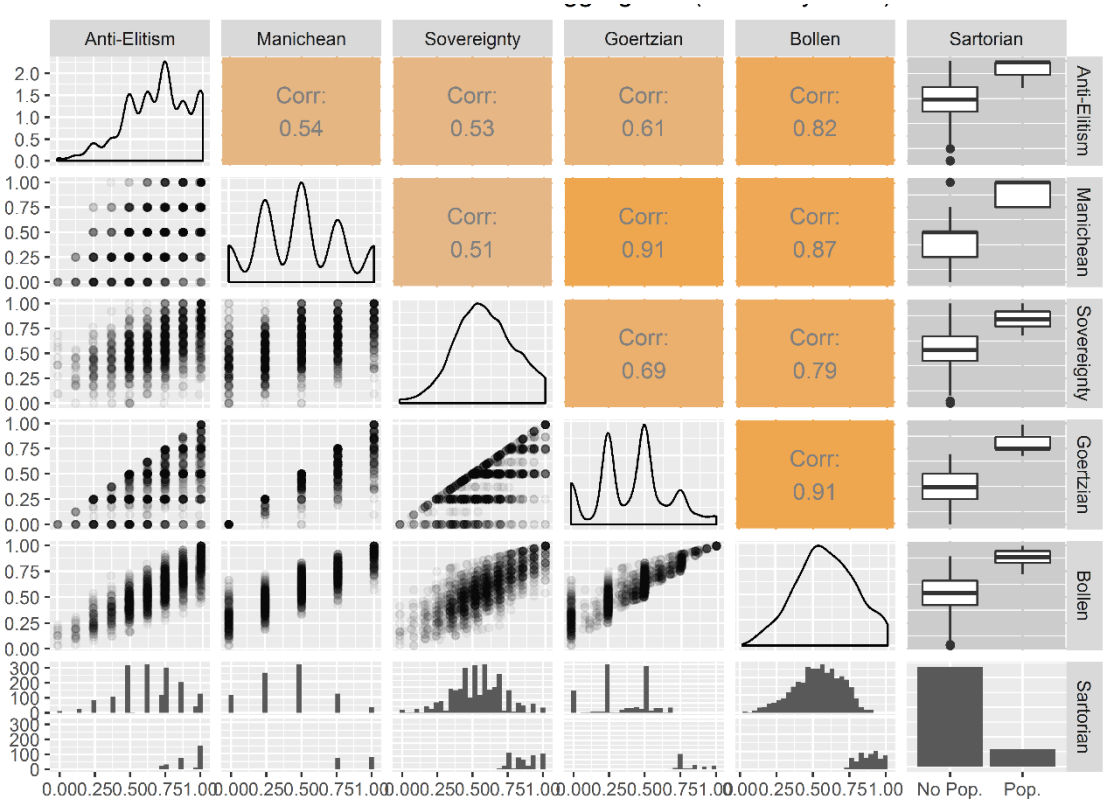


Figure S8-2. **Internal Structure of Akkerman et al. populism scale (Netherlands, LISS probability-based panel)**

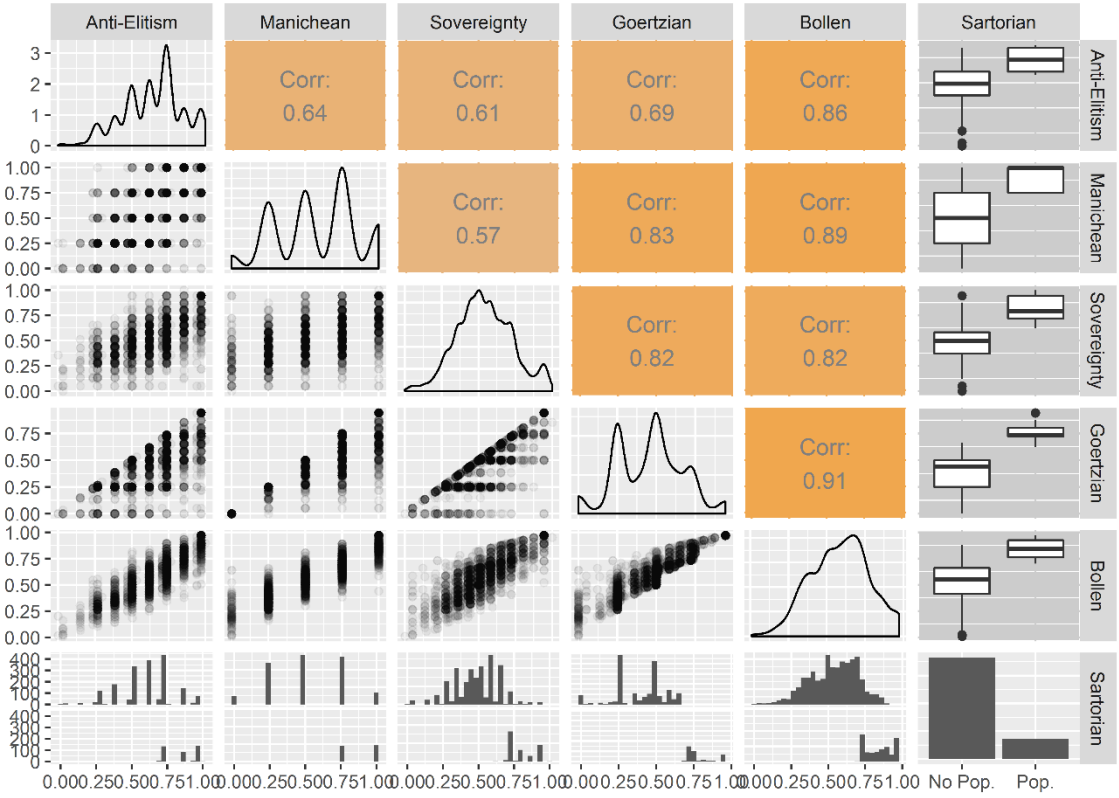

Anti-Elitism
Manichean
Sovereignty
Goertzian
Bollen
Sartorian
Corr: 0.64
Corr: 0.61
Corr: 0.69
Corr: 0.86
Corr: 0.57
Corr: 0.83
Corr: 0.89
Corr: 0.82
Corr: 0.82
Corr: 0.91
No Pop.
Pop.

Figure S8-3. **Internal Structure of Akkerman et al. populism scale (United Kingdom, Crowdflower non-probability panel)**

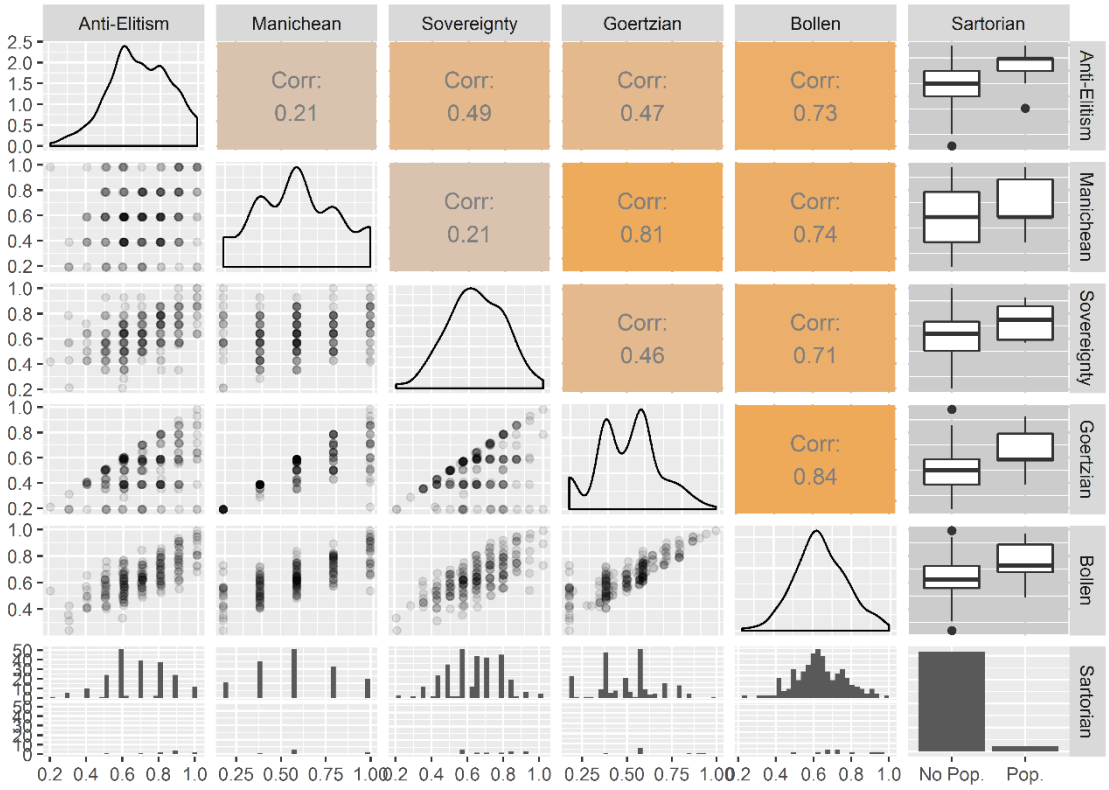


Figure S8-4. **Internal Structure of Akkerman et al. populism scale (USA, Crowdflower non-probability panel)**


Anti-Elitism
Manichean
Sovereignty
Goertzian
Bollen
Sartorian
Corr: 0.3
Corr: 0.54
Corr: 0.57
Corr: 0.77
Corr: 0.29
Corr: 0.79
Corr: 0.75
Corr: 0.64
Corr: 0.76
Corr: 0.89
No Pop.
Pop.

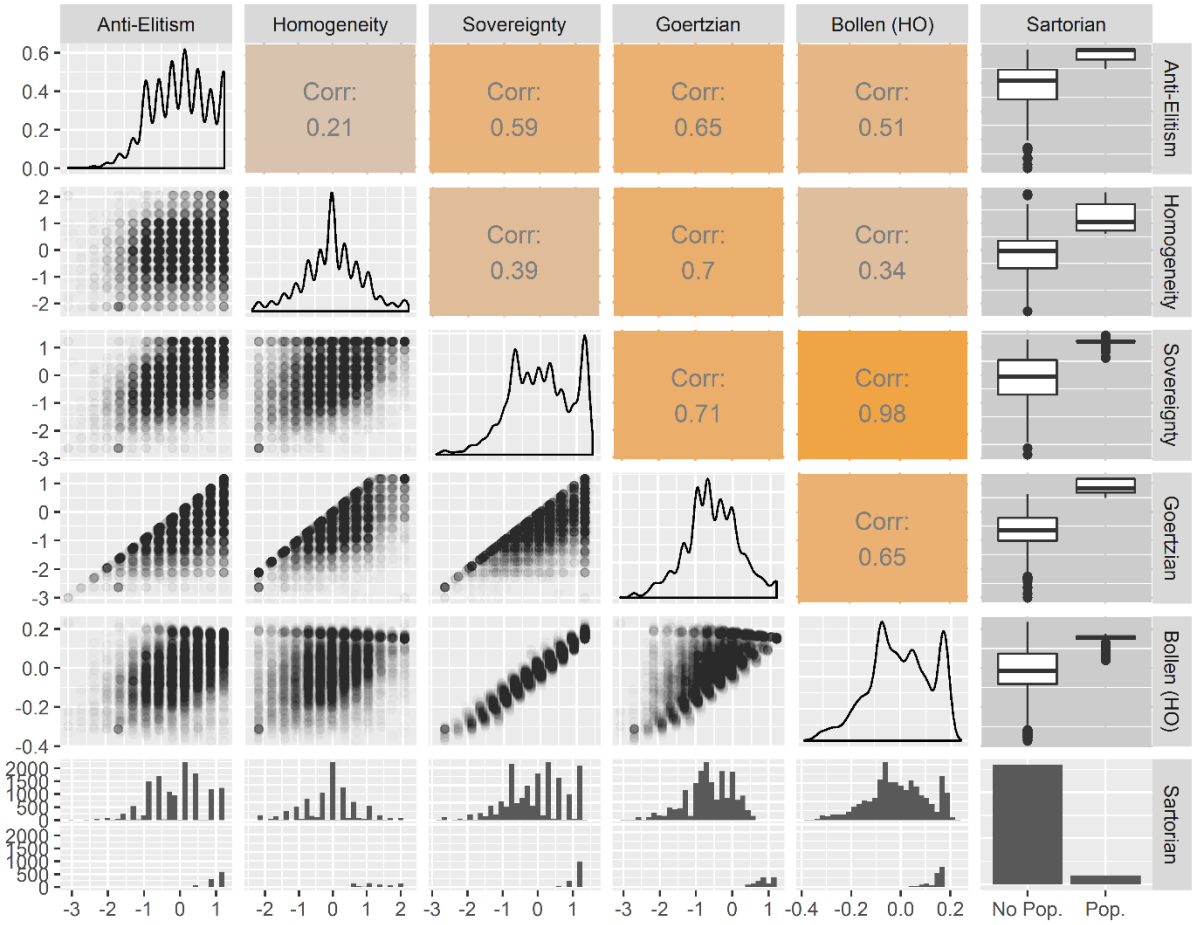
Supplement 9: Higher-order factor model

Because populist attitudes represent a multi-level construct with dimensions at the second level and indicators at the third level, many different operationalization techniques are conceivable. The main text focuses on the discussion of aggregation rules for the second level. For the aggregation of indicators into composite scores, we opted for a simple and transparent technique: unweighted summary scores. Yet, more sophisticated methods for aggregating indicators into dimensions are available. In particular, several validation studies of populism scales employ methods that are more technically advanced than the simple aggregation methods used in the main text of this study. Therefore, we replicated the analysis reported in Figure 2 of the main text, using the original authors' own methods. Specifically, following Schulz et al. (2018) we employed structural equation modeling to operationalize populist attitudes as a latent higher-order construct. That is, populist attitudes are modeled as a latent factor at the first level. The sub-dimensions are modeled as latent factors at the second level loading on the manifest indicators at the third level. In general, such a method can be subsumed under the Bollen approach because it treats the subdimensions as compensatory. Comparing the results reported in Figure S9-1 below with Figure 2 from the main text underscores the study's main finding: composite scores of populist attitudes do not yield identical results when then Goertz or Bollen operationalization strategies are applied. What is more, the discrepancy between the Goertz and the Bollen concept structure is even larger when we employ structural equation modeling to compute Bollen populism scores.

However, it should be noted that replicating the methods used by Schulz et al. (2018) on the Schulz et al. populism scale in the GLES dataset was not straightforward. The model did not converge using Stata 15.1 but it did converge using the R-lavaan package. Maybe the higher-order factor model is not well suited for the adapted Schulz et al. scale that was used in the German Campaign panel, which could explain the very strong correlation between the Bollen populism score and the sovereignty dimension.[21]

[21] Note that Hieda, Zenkyo and Nishikawa (2019) were also unable to reproduce a higher-order factor model in a sample of Japanese respondents based on the Schulz et al. scale.

Figure S9-1. **Internal structure of Schulz et al. populism scale using a Higher-Order Factor Model (GLES, Schulz et al. populism scale)**



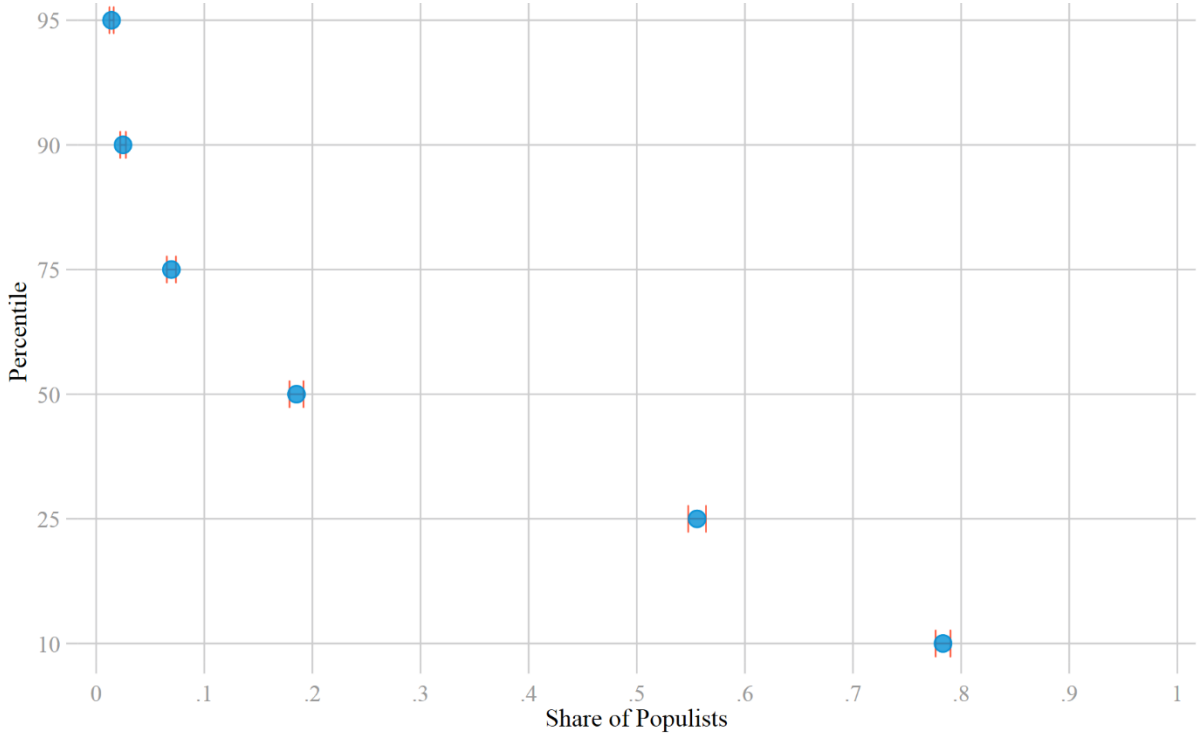
Supplement 10: Various operationalizations of the Sartori concept structure

To operationalize populist attitudes, according to the Sartori concept structure, we relied on the 75th percentile as a threshold for the classification of populist attitudes in the main text. According to this specification, individuals were classified as populists if, on each of the three populism subdimensions, they agree more strongly with the components of populism than the bottom 75% of respondents. Arguably, other thresholds are also plausible. Figure S10-1 demonstrates the empirical relationship between the threshold set on the subdimensions and the share of respondents who are classified as populists in a given survey sample. The results in Figure S10-1 are based on the Schulz et al. populism scale in the German campaign survey. Among other things, the plot shows the non-linear relationship between the threshold and the share of respondents who are classified as populists. For instance, using the 75th percentile instead of the 90th percentile as threshold multiplies the share of respondents who are classified as populists.

The plot visualizes the data structure of the German campaign survey and is intended to demonstrate the share of populists at various thresholds. However, it should be noted that theoretical reasoning about the essence of populist attitudes must guide the decisions regarding the adequate threshold.

Following the analyses in the main text, the results in Figure S10-1 rely on standardized indicators. As discussed in the main text, it is also possible to operationalize the Sartori concept structure using unstandardized indicators. Setting a meaningful threshold with unstandardized indicators requires to consider the scale of the indicators. The Schulz et al. populism scale, as it is used in the German Campaign Panel, provides three indicators for each of the three subdimensions of populist attitudes. The response scale of the indicators has the following values: strongly disagree; tend to disagree; neither agree nor disagree; tend to agree; strongly agree.

Figure S10-1. **Share of populists in the German Campaign Survey, by threshold**



We report results for the Sartori concept structure using unstandardized indicators based on two different thresholds. The more liberal threshold requires all individuals to "agree" or "strongly agree" with all indicators on all dimensions for the classification as populists. Applying this threshold, 4.5% of the respondents in the German campaign panel are categorized as populists. In other words, 4.5% of respondents agree with all components of populist attitudes.

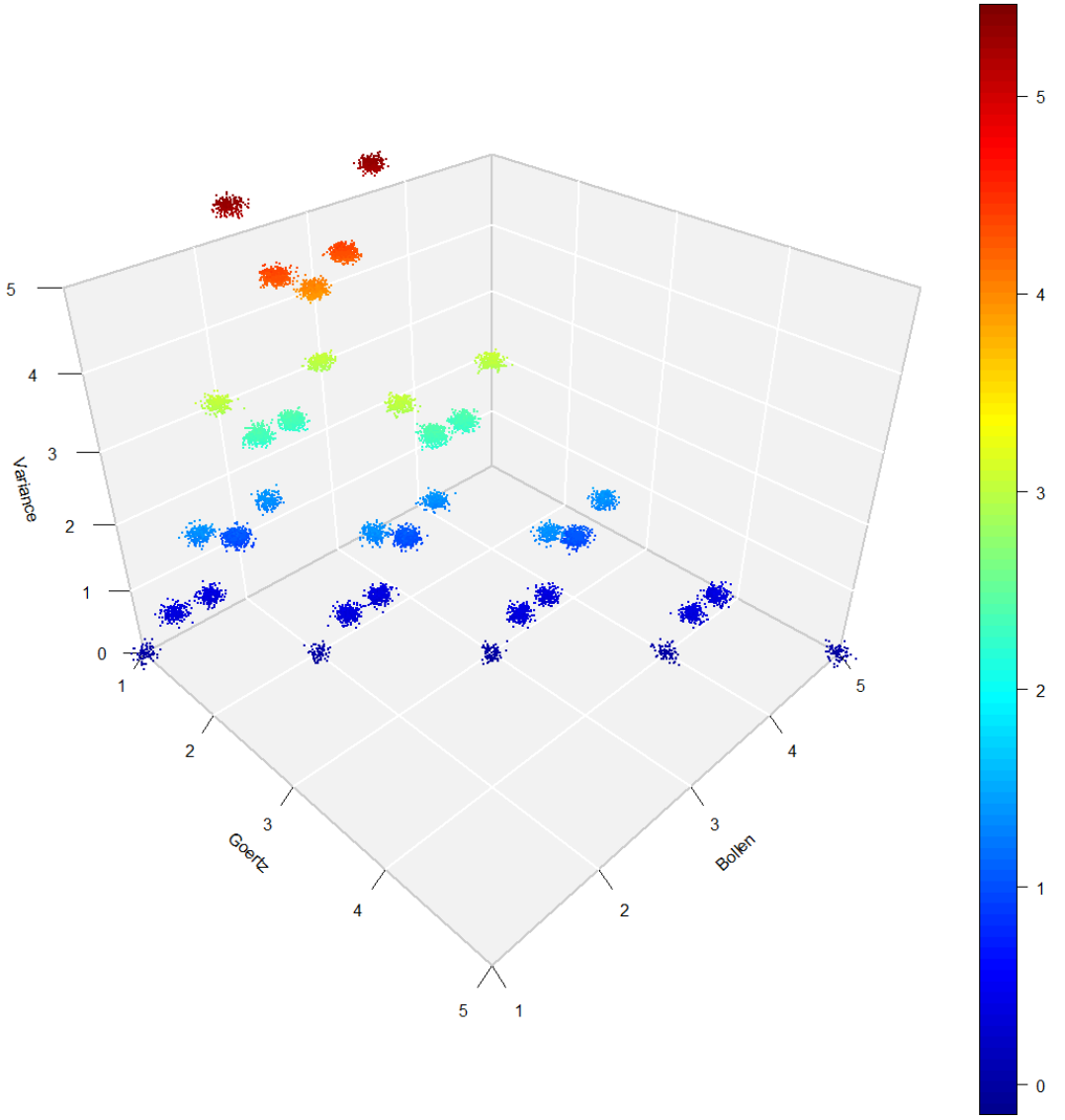
The stricter threshold requires individuals to "strongly agree" with the items. Yet, we still categorize individuals as populists if they "agree" (instead of "strongly agree") with one of the three items for each subdimension. Altogether, this threshold is higher than the liberal threshold because it requires stronger support of at least two components of populism on the mass level. In this specification, 3.0 % of the respondents are classified as populists. Altogether, mirroring the results in the main text using relative thresholds, these numbers suggest that a segment of the German sample of respondents exhibits a populist world view. However, the share of populist individuals is rather small.

Supplement 11: Simulation of the relationship between the Bollen and the Goertz composite scores

Which conditions determine the size of disparities between the Bollen and the Goertz constructs of populist attitudes? To examine this question, we simulated a dataset with 10,000 randomly drawn observations, assuming that populist attitudes comprise three subdimensions.

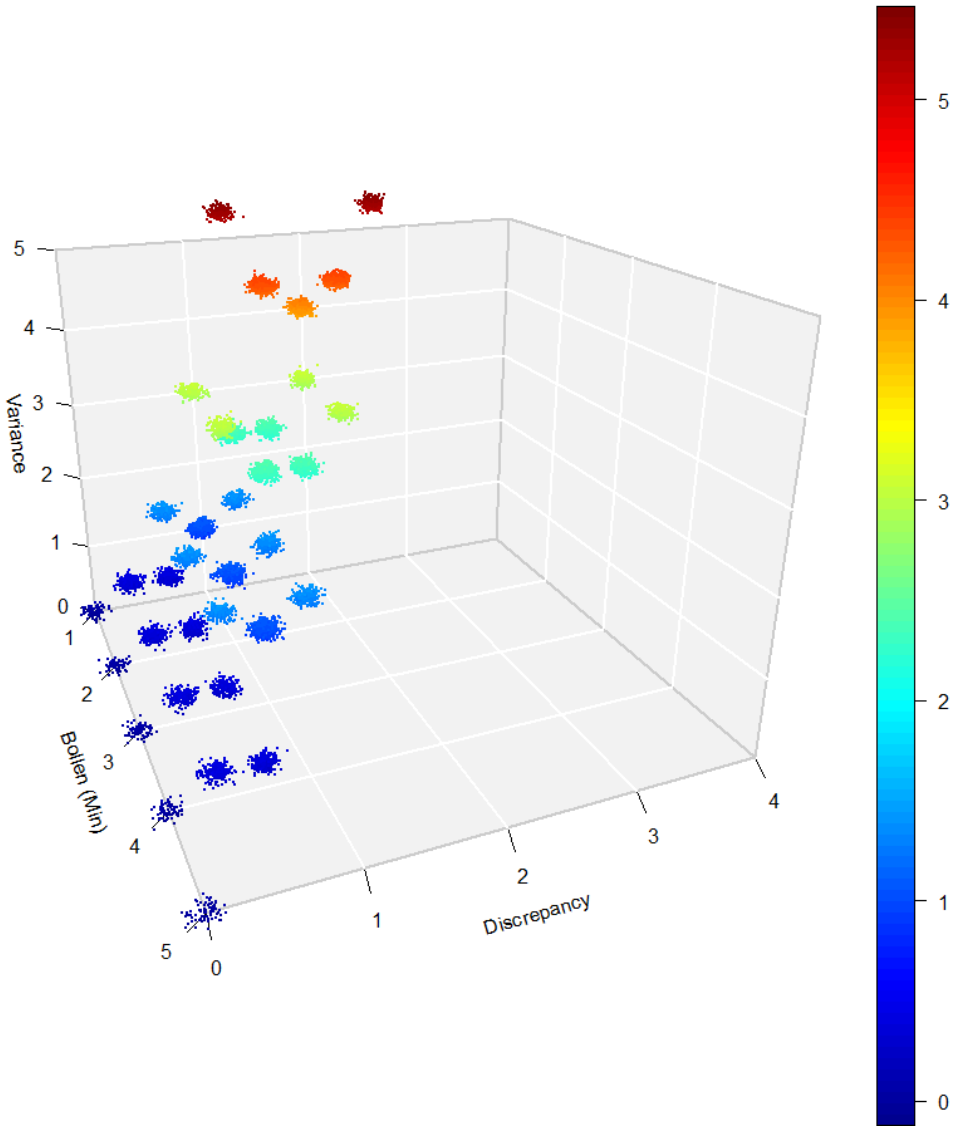
Figure S11-1 shows the empirical relationship between Bollen populism scores, Goertz populism scores, and the variance of the concept subdimensions. Figure S11-2 displays the empirical relationship between Goertz populism scores, the variance of the concept subdimensions, and the discrepancy between Bollen and Goertz scores in the simulated data set. The three-dimensional scatterplot demonstrates that the discrepancies between the Goertz and the Bollen populism scores increase when the Goertz scores increase. In addition, higher variances between the concept subdimensions also increase the discrepancy between the populism constructs.

Figure S11-1. **Relationship between Bollen populism scores, Goertz populism scores and the variance of concept subdimensions**



Notes: 10,000 randomly drawn observations, assuming three subdimensions of populist attitudes.

Figure S11-2. **Relationship between Goertz populism scores, the variance of the concept subdimensions and the discrepancy between Bollen and Goertz scores in a simulated data set**



Notes: 10,000 randomly drawn observations, assuming three subdimensions of populist attitudes.

Supplement 12: Correlations with conspiratorial thinking

Investigating how different concept structures of populist attitudes lead to different or similar correlations with substantive variables of interest, this appendix extends the analysis reported in Figure 5 in the main text. Specifically, it reports correlations of populist attitudes as derived with different operationalizations with conspiratorial thinking. Conspiratorial thinking is captured by a summary index of five indicators on the perception of hidden elite influences on the political process. The results vary between the four populism scales considerably.

The Castanho Silva et al. populism scale exhibits notable cross-national heterogeneity. People-centrism and Manichean outlook are positively correlated with conspiratorial thinking in some countries but negatively correlated in other countries. With respect to the disparities between the Goertz and the Bollen populism scores, the correlations with conspiratorial thinking differ significantly between both concept structures in all samples. In all but one case, the correlation is much weaker for the Goertz than for the Bollen concept structure. In several instances, only the correlation of the Bollen populism scores with conspiratorial thinking passes conventional levels of statistical significance, whereas the respective correlations of the Goertz populism scores do not.

Using the Oliver/Rahn populism scale, anti-elitist orientations are moderately correlated with conspiratorial thinking, but the other two subdimensions show hardly any association with the variable of interest. The disparities between the Goertz and Bollen concept structures are small to negligible. Using the Schulz et al. populism scale, all subdimensions are positively correlated with conspiratorial thinking in most of the samples. The evidence also exhibits small disparities between the Goertz and Bollen concept structures. Using the Akkerman et al. populism scale, the disparities between the Goertz and Bollen concept structures are negligible.

Altogether, whether and how the operationalization of populist attitudes makes a difference in investigating conspiratorial thinking depends a lot on the sample and the scale that is used.

Figure S12-1. **Bivariate correlations with conspiratorial thinking (Castanho Silva et al. populism scale)**

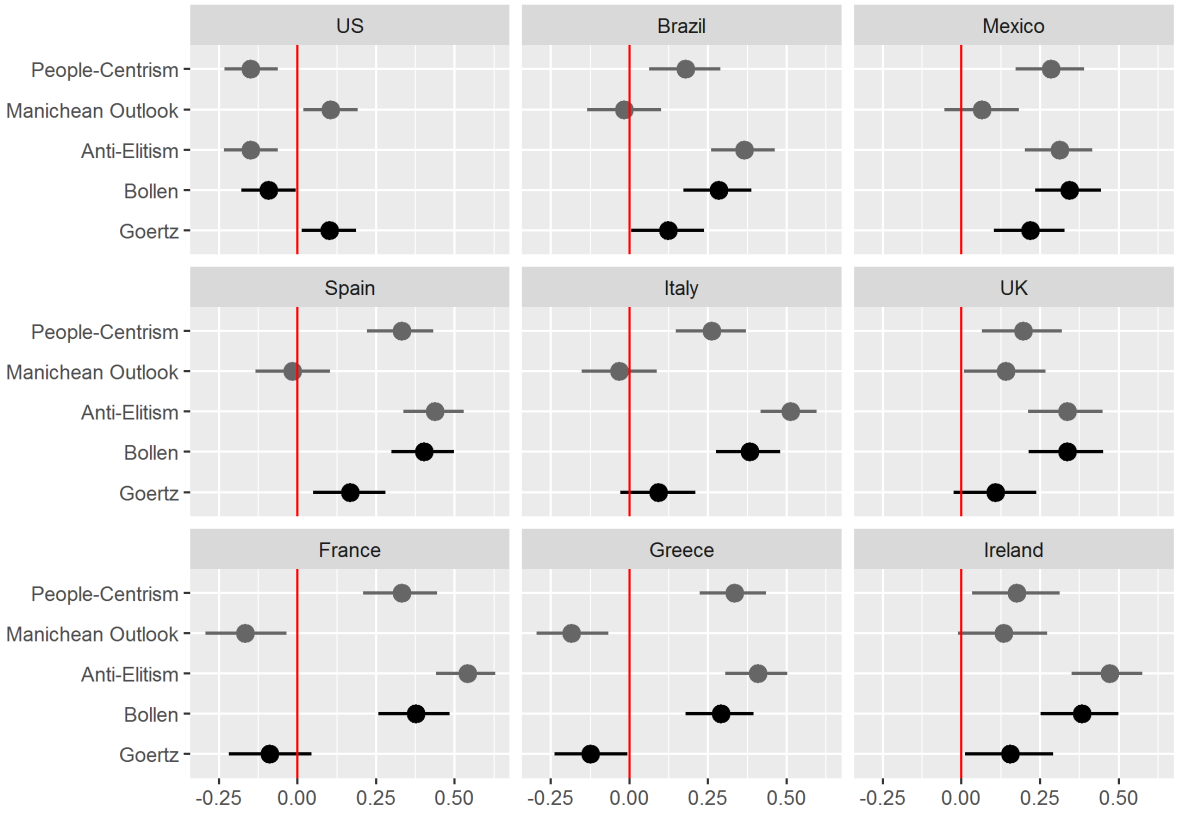


Figure S12-2. **Bivariate correlations with conspiratorial thinking (Oliver/Rahn populism scale)**

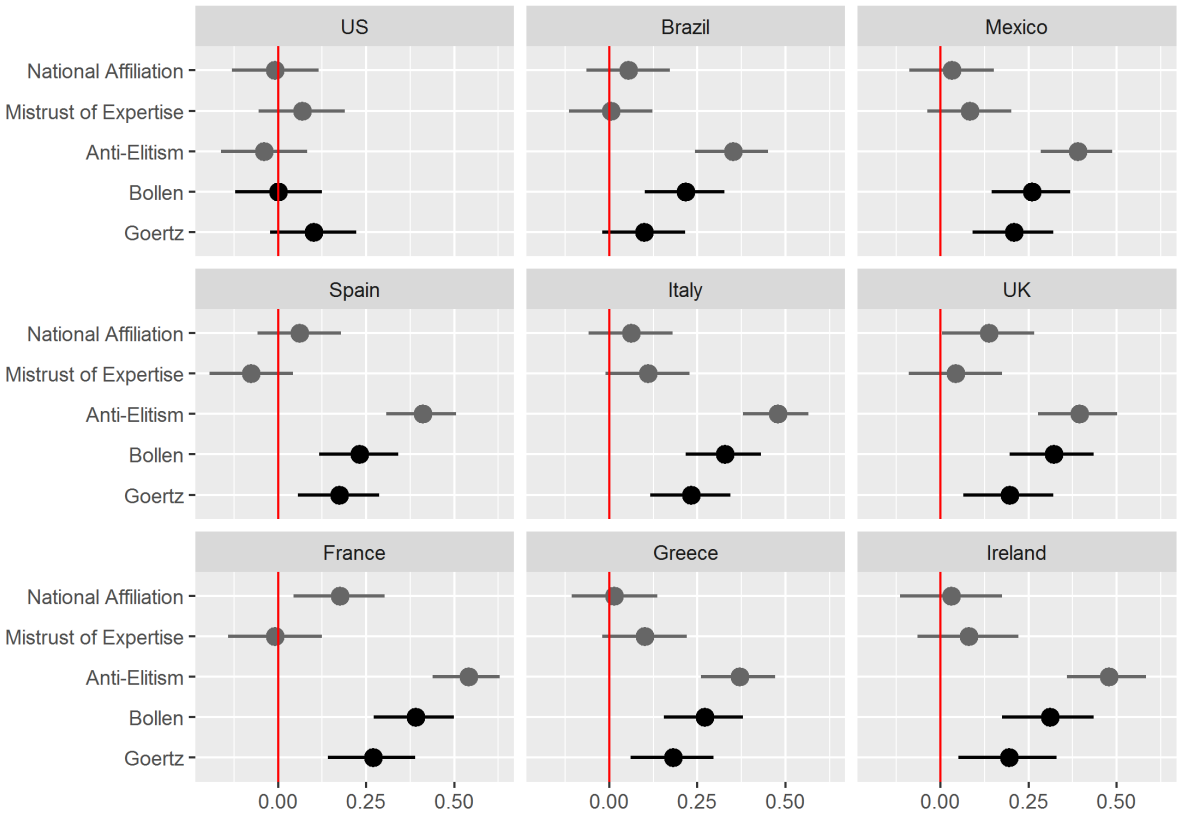


Figure S12-3. **Bivariate correlations with conspiratorial thinking (Schulz et al. populism scale)**

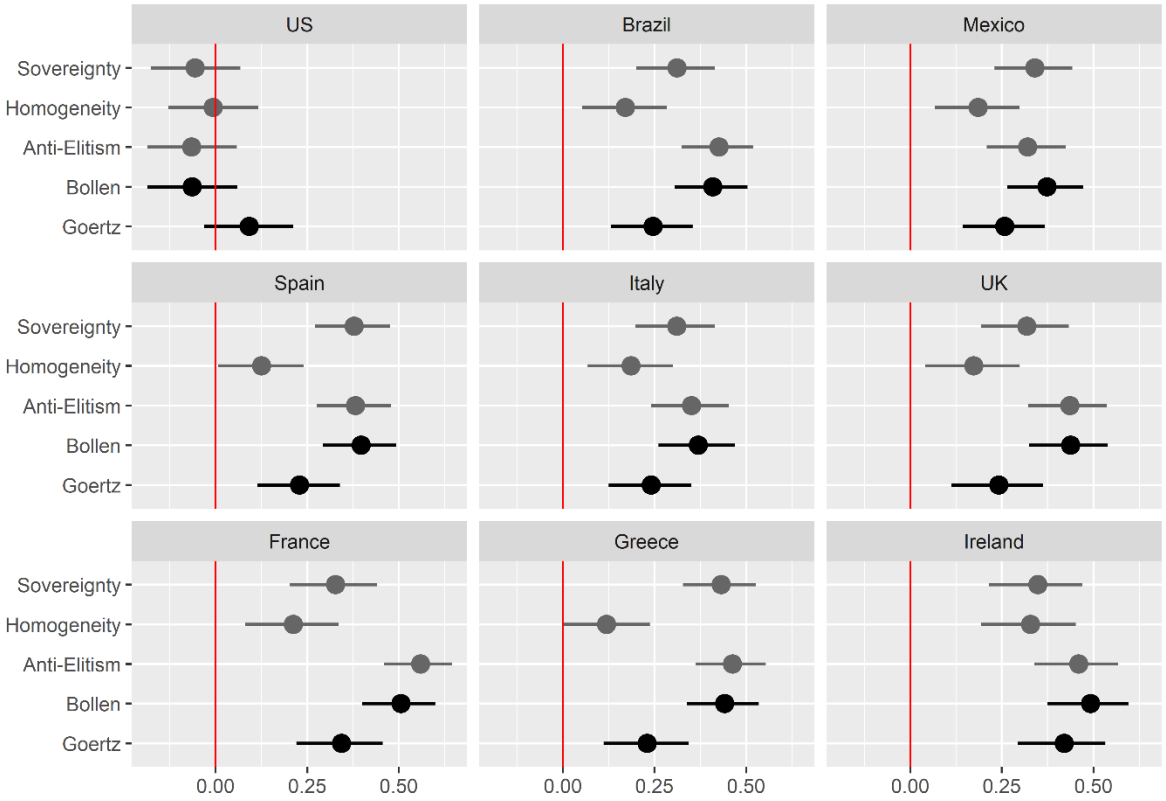
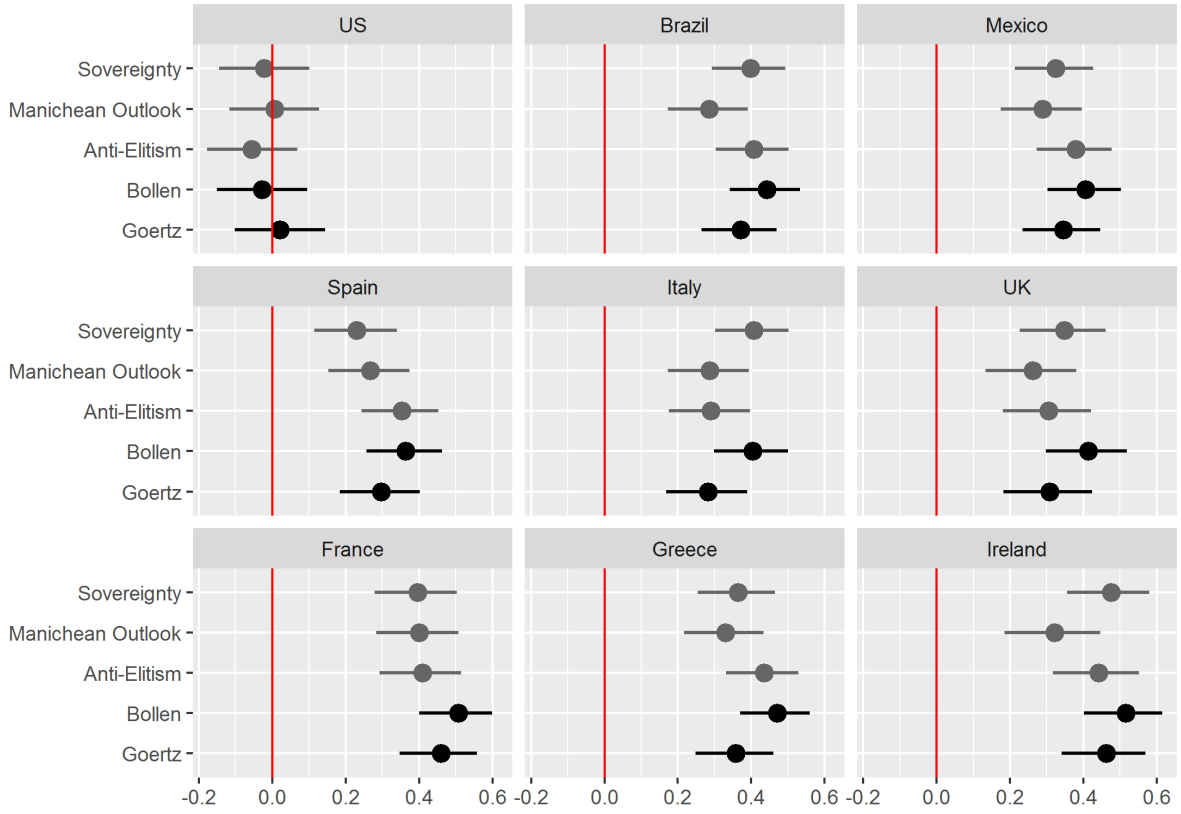


Figure S12-4. **Bivariate correlations with conspiratorial thinking (Akkerman et al. populism scale)**



Supplement 13: Correlations with institutional trust

Institutional trust was measured with three variables: Trust in the national government, trust in parliament and trust in political parties. In combining these indicators, we follow the practice by Castanho Silva et al. (2019). Substantively, the institutional trust variable reflects trust in political institutions. Table S13-1 shows that the internal consistency of the summary index is satisfactory in most samples, but it is considerably lower in the US and in the UK. This finding suggests that the empirical properties of the institutional trust variable vary between countries.

Table S13-1. **Cronbach's alpha of institutional trust across samples**

Sample	**Cronbach's alpha**
US	.62
Brazil	.86
Mexico	.81
Spain	.84

Italy	.90
UK	.69
France	.77
Greece	.83
Ireland	.78

In the main text, we reported correlations between institutional trust and different concept structures of populist attitudes using the Castanho Silva et al. populism scale. Figure S13-1 and Figure S13-2 replicate the analysis using the Oliver/Rahn populism scale and using the Schulz et al. populism scale. Compared to the results obtained with the Castanho Silva et al. scales, the discrepancies between populism scores are not as strong if the alternative scales are used. This is not surprising as the Castanho Silva et al. scale is particularly prone to concept-measurement inconsistencies of the Bollen approach due to the low covariances of the concept subdimensions. As written in the main text, thus, the extent of discrepancies between the Bollen and the Goertz concept structures depends on empirical properties of the scales and samples.

Figure S13-1. **Bivariate correlations with institutional trust (Oliver/Rahn populism scale)**

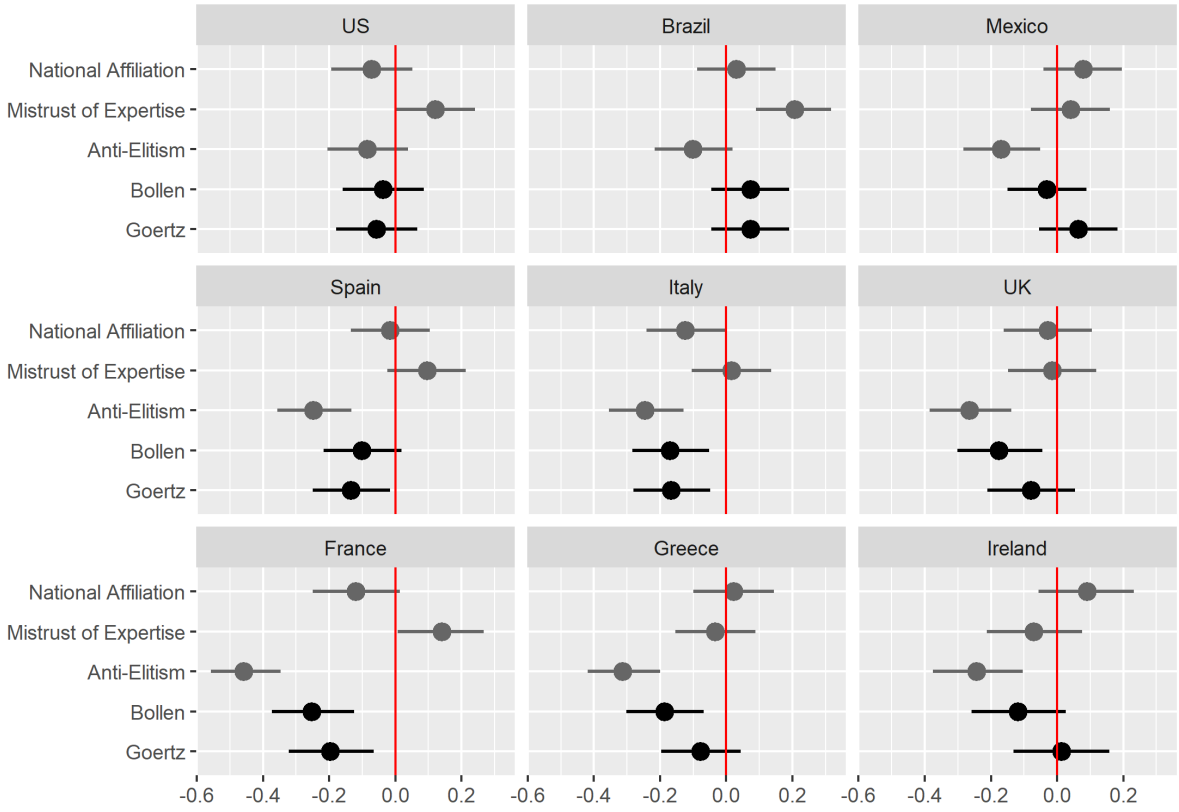
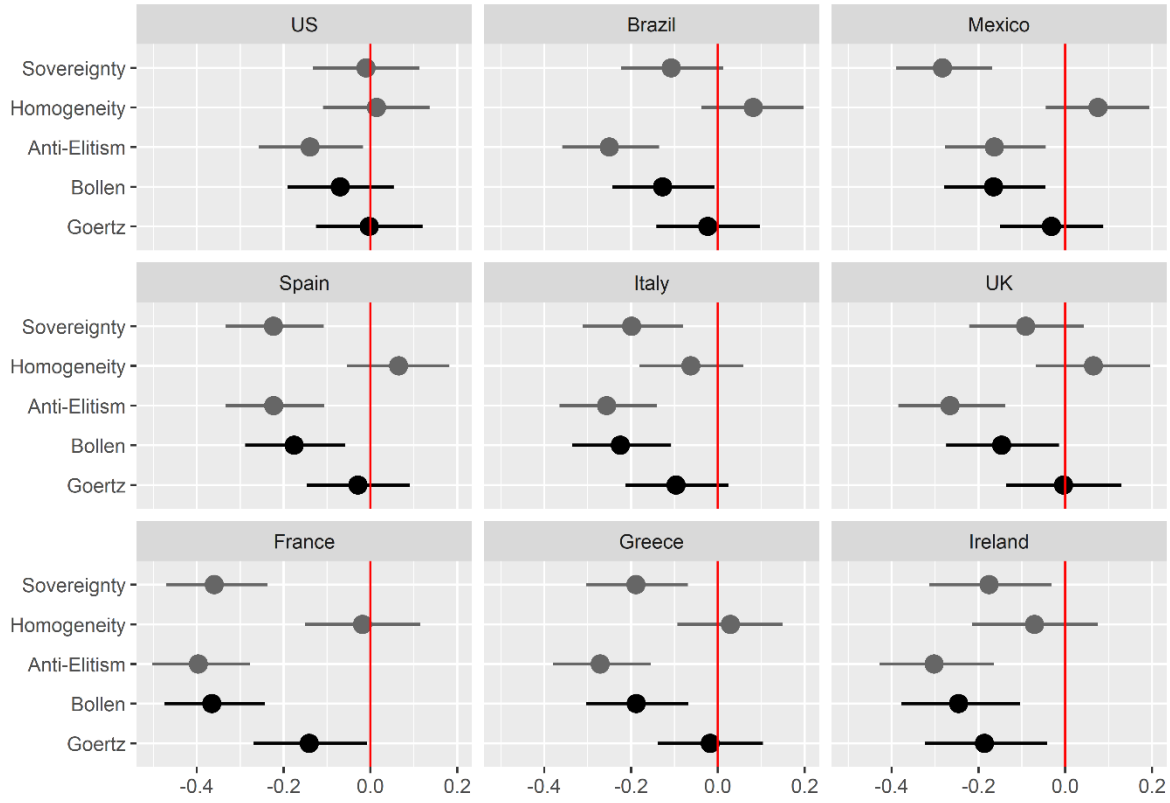


Figure S13-2. **Bivariate correlations with institutional trust (Schulz et al. populism scale)**



Supplement 14: Assessing non-compensatory multi-dimensional concepts already at the stage of measurement (Akkerman et al. scale)

The Akkerman et al. scale presents an interesting case which deserves further attention with regard to its logical organization and strategy of measurement. The scale differs crucially in one respect from other the scales reported in the main text as it employs a distinct strategy to assess the concept's multi-dimensionality.

In research on populist attitudes and in public opinion research, more generally, the most prevalent strategy to account for a concept's multi-dimensionality is to observe multiple, discrete subdimensions. Each of these subdimensions is then usually measured with one or multiple items where each item is reflective of one subdimension. As discussed in *Supplement* 2: Decisions to be made for the aggregation of multi-dimensional constructs, computational aggregation functions are then employed to combine these subdimensions into composite scores.

The Akkerman et al. scale, however, employs a different strategy (also see Elchardus and Spruyt 2016). Each indicator of the scale can be understood as intended to measure one concept subdimension but to incorporate multiple concept components into each single indicator (Akkerman, Mudde, and Zaslove 2014, 1332; Castanho Silva et al. 2019, 2). Specifically, some items of the Akkerman et al. scale are worded in a way so that respondents only agree with a statement if respondents concurrently agree with the tenets of multiple subdimensions (e.g., "*The people, and not politicians, should make our most important policy decisions*" which taps into sovereignty orientations and anti-elitist orientations, see Table S6-1 for an overview of the scale items). Using inherently multi-dimensional items to capture the multi-dimensionality of individual-level populism is thus crucially different from the approach to use uni-dimensional items to separate measure each subdimensions. Whereas the Akkerman et al. scale can be understood as following the first approach, the remaining populism scales can be understood as following the second approach. Note, however, that here we contrast the approach underlying the Akkerman et al. scale in a binary fashion with the approach underlying the other populism scales where, in fact, some of the items of the remaining scales are also not clearly one-dimensional.

Assessing multiple concept components simultaneously with one measurement instrument is an established research strategy in some disciplines. Consider educators or educational researchers who assess the competency of applying arithmetic skills to a problem that is described in plain text (Chalmers and Flora 2014, 341). Apparently, solving such math word test requires skills in two competency domains (maths and text comprehension), which are non-compensatory. Instead of assessing the intersection of these competency domains by separately measuring mathematical ability and text comprehension and then aggregating these two dimensions according to the Goertz or Sartori operationalization strategies (the classical strategy that is also employed on the populism scales in the main text), another option is to account for the non-substitutability already on the stage of measurement.

For assessing the non-substitutability already on the stage of measurement, it is conceivable to administer a math word test that can only be solved when pupils simultaneously possess both verbal and mathematical competencies. High values on such a math word test indicate that pupils score high on both dimensions because solving the test will be impossible if individuals either lack text comprehension skills or arithmetic competencies. Accounting for non-substitutability already

at the stage of measurement has at least two advantages: First, this strategy may achieve high congruence between measurement and the target concept (content validity). Second, this approach renders unnecessary any aggregation of distinct subdimensions. Hence, assessing the simultaneous presence of jointly necessary concept components appears to be generally feasible and in certain situations, potentially desirable for assessing multi-dimensional concepts with non-compensatory concept components.

However, this approach also comes with distinctive drawbacks. To begin with, this approach reduces the informational value of the assessment. Whereas the meaning of high values of such measurement is clear, the meaning and origins of assessments with low values are not clear. Researchers have no way of disentangling whether low values result from a lack of text comprehension or of arithmetic skills. Hence, the multi-concept measurement approach impedes in-depth analyses of the composition of individual competencies.

Things become worse once we turn to survey research. Researchers design survey questions to elicit object-specific responses from respondents. Differences in survey responses should reflect differences in respondents in terms of the target concept, i.e., the evaluation of an object. For comparability across individuals, survey researchers strive to administer stimuli that function similarly across individuals (Krosnick and Presser 2010). Equivalent item functioning is threatened if individuals do not interpret survey items in a similar way. This might be the case if a survey question alludes to a variety of different aspects of one phenomenon in order, e.g., to capture the response to a combination of two or more aspects. Some people may place greater relative weight on one aspect when constructing their answer, while other respondents place higher weight on another aspect. Therefore, the more complex a survey item is, the less certain a researcher can be how an item will be interpreted. In these cases, differences in a survey response may reflect substantive differences in the target concept or solely differences in how respondents interpreted the question stimulus. Consequently, in developing survey items psychometric textbooks usually advise to keep survey items as simple as possible and to avoid double-barreled items (Krosnick and Presser 2010). Otherwise, the meaning of all responses – and not only of some outcomes as in case of the competence measurement mentioned above – becomes unclear.

To conclude, assessing the simultaneous presence of jointly necessary concept components already at the stage of measurement process comes with drawbacks particularly in survey research,

but it is generally feasible and, in some research areas, an even promising approach. These considerations also apply to the Akkerman et al. scale and should be kept in mind when using the scale.

Supplement 15: Overview of data sources

Table S15-1 provides an overview of the survey datasets that underlie this study with further information on the scale that is contained in the dataset, in which country the dataset was surveyed and for which analysis the dataset was employed.

Table S15-1. **Data Sources**

Dataset	*Scale*	*Country*	*Analysis*
ANES 2016 Time Series	CSES	United States	Shiny Web Applet: How many populists, Correlations
AUTNES 2018 CSES Edition, doi:10.11587/W193UZ	CSES	Austria	Shiny Web Applet: How many populists, Correlations
BES 2017: Face-to-Face Post-Election Survey, doi:10.5255/UKDA-SN-8418-1	CSES	UK	Shiny Web Applet: How many populists, Correlations
GLES 2017 Post-election Cross Section, doi:10.4232/1.13235	CSES, Akkerman et al.	Germany	Shiny Web Applet: How many populists, Correlations
LISS 151.3 Measurement 3	Akkerman et al.	Netherlands	Shiny Web Applet: How many populists, Correlations
GLES Campaign Panel, doi:10.4232/1.13150	Schulz et al.	Germany	Shiny Web Applet: Correlations, Main text: Figure 2
Castanho Silva et al. replication material	Akkerman et al., Castanho Silva et al., Oliver/Rahn, Schulz et al.	Brazil, France, Greece, Ireland, Mexico, Spain, Italy, UK, United States	Main text: Figures 2-4

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.

Alkire, Sabina, and James Foster. 2011a. “Counting and multidimensional poverty measurement.” *Journal of Public Economics* 95 (7): 476–87. http://www.sciencedirect.com/science/article/pii/S0047272710001660.

———. 2011b. “Understandings and misunderstandings of multidimensional poverty measurement.” *The Journal of Economic Inequality* 9 (2): 289–314. https://link.springer.com/content/pdf/10.1007/s10888-011-9181-4.pdf (Accessed April 4, 2019).

Alkire, Sabina, James E. Foster, Suman Seth, Maria E. Santos, José M. Roche, and Paola Ballón. 2015. *Multidimensional poverty measurement and analysis*. Oxford: Oxford University Press.

Andreadis, Ioannis, Yannis Stavrakakis, and Giorgos Katsambekis. 2016. “Using Surveys to Measure the Populist Attitudes of Political Elites and Voters: A Greek Pilot Study of Supply and Demand.”

Atkinson, A. B. 2003. “Multidimensional Deprivation: Contrasting Social Welfare and Counting Approaches.” *The Journal of Economic Inequality* 1 (1): 51–65. https://doi.org/10.1023/A:1023903525276.

Babcock, Ben. 2011. “Estimating a noncompensatory IRT model using Metropolis within Gibbs sampling.” *Applied Psychological Measurement* 35 (4): 317–29.

Berinsky, Adam J., Gregory A. Huber, and Gabriel S. Lenz. 2012. “Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk.” *Political Analysis* 20 (3): 351–68. https://www.cambridge.org/core/services/aop-cambridge-core/content/view/348F95C0FBCF21C3B37D66EB432F3BA5/S1047198700013875a.pdf/div-class-title-evaluating-online-labor-markets-for-experimental-research-amazon-com-s-mechanical-turk-div.pdf.

Bieber, Ina E., and Evelyn Bytzek. 2012. “Online-Umfragen: eine geeignete Erhebungsmethode für die Wahlforschung? ; ein Vergleich unterschiedlicher Befragungsmodi am Beispiel der Bundestagswahl 2009.” *Methoden, Daten, Analysen (mda)* 6 (2): 185–211. https://www.ssoar.info/ssoar/bitstream/document/34341/1/ssoar-mda-2012-2-bieber_et_al-Online-Umfragen__eine_geeignete_Erhebungsmethode.pdf.

Bollen, Kenneth, and Richard Lennox. 1991. “Conventional wisdom on measurement: A structural equation perspective.” *Psychological Bulletin* 110 (2): 305–14.

Bollen, Kenneth A., and Judea Pearl. 2013. “Eight myths about causality and structural equation models.” In *Handbook of causal analysis for social research* [eng]. *Handbooks of sociology and social research*, ed. Stephen L. Morgan. Dordrecht: Springer, 301–28.

Bolt, Daniel M., and Venessa F. Lall. 2003. “Estimation of Compensatory and Noncompensatory Multidimensional Item Response Models Using Markov Chain Monte Carlo.” *Applied Psychological Measurement* 27 (6): 395–414.

Bonikowski, Bart, and Paul DiMaggio. 2016. “Varieties of American Popular Nationalism.” *American Sociological Review* 81 (5): 949–80.

Castanho Silva, Bruno, Ioannis Andreadis, Eva Anduiza, Nebojsa Blanusa, Yazmin M. Corti, Gisela Delfino, Guillem Rico, Saskia Ruth, Bram Spruyt, Marco R. Steenbergen, and Levente Littvay. 2018. “Public Opinion Surveys: A New Scale.” In *The Ideational Approach to Populism: Concept, Theory, and Analysis*, eds. Kirk A. Hawkins, Ryan Carlin, Levente Littvay and Cristóbal R. Kaltwasser. London: Routledge.

Castanho Silva, Bruno, Jungkunz, Marc Helbling, and Levente Littvay. 2019. “An Empirical Investigation of Seven Populist Attitudes Scales.” *Political Research Quarterly* First View: 1–16.

Chalmers, R. P. 2012. “mirt : A Multidimensional Item Response Theory Package for the R Environment.” *Journal of Statistical Software* 48 (6). https://www.researchgate.net/profile/R_Chalmers/publication/228534363_Mirt_A_Multidimensional_Item_Response_Theory_Package_for_the_R_Environment/links/0deec53c1e2b5c9303000000/Mirt-A-Multidimensional-Item-Response-Theory-Package-for-the-R-Environment.pdf (Accessed March 22, 2019).

———. 2018. “Model-Based Measures for Detecting and Quantifying Response Bias.” *Psychometrika* 83 (3): 696–732. https://doi.org/10.1007/s11336-018-9626-9.

Chalmers, R. P., and David B. Flora. 2014. “Maximum-Likelihood Estimation of Noncompensatory IRT Models With the MH-RM Algorithm.” *Applied Psychological Measurement* 38 (5): 339–58.

Collins, Linda M., and Stephanie T. Lanza. 2010. *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences* [eng]. v.718 of *Wiley series in probability and statistics*. Hoboken, NJ: Wiley.

DeMars, Christine E. 2016. “Partially Compensatory Multidimensional Item Response Theory Models: Two Alternate Model Forms.” [eng]. *Educational and Psychological Measurement* 76 (2): 231–57.

Edwards, Michael C. 2009. “An Introduction to Item Response Theory Using the Need for Cognition Scale.” *Social and Personality Psychology Compass* 3 (4): 507–29.

Elchardus, Mark, and Bram Spruyt. 2016. “Populism, Persistent Republicanism and Declinism: An Empirical Analysis of Populism as a Thin Ideology.” *Government and Opposition* 51 (01): 111–33.

Goertz, Gary. 2006. *Social science concepts: A user's guide*. Princeton, NJ: Princeton Univ. Press.

Greco, Salvatore, Alessio Ishizaka, Menelaos Tasiou, and Gianpiero Torrisi. 2019. “On the Methodological Framework of Composite Indices: A Review of the Issues of Weighting, Aggregation, and Robustness.” *Social Indicators Research* 141 (1): 61–94.

Hameleers, Michael, Linda Bos, and Claes H. de Vreese. 2017. “The Appeal of Media Populism: The Media Preferences of Citizens with Populist Attitudes.” *Mass Communication and Society* 20 (4): 481–504.

Hawkins, Kirk A., and Scott Riding. 2010. *Populist Attitudes and their Correlates among Citizens: Survey Evidence from the Americas*. Paper prepared for the ECPR Workshop “Disassembling Populism (and Putting It Back Together Again): Collaborative Empirical Research on Interactions among Populism’s Attributes,” March 22-25, 2010, Muenster, Germany.

Hawkins, Kirk A., Scott Riding, and Cas Mudde. 2012. “Measuring Populist Attitudes.” (55). IPSA Committee on Concepts and Methods Working Paper Series.

Hieda, Takeshi, Masahiro Zenkyo, and Masaru Nishikawa. 2019. “Do populists support populism? An examination through an online survey following the 2017 Tokyo Metropolitan Assembly election.” *Party Politics* 23 (4): 135406881984811.

Hobolt, Sara, Eva Anduiza, Ali Carkoglu, Georg Lutz, and Nicolas Sauger. 2016. *CSES Module 5 - Democracy Divided? People, Politicians and the Politics of Populism*. CSES Planning Committee Module 5 Final Report.

Hobolt, Sara B., and James Tilley. 2016. “Fleeing the centre: the rise of challenger parties in the aftermath of the euro crisis.” *West European Politics* 39 (5): 971–91.

Jacobs, Kristof, Agnes Akkerman, and Andrej Zaslove. 2018. “The voice of populist people? Referendum preferences, practices and populist attitudes.” *Acta Politica* 53 (4): 517–41.

Kaplan, D. 2008. *Structural Equation Modeling: Foundations and Extensions*. Sage Publications.

Kolen, Michael J., and Robert L. Brennan. 2004. *Test Equating, Scaling, and Linking: Methods and Practices* [eng]. *Statistics for Social Science and Public Policy*. New York, NY: Springer.

Krosnick, Jon A., and Stanley Presser. 2010. “Question and Questionnaire Design.” In *Handbook of survey research* [eng], eds. Peter v. Marsden and James D. Wright. Bingley: Emerald.

Liu, Chen-Wei, and R. P. Chalmers. 2018. “Fitting item response unfolding models to Likert-scale data using mirt in R.” *PLOS ONE* 13 (5): e0196292.

Mader, Matthias, and Harald Schoen. 2019. “The European refugee crisis, party competition, and voters’ responses in Germany.” *West European Politics* 42 (1): 67–90.

Mazziotta, Matteo, and Adriano Pareto. 2016. “On a Generalized Non-compensatory Composite Index for Measuring Socio-economic Phenomena.” *Social Indicators Research* 127 (3): 983–1003 (Accessed April 5, 2019).

Mohrenberg, Steffen, Robert A. Huber, and Tina Freyburg. 2019. “Love at first sight? Populist attitudes and support for direct democracy.” *Party Politics* Online First.

Møller, J., and S. E. Skaaning. 2012. *Requisites of Democracy: Conceptualization, Measurement, and Explanation*. Taylor & Francis.

Mudde, Cas. 2017. “Populism: An Ideational Approach.” In *The Oxford Handbook of Populism*. *Oxford handbook of politics*, eds. Cristobal R. Kaltwasser, Paul Taggart, Paulina O. Espejo and Pierre Ostiguy. New York NY: Oxford University Press, 27–47.

Mullinix, Kevin J., Thomas J. Leeper, James N. Druckman, and Jeremy Freese. 2015. “The Generalizability of Survey Experiments.” *Journal of Experimental Political Science* 2 (2): 109–38. https://www.cambridge.org/core/services/aop-cambridge-core/content/view/72D4E3DB90569AD7F2D469E9DF3A94CB/S2052263015000196a.pdf/div-class-title-the-generalizability-of-survey-experiments-a-href-afn1-ref-type-fn-a-div.pdf.

Munck, Gerardo L. 2009. *Measuring democracy : a bridge between scholarship and politics*. *Democratic transition and consolidation*. Baltimore, Md: Johns Hopkins Univ. Press.

Oliver, J. E., and Wendy M. Rahn. 2016. “Rise of the Trumpenvolk.” *The ANNALS of the American Academy of Political and Social Science* 667 (1): 189–206.

Pappas, Takis S. 2016. “The Specter Haunting Europe: Distinguishing Liberal Democracy's Challengers.” *Journal of Democracy* 27 (4): 22–36.

Plattner, Marc. F. 2010. “Populism, Pluralism, and Liberal Democracy.” *Journal of Democracy* 21 (1): 81–92.

Rico, Guillem, Marc Guinjoan, and Eva Anduiza. 2017. “The Emotional Underpinnings of Populism: How Anger and Fear Affect Populist Attitudes.” *Swiss Political Science Review* 23 (4): 444–61.

Schneider, Carsten Q., and Claudius Wagemann. 2012. *Set-theoretic methods for the social sciences: A guide to qualitative comparative analysis* [eng]. *Strategies for social inquiry*. Cambridge: Cambridge University Press.

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

Spierings, Niels, and Andrej Zaslove. 2017. “Gender, populist attitudes, and voting: Explaining the gender gap in voting for populist radical right and populist radical left parties.” *West European Politics* 40 (4): 821–47.

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.

Stanley, Ben. 2011. “Populism, nationalism, or national populism? An analysis of Slovak voting behaviour at the 2010 parliamentary election.” *Communist and Post-Communist Studies* 44 (4): 257–70.

Steiner, Nils D., and Claudia Landwehr. 2018. “Populistische Demokratiekonzeptionen und die Wahl der AfD: Evidenz aus einer Panelstudie.” *Politische Vierteljahresschrift*.

United Nations. 2019. “Why is the geometric mean used for the HDI rather than the arithmetic mean?” January 4. http://hdr.undp.org/en/content/why-geometric-mean-used-hdi-rather-arithmetic-mean.

Urbinati, N. 2014. *Democracy Disfigured*. Harvard University Press.

van Hauwaert, Steven M., and Stijn van Kessel. 2018. “Beyond protest and discontent: A cross-national analysis of the effect of populist attitudes and issue positions on populist party support.” *European Journal of Political Research* 57 (1): 68–92.

Vehrkamp, Robert, and Christopher Wratil. 2017. *A Populist Moment?: Populist Attitudes of Voters and Non-Voters before the German Federal Election 2017*. Bertelsmann.

Wuttke, Alexander. 2019. “Why Too Many Political Science Findings Cannot Be Trusted and What We Can Do About It: A Review of Meta-Scientific Research and a Call for Academic Reform.” *Politische Vierteljahresschrift* 60 (1): 1–19. https://doi.org/10.1007/s11615-018-0131-7.