Agree to Agree: Correcting Acquiescence Bias in the Case of Fully Unbalanced Scales with Application to UK Measurements of Political Beliefs

Phil Swatton
University of Essex
p.j.swatton@essex.ac.uk

Abstract: A methodologically important area in political science is measuring the ideology of voters. This task can be difficult, and researchers often rely on 'off the shelf' datasets. Many of these datasets contain unbalanced Likert scales, which risk acquiescence bias. This paper proposes a strategy for dealing with this issue. I first demonstrate using two comparable datasets from the UK how unbalanced scales produce distorted distributions and can affect regression results. Then, building on past research that utilises factor analysis to eliminate the influence of acquiescence bias, I demonstrate how researchers can utilise a person intercept confirmatory factor analysis model to obtain factor scores corrected for acquiescence in the case of fully unbalanced scales. I conclude with practical recommendations for researchers and survey designers moving forward.

Earlier version of this paper presented at APSA Conference, 1 - 3 Oct 2021

1 Introduction

In political science, a substantive area of interest is the ideology of voters. It follows that a methodologically important area of research is the measurement of voter ideology. In practice, due to the time and expense involved in collecting survey data, a great many political scientists (and indeed social scientists more broadly) are reliant on 'off the shelf' survey data produced by other researchers. A popular approach to measuring voter ideology in surveys is the Likert scale. Two types of Likert scale exist: balanced, which are built from an equal number of indicators for both 'sides' of the ideological dimension, and unbalanced, which have more indicators for one of the 'sides' of the ideological dimension. The indicators which are used to build Likert scales are typically subject to acquiescence bias and so where unbalanced scales are used, so too do the final scales. If researchers utilising these scales fail to consider the survey design and response stage of the data generating process (DGP), this will likely lead to incorrect research conclusions being drawn.

While past research has dealt with the problem of unbalanced scales, it has not dealt with the more difficult problem of fully unbalanced scales. In this paper I therefore discuss the problem of fully unbalanced scales, propose a model-based solution, and conclude by providing recommendations to both survey designers and users. The model-based solution I propose is an adaptation of a model previously proposed for balanced and partially unbalanced scales, which I label the person-intercept confirmatory factor analysis (CFA) approach. This approach leverges the common factor model to capture the acquiescence component latent in survey responses.

I begin with a substantive discussion of acquiescence bias in terms of the common factor model. This model divides latent variation between content factors, measurement factors, and unique variation. I use this model to discuss the assumptions underlying Likert scales, and how unbalanced scales introduce acquiescence bias. I proceed with a demonstration of acquiscence bias between two comparable datasets in the form of the British Election Study and the British Social Attitudes survey. These results serve as a baseline for the correction methods I apply. I discuss past work on person intercept CFA, and develop four variations of the model. I discuss the need for empirical identification in the case of fully unbalanced Likert scales. I then apply these to a third dataset in the form of the 14th wave of the British Election Study internet panel. The results broadly show that the correction methods applied succeed in producing results more akin to fully balanced scales. I conclude with recommendations for researchers and survey designers.

2 Measuring Voter Ideology

Voter ideology, political beliefs, or political attitudes represent an inherently ambiguous concept. Exactly what it is, what label to give it, how many dimensions it's composed of, and which of those dimensions we should be interested in are all contested. Even once researchers have agreed on a set of answers for the purpose of a given research project, it follows that it is not straightforward how to capture a given definition among survey respondents. One solution to this issue is the use of Likert scales. In this question format, respondents are shown a set of statements and given a range of responses, often five ranging from 'strongly disagree' to 'strongly agree'. Scores from these responses for each statement are then tallied to produce a final measurement of the concept of interest.

2.1 The Common Factor Model

For a given set of indicators of the same concept of interest, we can adopt a generative understanding of the measure. This means the indicator is assumed to be 'generated' by the target concept, and variation in the target concept causes variation in the indicator. One method of expressing this is via the common factor model [Brown, 2015]

$$x_{ij} = \lambda_{j1}\eta_{i1} + \ldots + \lambda_{jm}\eta_{im} + \epsilon_{ij} \tag{1}$$

where x_{ij} is the jth observed indicator for respondent i, η_{im} is the mth latent factor for respondent i, λ_{jm} is the loading on the mth factor for indicator j, and ϵ_{ij} is the unique factor for the jth observed indicator for respondent i.

The latent factors underlie the observed measurements. They can be further split between content factors capturing substantive variation and measurement factors capturing variation due to the measurement method of choice [Kenny and Kashy, 1992]. The unique factor captures variation in that indicator not found in any other indicators, which will be a mix of random noise and unique substantive variation. For my theoretical purposes in this paper, I adopt this common factor model.

2.2 Acquiescence Bias

Substantively, acquiescence bias can be described as a tendency to be more likely to 'agree' with survey statements regardless of their content. In terms of the common factor model, we can express acquiescence as a second common factor:

$$x_{ij} = \lambda_{jc}\eta_{ic} + \lambda_{ja}\eta_{ia} + \epsilon_{ij} \tag{2}$$

Here the c subscript denotes the concept of interest, while the a subscript denotes acquiescence. When aggregating indicators with this DGP into a Likert scale, we make several implicit assumptions. First, we assume that λ_{jc} is constant across indicators *except* that for some indicators its sign 'flips' depending on the direction of the statement. If for example we have a left-right factor, then we can imagine its sign being negative for left-wing indicators and positive for right-wing indicators. This assumption tends to be incorrect in practice [Billiet and Davidov, 2008, 545], but it need only be a reasonably close approximiation to be successful. Second, in the case of a balanced Likert scale we are assuming that λ_{ja} is also constant across indicators, albeit this time with its sign remaining positive regardless of the direction of the statement. Under these assumptions, acquiescence will 'cancel out' once the indicators are aggregated.

However, when the scale is unbalanced, acquiescence bias will shift the scale in the direction in which the scale is unbalanced. This is because where before the equal number of indicators in opposite directions 'cancelled' out the acquiescence in one another [see Cloud and Vaughan, 1970, Ray, 1979, Evans and Heath, 1995], in the unbalanced case there is leftover acquiescence. The more unbalanced the scale, the more bias leftover. This carries both descriptive and causal implications. Descriptively, this bias will shift the mean of the resultant scale in the direction of the imbalance. If we have a Likert scale comprised of more left-wing indicators than right-wing, then the resultant mean will be further to the left than it would be on a comparable balanced scale. Causally, on the same scale since acquiescence will point in the left-wing direction, variables that causally contribute to a respondent's level of acquiescence will appear to contribute to the scale. This can result in spurious causal associations (if the effect on the concept is 0), inflated causal associations (in the effect on the concept and acquiescence are in the same direction), or hidden causal associations (if the effect on the concept and acquiescence are in opposite directions).

2.3 Fully Unbalanced Likert Scales

A case not typically tackled in the literature on capturing acquiescence bias but which often arises in practice (both in political science and the broader social sciences) is that of a fully unbalanced Likert scale. Based on (2) and the subsequent discussion, it should become clear that in this case the acquiescence factor is impossible to *empirically* identify[1]. This is because at this point it becomes impossible to know which survey respondents are agreeing with the given statements because they sincerely agree with them; and which survey respondents are agreeing with them because they are acquiescent. Empirical identification of a model in this format requires *contradiction* in responses, which does not exist in a fully unbalanced scale.

It is arguably the case that this simple fact has led to some researchers mistakenly arguing that acquiescence bias is not a good explanation for the kind of results described above. For instance, Rodebaugh et al. [2007] find that removing reverse-scored (i.e. opposite) items improves the psychometric performance of their model; and argue that in fact these items were introducing an additional factor. I do not dispute their second claim, but instead point to the above: that an absence of contradiction in responses means acquiescence bias will become difficult if not impossible to directly identify. That the psychometric performance of the model is better after removing the contradicting items should be unsurprising: acquiescence is well-known among other things to be associated with inflated reliability coefficients and correlations [Winkler et al., 1982, Evans and Heath, 1995].

A separate but related line of argument argues that negatively worded items

[1] I use this term to distinguish from statistical identification of the model. The model may be statistically identified (i.e. a unique solution exists), but that is no reason to believe we have successfully captured the acquiescence component

are responsible for the item misresponse, rather than acquiescence bias [Swain et al., 2008]. The argument is that negatively worded items introduce additional cognitive complexity [Swain et al., 2008]. This point is not straightforwardly wrong: acquiescence bias is not the only potential problem that can occur in the survey design stage. The need for items on both sides of a scale does not mean that a simple negation route should be taken. However, as the demonstration below makes clear - acquiescence bias can still be observed in cases without negation. Researchers must therefore be prepared to tackle acquiescence bias within scales they are using.

In this paper, I tackle the specific problem of fully unbalanced Likert scales and offer some solutions for researchers utilising historical data.

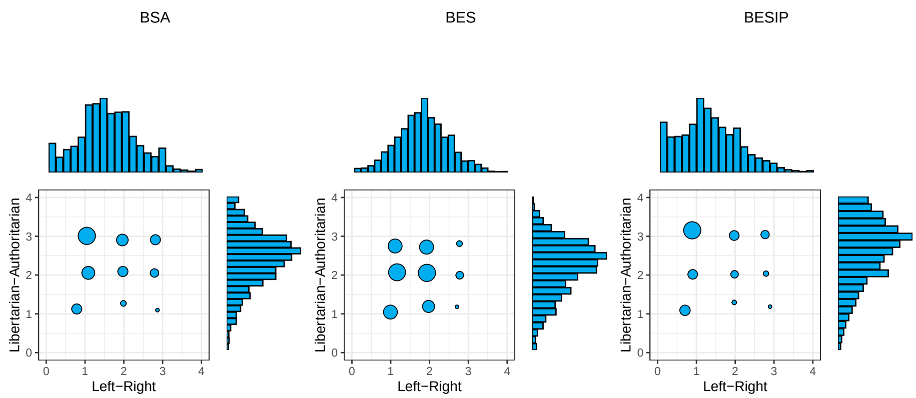
3 Case Selection and Datasets

I use public opinion datasets from Great Britain as my case study. This is because since the 1990s, almost all GB datasets contain 'left-right' and 'libertarian-authoritarian' Likert scales based on the work of Evans and Heath [Heath et al., 1994, Evans and Heath, 1995, Evans et al., 1996, see]. Several operationalisations of the same core concept therefore exist, creating an opportunity to assess how variations in measurement produce variations in research results. I use three datasets - two for demonstration, and one for developing a correction. For demonstration, I use the British Social Attitudes survey (BSA) [NatCen-Social-Research, 2017] and the British Election Study (BES) face-to-face survey [Fieldhouse et al., 2017]. Both surveys were collected in almost entirely the same time periods. Given this and the shared conceptual basis of their Likert scales, it is not unreasonable to expect reasonably similar distributions of attitudes in both surveys. Notably however, in the BSA, all items are left-wing or authoritarian. Insofar as acquiescence bias affects these scales, they should have a left-wing and authoritarian bias. The third dataset used is wave 14 of British Election Study internet panel (BESIP) [Fieldhouse et al., 2020], which is used as a cross-sectional dataset. Similar to the BSA, all items are worded in left-wing and authoritarian directions and it should therefore display a similar bias. Item wordings are available in appendix A.

Figure 1 presents both the joint and marginal distributions of the Likert scales from all three datasets. The BSA, BES, and BESIP scales are presented from left to right. The scales were constructed to range from 0 to 4. On each x-axis is the left-right scale and on each y-axis is the libertarian-authoritarian scale. The histograms opposite each axis capture the marginal distributions of these scales. To visualise the join distribution of the scales, respondents were divided into 'groups'. Those with scores ranging from 0 to 1.6 were placed in the 'left' and 'libertarian' groups of the respective dimensions. Correspondingly, those with scores ranging from 2.4 to 4 were placed in the 'right' and 'authoritarian' groups of the respective dimensions. Finally, those in-between these values were placed in the 'centre' group for each dimension. These groupings are of course arbitrary, but were chosen in part to resemble similar groupings utilised in research using these scales [see Surridge, 2018]. The mean of each group was plotted, while the size of the group's dot corresponds to the number of respondents in that group. Survey weights were used for the graphs.

Several notable differences emerge between the scales in figure 1. In line with the above predictions BSA and BESIP scales show clear left and authoritarian slants as compared to the BES. Indeed, the similarities between the BSA and BESIP plots are striking, both in terms of the marginal and joint distributions of the scales. By contrast, the BES plot shows both less left-wing and less authoritarian respondents. It retains a left-wing slant, but this is driven by the absence of right-wing respondents - it is still more balanced towards the

Figure 1: Joint Distribution of BSA and BES Scales



center of its scale relative to the BSA and BESIP plots. While the BES data therefore would offer firmer grounds for believing that the British electorate in 2017 was left-wing, some caution is still required. First, this may plausibly be a quirk of the sample in question. Second, it may be a function of the statements used to construct the Likert scale. Per (1), smaller loadings in the right-wing statements could produce a skewed result. Nonetheless, it is better evidence than that available in either of the other scales.

4 Demonstration

The impact of acquiescence bias on descriptive inference is straightforwardly demonstrated by the above graphs. However, its impact also extends to explanatory research. It is well-established that acquiescence has a strong negative relationship with education level [Ware Jr, 1978, Winkler et al., 1982], and so I take this as my example. Given the similar collection dates and conceptual overlap of the BSA and BES, I regress the scales contained within on a measure of education level. Since education also has some well-established results showing it has a negative relationship with authoritarianism [see Stubager, 2008, Surridge, 2016] but no well-established association with left-right attitudes, some predictions can be made. First, in the BES scales the results will be as described here. Second, in the BSA scale, a spurious positive association[2] between education level and left-right attitudes will be observed, while the negative association[3] between education level and libertarian-authoritarian attitudes will be stronger.

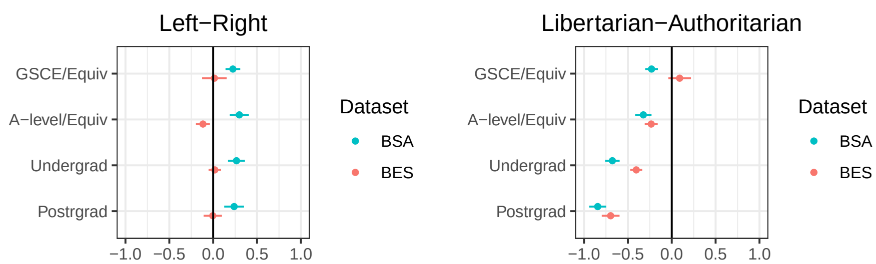
My interest here is not in offering some causal explanation of these scales, but rather to offer a clear example of results changing from unbalanced to balanced scales. I therefore do not include control variables as they do not add anything for the purposes of the demonstration. The education variables in both surveys were recoded such that the categories would match (full details of the recodes are in Appendix A). Most of the recodes will be uncontroversial and thus I do not discuss them further here, but the 'foreign' category in the BSA had to be treated as missing as it had no clear placement. Figure 2 shows the results of regressing the scales from the BSA and the BES on the recoded education variable. The left coefficient plot shows the results for the left-right scale, while the right coefficient plot shows the results for the libertarian-authoritarian scale. The reference category is possessing no education. 95% confidence intervals are included for each estimate. A full table of regression results is available in appendix B.

Figure 2 shows that the pattern of differences between the BSA and BES is as we'd expect given the expectations laid out above. First and least dramatically, the absolute size of the point estimates for the libertarian-authoritarian results

[2]Since the scale ranges from left (negative) to right (positive) and acquiescence points in the negative direction

[3]Since the scale ranges from libertarian (negative) to authoritarian (positive) and acquiescence points in the positive direction

Figure 2: Coefficient Plot of Demonstration Regressions



are larger in the BSA results. Moreover, two of the confidence intervals for BES coefficients (GSCE/Equiv and Undergrads) have no overlap with those of the BSA, indicating that they are significantly different from one another at the 95% confidence level. By contrast, the decision to use the BSA or BES dataset carries profound consequences for the results a researcher will find. The parameters for the BSA are all significant at the 95% confidence level and positive. By contrast, the parameters for the BES are not significant at the 95% confidence level with the exception of the A-levels parameter, which is negative. Only the parameters for GSCEs have overlapping 95% confidence intervals. The point of these results is not to offer some causal interpretation, but rather to highlight how scale construction can be the primary driver of research results.

All of the results demonstrated in this section have rely on an assumption that there should be no predictable differences from the BES to the BSA other than those caused by acquiescence. Given the importance of this assumption to my analysis above, I have performed two robustness checks to verify that these differences are likely driven by acquiescence bias, rather than any particular quirks of the samples.

First, I merged the five indicators common to the BES and BSA into a single dataset and created a binary variable denoting whether a respondent belonged to the BSA. I then regressed this binary variable on the five common indicators, and ran an OLS, Logit, and Probit model to check against model dependency. In all three cases, two indicators were significant[4]. However, their point estimates pointed in opposite directions, strongly suggesting that these were sample quirks most likely related to the unique componenets of these indicators. Certainly, it did not offer evidence against any of the above interpretation.

Next, to verify that there was no temporal instability in results, I regressed the scales from the BES on the survey month of the respondents. The result showed no association between interview month and scale score. I did not do the same for the BSA as the interview date is not included in the publicly available version of the dataset. Taken together, these two checks offer strong evidence that my assumption that differences between these two surveys are primarily driven by acquiescence bias is correct. Regression tables for both of these checks

[4]For some crimes, the death penalty is the most appropriate sentence; People who break the law should be given stiffer sentences

are available in appendix B.

5 Methodology

I now turn to the primary task of this paper, which is developing a methodology for the case of fully unbalanced Likert scales. Past research reviewing competing methodologies for modelling acquiescence bias have concluded that one of the most effective is an approach that treats acquiescence as a person-specific intercept across the scale items [Savalei and Falk, 2014, Primi et al., 2019b,a]. This model was developed in a Confirmatory Factor Analysis (CFA)/Structural Equation Modelling (SEM) context [Mirowsky and Ross, 1991, Billiet and McClendon, 2000] but later extended to a unidimensional Item Response Theory (IRT) context [Primi et al., 2019b,a]. For the sake of simplicity I focus on the CFA specification in this paper, but the general intuition translates to an IRT context. Here, I only briefly discuss the person intercept CFA model. A more complete background to CFA and its extension to include the person intercept is given in appendix C.

5.1 Person Intercept CFA

First used in Mirowsky and Ross' paper *Eliminating Defense and Agreement Bias from Measures of the Sense of Control: A 2 x 2 Index* [1991], the best exposition of the unit-intercept model is in Maydeu-Olivares and Coffman's paper *Random Intercept Item Factor Analysis* [2006]. The model is based on (2) and traditionally is estimated by setting λ_{ja} to 1. This essentially treats acquiescence bias as a form of *differential person functioning*, where there is a constant difference *between* respondents but not *within* them as to how they respond to the survey items. This assumption is therefore kept from the balanced Likert scales, but the assumption that each item equally captures the concept of interest is relaxed. Although the assumption of equal loadings for the acquiescence component is a strong one, simulations do suggest that the model is robust to violations of this assumption [Savalei and Falk, 2014]. Similarly, since the model is being estimated, the unique variation is also stripped from each item - a further relaxation relative to the balanced Likert scale. To identify the scales, the variance of η_{ic} is constrained to 1 while the variance of η_{ia} is freely estimated, producing the following model [Maydeu-Olivares and Coffman, 2006]:

$$x_{ij} = \lambda_{jc}\eta_{ic} + 1\eta_{ia} + \epsilon_{ij} \tag{3}$$

For the purposes of this paper I label this version of unit intercept CFA as CFA1. An alternative specification can be achieved by constraining the variances of both η_{ic} and the η_{ia} to 1, while freely estimating their loadings. However, a constraint is still placed on λ_{ja}, in that it must be equal across indicators. The linear form of this version of the model can thus be given as:

$$x_{ij} = \lambda_{jc}\eta_{ic} + \lambda_{a}\eta_{ia} + \epsilon_{ij} \tag{4}$$

For the purposes of this paper I label this version of unit intercept CFA as CFA2. The full set of assumptions for both CFA in general and unit intercept CFA are given in appendix C of this paper. However, one crucial difference in my definition of the model to Maydeu-Olivares and Coffman's is that I drop the language of 'random intercepts'. Here, they are drawing a parallel with hierarchical regression modelling in their description of the unit intercept. However, the comparison is not necessary and more importantly undermines the utility of the model. In a random-intercepts regression model, the random intercepts are estimated as an error component. The unit-intercept here is not being estimated as an error term - it is being estimated as another common factor.

The orthogonality assumption is thus not required for identification purposes (as other assumptions in the model are), but rather is made for the purpose of this comparison. This unnecessarily confuses things and potentially reduces the desirability of the model. In their review, Salvei and Falk suggest more work is required to explore potential relaxations of the orthogonality assumption. This assumption however is unnecessary to begin with, and I therefore drop it and utilise the terminology person-intercept instead of random-intercept.

In theory, the main difference between the specifications in (3) and (4) is their interpretability. Since the variances of both factors are the same in (4), the main advantage is that the model allows more direct comparison of the respective loadings - it is immediately clear how acquiescence bias compares to the content factors of interest in its effect on the scales. An advantage of the person intercept approach in general is that it does not require a balanced scale to work. Instead, the unit intercept merely acts to capture inconsistency in observed responses and thus in theory only requires at least one opposite-worded indicator in order to successfully capture acquiescence bias. I also consider ordinal versions of CFA1 and CFA2 in this paper, and I label them as OCFA1 and OCFA2 respectively. Their specifications are also detailed in appendix C.

5.2 Fully Unbalanced Scales

Past simulation studies suggest that unit-intercept models are robust to unbalanced scales where other acquiescence-correction methods require balanced scales [Savalei and Falk, 2014]. However, the crucial point made above is that the unit-intercept requires *contradiction* in order to empirically identify the acquiescence component, which is lacking in fully unbalanced scales. If for instance we take the BSA left-right scale, it is impossible to try and tell apart those who are agreeing with left-wing statements because they agree with them and those who are agreeing with the same statements because they are acquiescent. *There is no information available to distinguish the two kinds of agreement.*

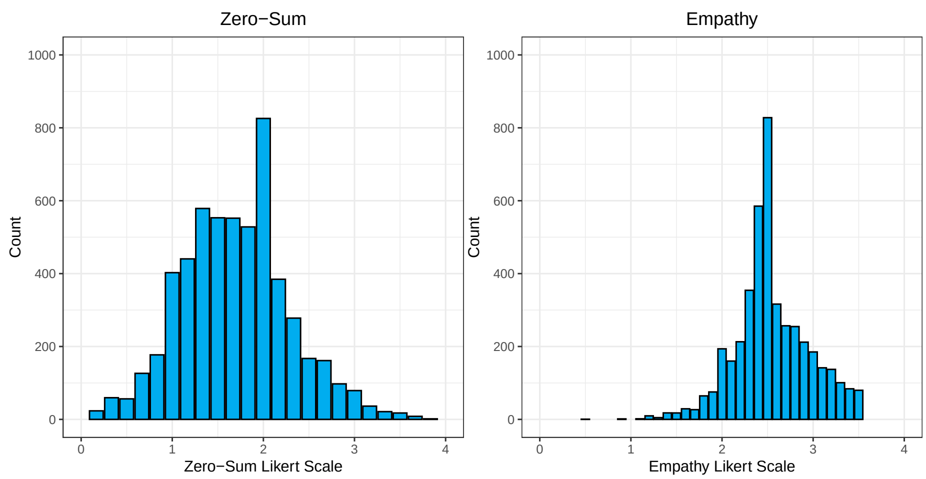
To solve this problem and empirically identify the acquiescence component, I use Watson's idea of introducing further information in the model [1992]. Specifically, if a scale which contains statements for which it would be contradictory to agree to all of them, it can be used to identify the acquiescence component in itself and thus also in the fully unbalanced scale. It is unfortunate that a strategy does not exist based on the fully unbalanced scale alone, but it should be clear that it is not possible to identify acquiescence bias in such a scale *without additional information being introduced in some form.* While the same simulations suggest that the unit intercept CFA model is robust to differing levels of acquiescence bias in each indicator [Savalei and Falk, 2014], this approach necessarily strengthens the assumption, as it assumes not only the same level of acquiescence for each respondent on one scale, but on all scales in the model.

5.3 Identifying Scales in BESIP

The reason I chose the fourteenth wave of BESIP is that it contains two balanced Likert scales which could be used to identify the acquiescence component in the manner described above. This is the May 2018 wave of BESIP and thus some comparability to the other two surveys in this paper is lost. However, as seen in 1 there is nonetheless enough similarity in the scales in BESIP and the BSA for the dataset to be suitable for my purposes. The two additional scales cover zero-sum approaches to life and second on personal empathy respectively. The individual item wordings are available in appendix A. The two scales were asked in two separate subsamples of BESIP wave 14. This creates two separate opportunities to test the model, and so I test the four model types across the two BESIP subsamples. Figure 3 shows bar plots of the two balanced scales.

The zero-sum scale ranges from everyone can win (0) to zero-sum (4), while the empathy scale runs from unempathetic (0) to empathetic (4). After filtering for missing data, the zero-sum subsample has 5836 respondents while the empathy subsample has 4478 respondents.

Figure 3: Bar Plots of the Balanced Scales



The empathy scale in figure 3 is notably less dispersed than the other scales discussed in this paper. This may be a function of the fact that it is comprised of a higher number of indicators than any of the others (10, as opposed to 5 or 6). It may also be however that given individuals are generally predisposed to view themselves as empathetic that there is less noise - and overall acquiescence - in the empathy scale. To test this second point, I ran person intercept CFA models on each of these scales alone, the full results for which are available in appendix D. The estimated variance for the acquiescence component in the zero-sum was larger than in the empathy model, suggesting that there is less acquiescence in the empathy scale. The extent to which the corrections are successful likely depends in part on which of these scales is a closer match in terms of acquiescence to the acquiescence in the left-right and libertarian-authoritarian scales. I therefore test all four variations of the person intercept on both subsamples of BESIP wave 14, using the respective additional scales to empirically identify the model.

5.4 Estimation

To estimate the CFA1 and CFA2 models, I use robust maximum likelihood (MLR) estimation. MLR returns the same point estimates as ML estimation but adjusts standard errors and test statistics for violations of the normality assumption. A rough rule of thumb suggests it's a reasonable approximation once at least 5 response categories exist. To estimate the OCFA1 and OCFA2 models, I use unweighed least squares estimation (ULS). In a comparison between MLR and diagonally weighted least squares (DWLS) estimation found in favour in DWLS for ordinal data [Li, 2016]. However, simulations comparing DWLS to ULS have in turn found in favour of ULS, with the caveat that DWLS may converge in situations where ULS does not [Forero et al., 2009]. I therefore utilise ULS estimation. All CFA models in this paper were estimated using lavaan version 0.6-9 [Rosseel, 2012] using code adapted from the appendix of Savalei et al [2014]. Since lavaan does not currently support survey weights for ordinal CFA models I have not used them in the CFA models themselves, but they were used in producing distributions from the predicted factor scores of the models. Since the associations between variables should be reasonably robust to weighting, this is likely unproblematic.

6 Results

In this section I present a series of results demonstrating the comparative performance of the methods. Since my emphasis as throughout the paper is on obtaining corrected measurements, the plots presented here pertain to the predicted factor scores. Tables containing results for the CFA models can be found in appendix D. To verify that the scales were broadly capturing the same content, their correlation matrices were checked. These tables are also available in appendix D.

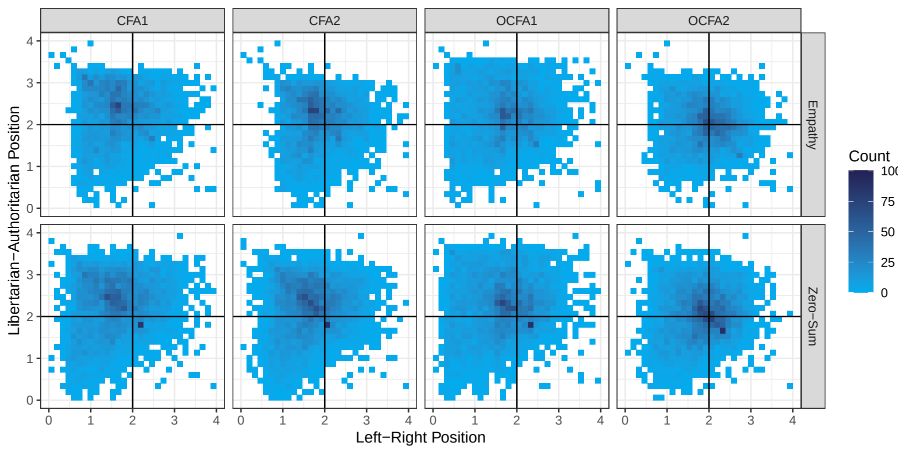
6.1 Distributions

Since my demonstration is preceded by distributional differences, I also begin my presentation of correction results with the distributions of the predicted factor scores. Figure 4 shows two-dimensional binplots of the resultant measurements from the four variations and the two subsets of the BESIP dataset. The extracted measures were rescaled to range from 0 to 4 to facilitate comparability with one another[5]. The 'left' factor extracted was flipped to range from left to right, rather than right to left. Libertarian-authoritarian factors are on the y-axis, and left-right factors are on the x-axis. The colour of the bins change from light blue to dark blue as the count of respondents in that bin increases. The plots are organised in columns for correction method, and by row for the two BESIP subsets. Plots of the marginal distributions of the predicted factor scores can be viewed in appendix D.

Alongside the marginal distributions available in appendix D, figure 4 shows that for all correction methods, relative to figure 1 there is a shift towards a more normal distribution. The scale is broadly more evenly distributed (especially in the case of OCFA2). There is a starker effect for the left-right scale, which in some cases appears to remain somewhat left-leaning. These differences are carried into the association between the two scales. The extent to which the joint

[5] An identifying constraint on the scales is that they are mean 0. However, this won't necessarily be a meaningful midpoint, especially if the distributions are skewed. Rescaling in this way establishes the midrange point as the central point of each scale, which is no less arbitrary in theory but in practice may be a better approximation to a 'true' midpoint

Figure 4: Two-Dimensional Bin Plot of Voter Beliefs



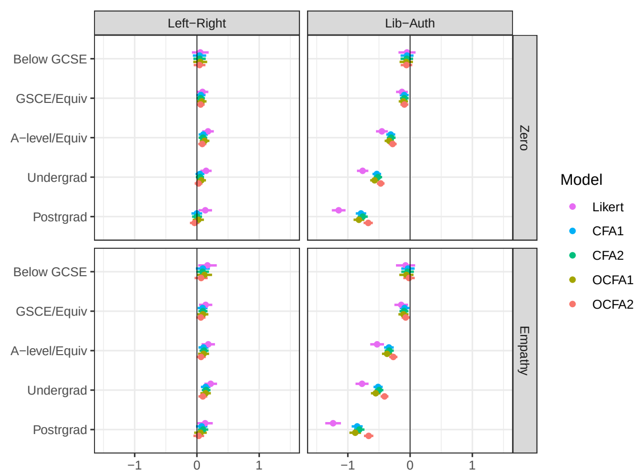
distribution is even across the four quadrants varies from correction method to correction method, but in all cases similarly appears more evenly distributed than in figure 1. These results would therefore indicate that once acquiescence bias is accounted for, the distribution of voter ideology on both scales is closer to a normal distribution. Caution is required in interpreting the midpoint of these scales, but nonetheless the extract factor scores do appear more evenly distributed.

6.2 Regression Results

With the distributional results of the correction methods established, I now turn to examining how explanatory research results are changed by using the predicted factor scores. I regressed the raw Likert scales and the predicted factor scores in each subsample on the education level variable available in the BESIP dataset. Once again, the reference category for the education level variable is 'no qualifications'. 95% confidence intervals are included in the plot. Figure

5 shows coefficient plots for each of these regression results. Tables for each regression are available in appendix D. For the results displayed in the main body of the paper, I have avoided recoding the education level variable as this recoding makes some results appear to be somewhat better than they are. I have however included regression results with the recoded education level variable in appendix D.

Figure 5: Coefficient Plots of Scales Regressed on Education



In the case of the results for the zero-sum left-right scales, the confidence intervals for each model fully overlap in both datasets. However, the point estimates shift towards 0 once correction methods are used; and more improtantly inferential differences emerge. If a correction method is used, it becomes the case that a researcher using null hypothesis significance testing will reach the same conclusions using the BESIP data as they would using the balanced scales in the BES. The correction is sharper in the zero-sum subsample, as predicted by the differences in acquiescence between the zero-sum and empathy scales. However, the OCFA2 model sufficiently shifts the point estimates in both subsamples that the same inferences will be produced in BESIP regardless of the subsample of choice. In terms of the libertarian-authoritarian scales, the gap in point estimates are considerably larger. In several cases, the 95% confidence intervals do not overlap at all. However, in line with the differences between the BES and BSA, a researcher using null hypothesis significance testing will reach the same inferential results - albeit with smaller point estimates for the corrected scales. The correction methods therefore broadly produce the same inferences as the balanced scales in the BES, while utilising biased data as in the BSA.

7 Conclusion

In this paper, I have set out the specific problem of fully unbalanced Likert scales in the context of wider work on acquiescence bias. Fully unbalanced Likert scales carry the particular problem of rendering the acquiescence within impossible to empirically identify without the introduction of additional information. In this paper, I have further clarified the different versions of person intercept CFA relative to Likert scales, relaxed the unnecessary orthogonality assumption, and developed a strategy for identifying a person intercept CFA model in the case of fully unbalanced Likert scales. The OCFA2 approach appears to work best for fully unbalanced scales, but it is not immediately clear why this should be the case. Researchers using these approaches should run all four and compare the results until further research can be conducted on the relative performance of the four methods.

A clear limitation of the correction methods used in this paper is the data requirements they impose on the user in the case of fully unbalanced scales. However, these limtis are necessary: acquiescence bias is not empirically identifiable without some degree of contradiction. Where researchers cannot utilise the corrections, they should at minimum be conscious of the role that acquiescence bias is likely to be playing in their results. Even where Likert scales are fully balanced, their use entails strong assumptions about the data generating process that can be relaxed by the use of person intercept CFA. There is no clear case where the use of these models if possible is not preferable to a raw likert scale. For survey designers, two main points should be taken from this paper. First, as far as possible they should seek to design Likert scales that are fully balanced. Where this is not entirely possible, whether due to difficulties in designing reverse-keyed items or the need for backwards compatability, they should instead try to include other, substantively unrelated scales for the purposes of identifying person intercept CFA models.

References

Jaak B Billiet and Eldad Davidov. Testing the stability of an acquiescence style factor behind two interrelated substantive variables in a panel design. *Sociological Methods & Research*, 36(4):542–562, 2008.

Jaak B Billiet and McKee J McClendon. Modeling acquiescence in measurement models for two balanced sets of items. *Structural equation modeling*, 7(4):608–628, 2000.

Timothy A Brown. *Confirmatory factor analysis for applied research.* Guilford publications, 2015.

Jonathan Cloud and Graham M Vaughan. Using balanced scales to control acquiescence. *Sociometry*, pages 193–202, 1970.

Geoffrey Evans, Anthony Heath, and Mansur Lalljee. Measuring left-right and libertarian-authoritarian values in the british electorate. *British Journal of Sociology*, pages 93–112, 1996.

Geoffrey A Evans and Anthony F Heath. The measurement of left-right and libertarian-authoritarian values: A comparison of balanced and unbalanced scales. *Quality and Quantity*, 29(2):191–206, 1995.

E. Fieldhouse, J. Green, G. Evans, H. Schmitt, C. van der Eijk, J. Mellon, and C. Prosser. British election study 2017 face-to-face post-election survey, 2017. URL `https://www.britishelectionstudy.com/data-objects/cross-sectional-data/`.

E. Fieldhouse, J. Green, G. Evans, J. Mellon, and C. Prosser. British election study internet panel, 2020. URL `https://www.britishelectionstudy.com/data-objects/panel-study-data/`.

Carlos G Forero, Alberto Maydeu-Olivares, and David Gallardo-Pujol. Factor analysis with ordinal indicators: A monte carlo study comparing dwls and uls estimation. *Structural Equation Modeling*, 16(4):625–641, 2009.

Anthony Heath, Geoffrey Evans, and Jean Martin. The measurement of core beliefs and values: The development of balanced socialist/laissez faire and libertarian/authoritarian scales. *British Journal of Political Science*, 24(1): 115–132, 1994.

David A Kenny and Deborah A Kashy. Analysis of the multitrait-multimethod matrix by confirmatory factor analysis. *Psychological Bulletin*, 112(1):165, 1992.

Cheng-Hsien Li. Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior research methods*, 48(3):936–949, 2016.

Albert Maydeu-Olivares and Donna L Coffman. Random intercept item factor analysis. *Psychological methods*, 11(4):344, 2006.

John Mirowsky and Catherine E Ross. Eliminating defense and agreement bias from measures of the sense of control: A 2 x 2 index. *Social Psychology Quarterly*, pages 127–145, 1991.

NatCen-Social-Research. British social attitudes survey, 2017, 2017. URL `https://www.ukdataservice.ac.uk/`.

Ricardo Primi, Nelson Hauck-Filho, Felipe Valentini, Daniel Santos, and Carl F Falk. Controlling acquiescence bias with multidimensional irt modeling. In *The Annual Meeting of the Psychometric Society*, pages 39–52. Springer, 2019a.

Ricardo Primi, Daniel Santos, Filip De Fruyt, and Oliver P John. Comparison of classical and modern methods for measuring and correcting for acquiescence. *British Journal of Mathematical and Statistical Psychology*, 72(3):447–465, 2019b.

John J Ray. Is the acquiescent response style problem not so mythical after all? some results from a successful balanced f scale. *Journal of Personality Assessment*, 43(6):638–643, 1979.

Thomas L Rodebaugh, Carol M Woods, and Richard G Heimberg. The reverse of social anxiety is not always the opposite: The reverse-scored items of the social interaction anxiety scale do not belong. *Behavior Therapy*, 38(2):192–206, 2007.

Yves Rosseel. Lavaan: An r package for structural equation modeling and more. version 0.5–12 (beta). *Journal of statistical software*, 48(2):1–36, 2012.

Victoria Savalei and Carl F Falk. Recovering substantive factor loadings in the presence of acquiescence bias: A comparison of three approaches. *Multivariate behavioral research*, 49(5):407–424, 2014.

Rune Stubager. Education effects on authoritarian–libertarian values: a question of socialization1. *The British journal of sociology*, 59(2):327–350, 2008.

Paula Surridge. Education and liberalism: pursuing the link. *Oxford Review of Education*, 42(2):146–164, 2016.

Paula Surridge. The fragmentation of the electoral left since 2010. *Renewal: a Journal of Labour Politics*, 26(4):69–78, 2018.

Scott D Swain, Danny Weathers, and Ronald W Niedrich. Assessing three sources of misresponse to reversed likert items. *Journal of Marketing Research*, 45(1):116–131, 2008.

John E Ware Jr. Effects of acquiescent response set on patient satisfaction ratings. *Medical care*, pages 327–336, 1978.

Dorothy Watson. Correcting for acquiescent response bias in the absence of a balanced scale: An application to class consciousness. *Sociological Methods & Research*, 21(1):52–88, 1992.

John D Winkler, David E Kanouse, and John E Ware. Controlling for acquiescence response set in scale development. *Journal of Applied Psychology*, 67 (5):555–561, 1982.

Agree to Agree: Appendix

Appendix A: Variable Codings

For the BES, it is noted next to each item whether the statement is left-wing, right-wing, libertarian, or authoritarian in its direction. All of the scales where constructed to range from 0 (left/libertarian) to 4 (right/authoritarian). Since survey weights are used throughout unless otherwise stated, the BES respondents without survey weights were not included all parts of the following analysis.

BSA Likert Scales

The statements utilised in the BSA economic dimension (ranging from Disagree Strongly to Agree Strongly) are as follows:

- Government should redistribute income from the better off to those who are less well off
- Big business benefits owners at the expense of workers
- Ordinary working people do not get their fair share of the nation's wealth
- There is one law for the rich and one for the poor
- Management will always try to get the better of employees if it gets the chance

The statements utilised in the BSA second dimension (ranging from Disagree Strongly to Agree Strongly) are as follows:

- Young people today don't have enough respect for traditional British values
- People who break the law should be given stiffer sentences
- For some crimes, the death penalty is the most appropriate sentence
- Schools should teach children to obey authority
- The law should always be obeyed, even if a particular law is wrong
- Censorship of films and magazines is necessary to uphold moral standards

BES Likert Scales

The statements utilised in the BES economic dimension (ranging from Strongly Disagree to Strongly Agree) are as follows:

- Ordinary working people get their fair share of the nation's wealth (right)
- There is one law for the rich and one for the poor (left)
- There is no need for strong trade unions to protect employees' working conditions and wages (right)
- Private enterprise is the best way to solve Britain's economic problems (right)
- Major public services and industries ought to be in state ownership (left)
- It is the government's responsibility to provide a job for everyone who wants one (left)

The statements utilised in the BES second dimension (ranging from Strongly Disagree to Strongly Agree) are as follows:

- Young people today don't have enough respect for traditional British values (auth)
- Censorship of films and magazines is necessary to uphold moral standards (auth)
- People should be allowed to organise public meetings to protest against the government (lib)
- People in Britain should be more tolerant of those who lead unconventional lives (lib)
- For some crimes, the death penalty is the most appropriate sentence (auth)
- People who break the law should be given stiffer sentences (auth)

BESIP Likert Scales

- **lr1:** Government should redistribute income from the better off to those who are less well off
- **lr2:** Big business takes advantage of ordinary people
- **lr3:** Ordinary working people do not get teir fair share of the nation's wealth
- **lr4:** There is one law for the rich and one for the poor
- **lr5:** Management will always try to get the better of employees if it gets the chance

The wordings of the libertarian-authoritarian statements are:

- **al1:** Young people today don't have enough respect for traditional authority
- **al2:** For some crimes, the death penalty is the most appropriate sentence
- **al3:** Schools should teach children to obey authority
- **al4:** Censorship of films and magazines is necessary to uphold moral standards
- **al5:** People who break the law should be given stiffer sentences

BESIP Extra Likert Scales

The statements on the zero-sum scale are:

- **zero1:** One person's loss is another person's gain (zero-sum)
- **zero4:** There's only so much to go around. Life is about how big a slice of the pie you can get. (zero-sum)
- **zero5:** Life isn't about winners and losers, everyone can do well (everyone can win)
- **zero7:** The only way to make someone better off is to make someone else worse off (zero-sum)
- **zero9:** There are ways to make everyone better off without anyone losing out (everyone can win)
- **zero11:** Everyone can be a winner at the same time (everyone can win)

The statements from the empathy scale are:

- **empathy1:** I can usually figure out when my friends are scared (empathetic)

- **empathy2:** I can usually realize quickly when a friend is angry (empathetic)
- **empathy3:** I can usually figure out when people are cheerful (empathetic)
- **empathy4:** I am not usually aware of my friends' feelings (unempathetic)
- **empathy5:** When someone is feeling 'down' I can usually understand how they feel (empathetic)
- **empathy6:** After being with a friend who is sad about something, I usually feel sad (empathetic)
- **empathy7:** My friends' unhappiness doesn't make me feel anything (unempathetic)
- **empathy8:** Other people's feelings don't bother me at all (unempathetic)
- **empathy9:** I don't become sad when I see other people crying (unempathetic)
- **empathy10:** My friends' emotions don't affect me much (unempathetic)

Education Recodes

Table A1: BSA Education Recode

Original Coding	New Coding
Postgraduate degree	Postgrad
First degree	Undergrad
Higher educ below degree	A-level/equiv
A level or equiv	A-level/equiv
O level or equiv	GCSE/equiv
CSE or equiv	GCSE/equiv
Foreign or other	Missing
No qualification	No Qualification

Table A2: BES Education Recode

Original Coding	New Coding
No qualifications	No qualification
Below GCSE	No qualification
GCSE	GCSE/equiv
A-level	A-level/equiv
Undergraduate	Undergrad
Postgrad	Postgrad

Appendix B: Demonstration

Regression results

Table B1: BSA and BES Scales Regressed on Education

	BSA Left-Right	BES Left-Right	BSA Lib-Auth	BES Lib-Auth
Intercept	1.31***	1.65***	2.82***	2.26***
	(0.04)	(0.02)	(0.03)	(0.02)
GCSE/Equiv	0.22***	0.01	−0.23***	0.09
	(0.04)	(0.07)	(0.04)	(0.07)
A-level/Equiv	0.30***	−0.12**	−0.32***	−0.23***
	(0.06)	(0.04)	(0.05)	(0.04)
Undergrad	0.27***	0.02	−0.68***	−0.40***
	(0.05)	(0.04)	(0.04)	(0.03)
Postgrad	0.24***	−0.00	−0.84***	−0.70***
	(0.06)	(0.05)	(0.05)	(0.05)
R^2	0.01	0.01	0.13	0.12
Adj. R^2	0.01	0.00	0.13	0.12
Num. obs.	3123	1806	3125	1931

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Demonstration Robustness

Indicators common to both datasets:

- **Ind1:** There is one law for the rich and one for the poor
- **Ind2:** Young people today don't have enough respect for traditional British values
- **Ind3:** Censorship of films and magazines is necessary to uphold moral standards
- **Ind4:** For some crimes, the death penalty is the most appropriate sentence
- **Ind5:** People who break the law should be given stiffer sentences

Table B2: Regression of Survey Membership on Common Indicators

	OLS	Logit	Probit
Intercept	0.56	0.25	0.16
	(0.03)	(0.11)	(0.07)
Ind1	0.01	0.05	0.03
	(0.01)	(0.03)	(0.02)
Ind2	−0.00	−0.02	−0.01
	(0.01)	(0.03)	(0.02)
Ind3	−0.01	−0.05	−0.03
	(0.01)	(0.03)	(0.02)
Ind4	−0.02	−0.06	−0.04
	(0.01)	(0.02)	(0.01)
Ind5	0.04	0.16	0.10
	(0.01)	(0.04)	(0.02)
R^2	0.01		
Adj. R^2	0.00		
Num. obs.	5170	5170	5170
AIC		7039.76	7040.06
BIC		7079.07	7079.37
Log Likelihood		−3513.88	−3514.03
Deviance		6841.30	6841.57

Table B3: Regression of Scales on Survey Month

	Left-Right	Lib-Auth
Intercept	1.64	2.04
	(0.02)	(0.02)
Aug	−0.03	0.00
	(0.03)	(0.03)
Sep	−0.04	0.04
	(0.04)	(0.05)
R^2	0.00	0.00
Adj. R^2	−0.00	−0.00
Num. obs.	1789	1914

Appendix C: Unit Intercept Confirmatory Factor Analysis

The standard confirmatory factor analysis model is given in its linear form as:

$$x_{ij} = \lambda_{j1}\eta_{i1} + ... + \lambda_{jm}\eta_{im} + \epsilon_{ij} \qquad \text{(?? revisited)}$$

Which is the common factor model discussed in the main body of the paper. The assumptions of this model are:

1. The means of the common factors are 0
2. The common factors are normally distributed
3. The means of the unique components are 0
4. The unique components are normally distributed
5. The unique components are uncorrelated with the common factors
6. The unique components are uncorrelated with each other

The model can be expressed in a more compact matrix form:

$$\mathbf{x} = \boldsymbol{\Lambda}\boldsymbol{\eta} + \boldsymbol{\epsilon} \tag{5}$$

Where $\mathbf{x}$ is the $p \times 1$ vector of indicators, $\boldsymbol{\Lambda}$ is the $p \times m$ matrix of factor loadings, $\boldsymbol{\eta}$ is the $m \times 1$ vector of factor scores, and $\boldsymbol{\epsilon}$ is the $p \times 1$ vector of unique components. In turn, we can further express the model in terms of covariance matrices:

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{\Lambda}' + \boldsymbol{\Theta_{\epsilon}} \tag{6}$$

Where $\boldsymbol{\Sigma}$ is the $p \times p$ variance-covariance matrix of the indicators, $\boldsymbol{\psi}$ is the $m \times m$ variance-covariance matrix of the common factors, and $\boldsymbol{\Theta_{\epsilon}}$ is the $p \times p$ variance-covariance matrix of unique components which by assumption 6 is a diagonal matrix. When estimated with maximum likelihood (ML), assuming no (further) restrictions are placed on the latent variables means the discrepancy function minimised is:

$$F_{ML} = ln|\mathbf{S}| - ln|\boldsymbol{\Sigma}| + trace(\boldsymbol{S\Sigma^{-1}}) - p \tag{7}$$

Where $\mathbf{S}$ is the model-implied variance-covariance matrix and p is the number of indicators.

Person Intercept CFA

As discussed in the main body of the paper, unit intercept CFA is given by

$$x_{ij} = \lambda_{jc}\eta_{ic} + 1\eta_{ia} + \epsilon_{ij} \qquad \text{(\textbf{??} revisited)}$$

Where factor c would be the common factor and factor a would be the person intercept factor. Maydeu-Olivares and Coffman introduce three further assumptions for this model relative to regular CFA, which deserve discussion. The first two are:

7. The mean of the unit-intercepts is 0

8. The unit intercepts are uncorrelated with the unique components

Thus far, these are simply assumptions 1 and 5 repackaged for treating the unit-intercept factor separately. However, Maydeu-Olivares and Coffman make a further assumption:

9. The unit intercepts are uncorrelated with the common factor(s)

This assumption is explained in part by Maydeu-Olivares and COffman's choice of language for the model. As discussed in the main body of the paper, they specifically refer to the model as a *random-intercept* model and clearly are aiming to draw a parallel with multilevel regression modelling in their description of the unit-intercept confirmatory factor analysis model (indeed, their formulae reflect this too). However, as discussed in the main body of the paper, this comparison is not only unnecessary but arguably limits the utility of the model. I therefore drop this assumption and utilise the terminology person intercept instead.

To identify the scales of the common factors in the person intercept model, the variances of the common factors are constrained to 1 (as opposed to their first loading being constrained to 1). By contrast, the variance of the unit-intercept is freely estimated. The important feature of such a model is that the loading of the unit-intercept factor is constrained across indicators. A method of creating such an intercept while constraining the unit-intercept variance to 1 would simply be to apply equality constraints to the unit-intercept loadings, such that they were equal across all indicators:

$$x_{ij} = \lambda_{jc}\eta_{ic} + \lambda_a\eta_{ia} + \epsilon_{ij} \tag{?? revisited}$$

As stated in the main body of the paper, the difference between (**??**) and (**??**) is that instead of a loading of '1' on η_{ia}, there is now a freely estimated loading lacking a 'j' subscript as it is common to all indicators.

Ordinal Confirmatory Factor Analysis

One potential flaw of the person intercept CFA model is that it does not fully take into account the ordinal nature of the indicator variables typical for Likert scales. In ordinal CFA, the relationship between the latent variables and the observed categories are assumed to exist via a threshold relationship:

$$\begin{aligned} x^*_{ij} &= \lambda_{j1}\eta_{i1} + ... + \lambda_{jm}\eta_{im} + \epsilon_{ij} \\ x_{ij} &= K \quad if \quad \tau_{jk} < x^*_{ij} < \tau_{jk+1} \end{aligned} \tag{8}$$

Where x^*_{ij} is the latent variable underlying x_{ij}, K is one of the t values x_{ij} can take on, τ_{jk} is the kth threshold for indicator j, $\tau_{j0} = -\infty$ and $\tau_{jt} = \infty$.

Ordinal CFA makes similar assumptions to continuous CFA:

1. The means of the common factors are 0
2. The common factors are normally distributed
3. The means of the unique components are 0
4. The unique components are normally distributed
5. The unique components are uncorrelated with the common factors
6. The unique components are uncorrelated with each other

It follows that x^*_{ij} is normally distributed with mean 0 and the covariance matrix:

$$\mathbf{\Sigma} = \mathbf{\Lambda\Psi\Lambda}' + \mathbf{\Theta}_{\boldsymbol{\epsilon}} \tag{9}$$

To identify the variances of the unique components, we set

$$\mathbf{\Theta}_{\boldsymbol{\epsilon}} = \mathbf{I} - diag(\mathbf{\Lambda\Psi\Lambda}') \tag{10}$$

such that the covariance matrix becomes a correlation matrix $\mathbf{P}$.

Ordinal CFA is often estimated in a three-step procedure. First, the thresholds are estimated alone using maximum likelihood. The thresholds are often estimated by the corresponding percentage of respondents in each category of the ordinal variable. Second, the polychoric correlation matrix of the observed indicators is estimated via maximum likelihood. Third, assuming no restrictions are placed on the thresholds, a least squares discrepancy function based on the polychoric correlations can be used:

$$F_{LS} = (\hat{\boldsymbol{p}} - \boldsymbol{p(\theta)})'\boldsymbol{V}(\hat{\boldsymbol{p}} - \boldsymbol{p(\theta)}) \tag{11}$$

Where $\hat{p}$ is the polychoric correlation matrix estimated in the second step, $p(\theta)$ is the model-implied correlation matrix, $\boldsymbol{\theta}$ represents the parameters of the model, and $\boldsymbol{V}$ is a weighting matrix. The choice of weighting matrix determines the exact estimation method being used. If $\hat{\mathbf{\Gamma}}$ is an estimate of the asymptotic covariance matrix of estimated polychoric correlations, then:

- Weighted Least Squares (WLS): $\boldsymbol{V} = \hat{\mathbf{\Gamma}}$
- Diagonally Weighted Least Squares (DWLS): $\mathbf{V} = diag(\hat{\mathbf{\Gamma}})^{-1/2}$
- Unweighted Least Squares (ULS): $\boldsymbol{V} = \boldsymbol{I}$

Similarly to regular CFA, implementing the unit intercept in ordinal CFA is relatively straightforward. We can either set the loadings of the unit-intercept factor to 1 while freeing its variance:

$$x^*_{ij} = \lambda_{jc}\eta_{ic} + 1\eta_{ia} + \epsilon_{ij} \tag{12}$$

Or alternatively we can can constrain its variance to 1 while constraining the loadings to be equal but freely estimating their value:

$$x^*_{ij} = \lambda_{jc}\eta_{ic} + \lambda_a\eta_{ia} + \epsilon_{ij} \tag{13}$$

Continuing with the convention established above, for the remainder of this paper I refer to these models as (12) OCFA1 and (13) OCFA2.

Appendix D: Correction

Identifying Scale CFA

Table D1: Zero CFA Check

	Model	
	Estimate	Std. Err.
	Loadings	
Zero		
zero1	0.41	0.01
zero4	0.49	0.02
zero5	-0.53	0.02
zero7	0.61	0.01
zero9	-0.59	0.01
zero11	-0.58	0.02
Acq		
zero1	1.00^{+}	
zero4	1.00^{+}	
zero5	1.00^{+}	
zero7	1.00^{+}	
zero9	1.00^{+}	
zero11	1.00^{+}	
	Latent Variances	
Zero	1.00^{+}	
Acq	0.10	0.00
	Fit Indices	
χ^2(df)	253.63	
CFI	0.96	
TLI	0.93	
RMSEA	0.07	
Scaled χ^2(df)	181.16(8)	

$^{+}$Fixed parameter

CFA Results

Zero-Sum CFA Results

Table D3: Zero-Sum CFA1

	Model	
	Estimate	Std. Err.
	Loadings	
Z		
zero7	0.58	0.01

zero1	0.40	0.01
zero4	0.48	0.02
zero11	-0.59	0.02
zero5	-0.55	0.02
zero9	-0.60	0.01
<u>LeftCorrected</u>		
lr1	0.81	0.02
lr2	0.70	0.01
lr3	0.81	0.01
lr4	0.83	0.01
lr5	0.61	0.01
<u>AuthCorrected</u>		
al1	0.85	0.01
al2	0.99	0.02
al3	0.73	0.01
al4	0.56	0.02
al5	0.72	0.01
<u>Acq</u>		
zero7	1.00^{+}	
zero1	1.00^{+}	
zero4	1.00^{+}	
zero11	1.00^{+}	
zero5	1.00^{+}	
zero9	1.00^{+}	
lr1	1.00^{+}	
lr2	1.00^{+}	
lr3	1.00^{+}	

lr4	1.00+	
lr5	1.00+	
al1	1.00+	
al2	1.00+	
al3	1.00+	
al4	1.00+	
al5	1.00+	
	Latent Variances	
Z	1.00+	
LeftCorrected	1.00+	
AuthCorrected	1.00+	
Acq	0.08	0.00
	Fit Indices	
χ^2(df)	3134.95	
CFI	0.90	
TLI	0.89	
RMSEA	0.07	
Scaled χ^2(df)	2641.83(103)	

+Fixed parameter

Table D4: Zero-Sum CFA2

	Model	
	Estimate	Std. Err.
	Loadings	
Z		
zero7	0.67	0.05
zero1	0.50	0.05

zero4	0.60	0.05
zero11	-0.49	0.05
zero5	-0.44	0.05
zero9	-0.50	0.05
LeftCorrected		
lr1	0.83	0.02
lr2	0.74	0.02
lr3	0.85	0.01
lr4	0.87	0.02
lr5	0.65	0.02
AuthCorrected		
al1	0.87	0.02
al2	1.02	0.03
al3	0.75	0.02
al4	0.57	0.02
al5	0.74	0.02
Acq		
zero7	0.32	0.02
zero1	0.32	0.02
zero4	0.32	0.02
zero11	0.32	0.02
zero5	0.32	0.02
zero9	0.32	0.02
lr1	0.32	0.02
lr2	0.32	0.02
lr3	0.32	0.02
lr4	0.32	0.02

lr5	0.32	0.02
al1	0.32	0.02
al2	0.32	0.02
al3	0.32	0.02
al4	0.32	0.02
al5	0.32	0.02
	Latent Variances	
Z	1.00^{+}	
LeftCorrected	1.00^{+}	
AuthCorrected	1.00^{+}	
Acq	1.00^{+}	
	Fit Indices	
χ^2(df)	2705.09	
CFI	0.92	
TLI	0.90	
RMSEA	0.07	
Scaled χ^2(df)	2307.25(97)	

$^{+}$Fixed parameter

Table D5: Zero-Sum OCFA1

	Model	
	Estimate	Std. Err.
	Loadings	
Z		
zero7	0.70	0.01
zero1	0.45	0.01
zero4	0.52	0.01

zero11	-0.57	0.01
zero5	-0.59	0.01
zero9	-0.67	0.01
<u>LeftCorrected</u>		
lr1	0.67	0.01
lr2	0.81	0.01
lr3	0.83	0.01
lr4	0.81	0.01
lr5	0.67	0.01
<u>AuthCorrected</u>		
al1	0.80	0.01
al2	0.70	0.01
al3	0.75	0.01
al4	0.50	0.01
al5	0.79	0.01
<u>Acq</u>		
zero7	1.00^{+}	
zero1	1.00^{+}	
zero4	1.00^{+}	
zero11	1.00^{+}	
zero5	1.00^{+}	
zero9	1.00^{+}	
lr1	1.00^{+}	
lr2	1.00^{+}	
lr3	1.00^{+}	
lr4	1.00^{+}	
lr5	1.00^{+}	

al1	1.00^{+}	
al2	1.00^{+}	
al3	1.00^{+}	
al4	1.00^{+}	
al5	1.00^{+}	
	Latent Variances	
Z	1.00^{+}	
LeftCorrected	1.00^{+}	
AuthCorrected	1.00^{+}	
Acq	0.05	0.00
	Fit Indices	
χ^2(df)	6344.31	
CFI	0.90	
TLI	0.92	
RMSEA	0.08	
Scaled χ^2(df)	1855.01(167)	

$^{+}$Fixed parameter

Table D6: Zero-Sum OCFA2

	Model	
	Estimate	Std. Err.
	Loadings	
Z		
zero7	0.69	0.03
zero1	0.49	0.03
zero4	0.61	0.03
zero11	-0.54	0.03

zero5	-0.55	0.03
zero9	-0.70	0.03
LeftCorrected		
lr1	0.78	0.01
lr2	0.86	0.01
lr3	0.90	0.01
lr4	0.85	0.01
lr5	0.68	0.01
AuthCorrected		
al1	0.81	0.01
al2	0.74	0.01
al3	0.76	0.01
al4	0.49	0.01
al5	0.78	0.01
Acq		
zero7	0.35	0.01
zero1	0.35	0.01
zero4	0.35	0.01
zero11	0.35	0.01
zero5	0.35	0.01
zero9	0.35	0.01
lr1	0.35	0.01
lr2	0.35	0.01
lr3	0.35	0.01
lr4	0.35	0.01
lr5	0.35	0.01
al1	0.35	0.01

al2	0.35	0.01
al3	0.35	0.01
al4	0.35	0.01
al5	0.35	0.01
	Latent Variances	
Z	1.00^{+}	
LeftCorrected	1.00^{+}	
AuthCorrected	1.00^{+}	
Acq	1.00^{+}	
	Fit Indices	
χ^2(df)	3190.44	
CFI	0.95	
TLI	0.94	
RMSEA	0.07	
Scaled χ^2(df)	3933.42(97)	

$^{+}$Fixed parameter

Table D2: Empathy CFA Check

	Model	
	Estimate	Std. Err.
	Loadings	
Empathy		
em1	0.30	0.01
em2	0.32	0.01
em3	0.30	0.01
em4	-0.34	0.01
em5	0.29	0.01
em6	0.25	0.01
em7	-0.45	0.01
em8	-0.48	0.01
em9	-0.39	0.01
em10	-0.47	0.01
Acq		
em1	1.00^{+}	
em2	1.00^{+}	
em3	1.00^{+}	
em4	1.00^{+}	
em5	1.00^{+}	
em6	1.00^{+}	
em7	1.00^{+}	
em8	1.00^{+}	
em9	1.00^{+}	
em10	1.00^{+}	
	Latent Variances	
Empathy	1.00^{+}	
Acq	0.05	0.00
	Fit Indices	
χ^2(df)	2268.03	
CFI	0.87	
TLI	0.83	
RMSEA	0.12	
Scaled χ^2(df)	1513.06(34)	

$^{+}$Fixed parameter

Empathy CFA Results

Table D7: Empathy CFA1

	Model

	Estimate	Std. Err.
	Loadings	
E		
em1	0.29	0.01
em2	0.31	0.01
em3	0.29	0.01
em4	-0.34	0.01
em5	0.28	0.01
em6	0.25	0.01
em7	-0.46	0.01
em8	-0.49	0.01
em9	-0.40	0.01
em10	-0.48	0.01
LeftCorrected		
lr1	0.83	0.02
lr2	0.70	0.01
lr3	0.84	0.01
lr4	0.82	0.01
lr5	0.65	0.02
AuthCorrected		
al1	0.90	0.02
al2	1.03	0.02
al3	0.77	0.02
al4	0.63	0.02
al5	0.77	0.01
Acq		
em1	1.00^{+}	

em2	1.00^{+}	
em3	1.00^{+}	
em4	1.00^{+}	
em5	1.00^{+}	
em6	1.00^{+}	
em7	1.00^{+}	
em8	1.00^{+}	
em9	1.00^{+}	
em10	1.00^{+}	
lr1	1.00^{+}	
lr2	1.00^{+}	
lr3	1.00^{+}	
lr4	1.00^{+}	
lr5	1.00^{+}	
al1	1.00^{+}	
al2	1.00^{+}	
al3	1.00^{+}	
al4	1.00^{+}	
al5	1.00^{+}	
	Latent Variances	
E	1.00^{+}	
LeftCorrected	1.00^{+}	
AuthCorrected	1.00^{+}	
Acq	0.04	0.00
	Fit Indices	
χ^2(df)	4227.66	
CFI	0.89	

TLI	0.87	
RMSEA	0.07	
Scaled χ^2(df)	3470.77(169)	

$^+$Fixed parameter

Table D8: Empathy CFA2

	Model	
	Estimate	Std. Err.
	Loadings	
E		
em1	0.13	0.05
em2	0.16	0.05
em3	0.14	0.05
em4	-0.49	0.05
em5	0.13	0.05
em6	0.10	0.05
em7	-0.61	0.05
em8	-0.64	0.05
em9	-0.55	0.05
em10	-0.63	0.05
LeftCorrected		
lr1	0.85	0.02
lr2	0.72	0.02
lr3	0.85	0.01
lr4	0.83	0.02
lr5	0.67	0.02
AuthCorrected		

al1	0.92	0.02
al2	1.08	0.03
al3	0.79	0.02
al4	0.65	0.02
al5	0.80	0.02
<u>Acq</u>		
em1	0.27	0.03
em2	0.27	0.03
em3	0.27	0.03
em4	0.27	0.03
em5	0.27	0.03
em6	0.27	0.03
em7	0.27	0.03
em8	0.27	0.03
em9	0.27	0.03
em10	0.27	0.03
lr1	0.27	0.03
lr2	0.27	0.03
lr3	0.27	0.03
lr4	0.27	0.03
lr5	0.27	0.03
al1	0.27	0.03
al2	0.27	0.03
al3	0.27	0.03
al4	0.27	0.03
al5	0.27	0.03
	<u>Latent Variances</u>	

E	1.00^{+}
LeftCorrected	1.00^{+}
AuthCorrected	1.00^{+}
Acq	1.00^{+}
	Fit Indices
χ^2(df)	3846.21
CFI	0.90
TLI	0.88
RMSEA	0.07
Scaled χ^2(df)	3164.99(163)

$^{+}$Fixed parameter

Table D9: Empathy OCFA1

	Model	
	Estimate	Std. Err.
	Loadings	
E		
em1	0.61	0.01
em2	0.69	0.01
em3	0.66	0.01
em4	-0.53	0.01
em5	0.63	0.01
em6	0.48	0.01
em7	-0.77	0.01
em8	-0.78	0.01
em9	-0.56	0.01
em10	-0.78	0.01

<u>LeftCorrected</u>		
lr1	0.68	0.01
lr2	0.80	0.01
lr3	0.85	0.01
lr4	0.81	0.01
lr5	0.68	0.01
<u>AuthCorrected</u>		
al1	0.81	0.01
al2	0.73	0.01
al3	0.76	0.01
al4	0.50	0.01
al5	0.81	0.01
<u>Acq</u>		
em1	1.00^{+}	
em2	1.00^{+}	
em3	1.00^{+}	
em4	1.00^{+}	
em5	1.00^{+}	
em6	1.00^{+}	
em7	1.00^{+}	
em8	1.00^{+}	
em9	1.00^{+}	
em10	1.00^{+}	
lr1	1.00^{+}	
lr2	1.00^{+}	
lr3	1.00^{+}	
lr4	1.00^{+}	

lr5	1.00⁺	
al1	1.00⁺	
al2	1.00⁺	
al3	1.00⁺	
al4	1.00⁺	
al5	1.00⁺	
	Latent Variances	
E	1.00⁺	
LeftCorrected	1.00⁺	
AuthCorrected	1.00⁺	
Acq	0.04	0.00
	Fit Indices	
χ^2(df)	7500.96	
CFI	0.90	
TLI	0.92	
RMSEA	0.08	
Scaled χ^2(df)	1682.38(239)	

⁺Fixed parameter

Table D10: Empathy OCFA2

	Model	
	Estimate	Std. Err.
	Loadings	
E		
em1	0.65	0.04
em2	0.73	0.04
em3	0.71	0.04

em4	-0.49	0.04
em5	0.68	0.04
em6	0.53	0.04
em7	-0.73	0.04
em8	-0.74	0.04
em9	-0.52	0.04
em10	-0.74	0.04
LeftCorrected		
lr1	0.79	0.01
lr2	0.88	0.01
lr3	0.95	0.01
lr4	0.89	0.01
lr5	0.73	0.01
AuthCorrected		
al1	0.87	0.01
al2	0.75	0.01
al3	0.83	0.01
al4	0.59	0.01
al5	0.87	0.01
Acq		
em1	0.33	0.01
em2	0.33	0.01
em3	0.33	0.01
em4	0.33	0.01
em5	0.33	0.01
em6	0.33	0.01
em7	0.33	0.01

em8	0.33	0.01
em9	0.33	0.01
em10	0.33	0.01
lr1	0.33	0.01
lr2	0.33	0.01
lr3	0.33	0.01
lr4	0.33	0.01
lr5	0.33	0.01
al1	0.33	0.01
al2	0.33	0.01
al3	0.33	0.01
al4	0.33	0.01
al5	0.33	0.01
	Latent Variances	
E	1.00^{+}	
LeftCorrected	1.00^{+}	
AuthCorrected	1.00^{+}	
Acq	1.00^{+}	
	Fit Indices	
χ^2(df)	4465.90	
CFI	0.94	
TLI	0.93	
RMSEA	0.08	
Scaled χ^2(df)	4111.71(163)	

$^{+}$Fixed parameter

Correlations

Table D11: Zero-Sum Left-Right

	CFA1	OCFA1	CFA2	OCFA2
CFA1				
OCFA1	0.991			
CFA2	0.987	0.974		
OCFA2	0.984	0.984	0.984	

Table D12: Zero-Sum Left-Right

	CFA1	OCFA1	CFA2	OCFA2
CFA1				
OCFA1	0.991			
CFA2	0.987	0.974		
OCFA2	0.984	0.984	0.984	

Table D13: Zero-Sum Left-Right

	CFA1	OCFA1	CFA2	OCFA2
CFA1				
OCFA1	0.991			
CFA2	0.987	0.974		
OCFA2	0.984	0.984	0.984	

Table D14: Zero-Sum Left-Right

	CFA1	OCFA1	CFA2	OCFA2
CFA1				
OCFA1	0.991			
CFA2	0.987	0.974		
OCFA2	0.984	0.984	0.984	

Marginal Distributions

Figure D1: Density Plots of Left-Right Factors from Correction Models

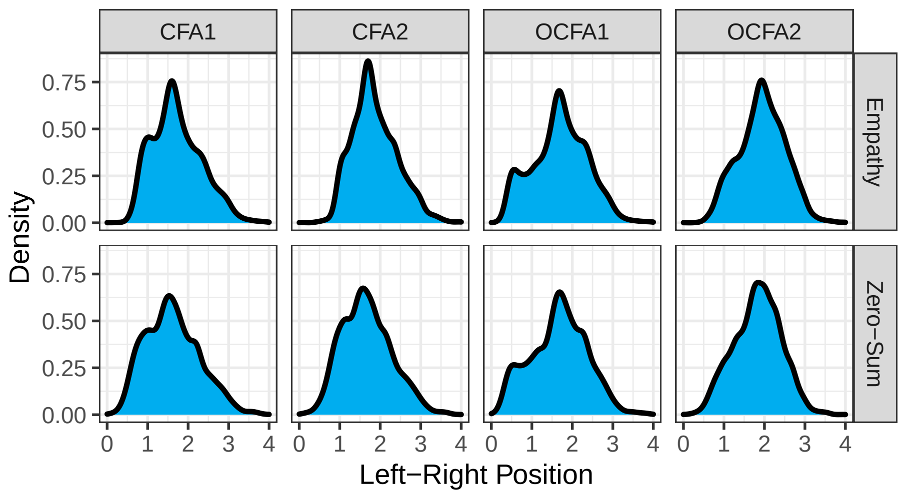
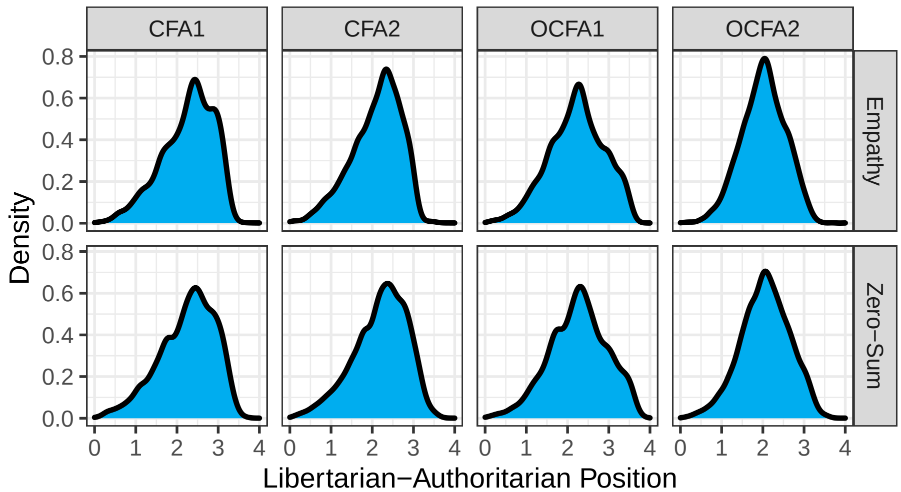


Figure D2: Density Plots of Lib-Auth Factors from Correction Models



Regression Results

Table D15: Zero-Sum Left-Right

	Raw	CFA1	CFA2	OCFA1	OCFA2
Intercept	1.12	1.55	1.61	1.58	1.78
	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
Below GCSE	0.05	0.04	0.04	0.05	0.04
	(0.07)	(0.05)	(0.05)	(0.06)	(0.05)
GCSE/Equiv	0.09	0.06	0.06	0.08	0.06
	(0.05)	(0.04)	(0.03)	(0.04)	(0.03)
A-level/Equiv	0.18	0.10	0.10	0.12	0.09
	(0.05)	(0.04)	(0.04)	(0.04)	(0.03)
Undergrad	0.15	0.05	0.05	0.07	0.03
	(0.05)	(0.04)	(0.03)	(0.04)	(0.03)
Postgrad	0.13	−0.00	0.00	0.02	−0.04
	(0.06)	(0.04)	(0.04)	(0.05)	(0.04)
R^2	0.00	0.00	0.00	0.00	0.00
Adj. R^2	0.00	0.00	0.00	0.00	0.00
Num. obs.	4965	4965	4965	4965	4965

Table D16: Empathy Left-Right

	Raw	CFA1	CFA2	OCFA1	OCFA2
Intercept	1.08	1.62	1.75	1.59	1.87
	(0.05)	(0.04)	(0.03)	(0.04)	(0.03)
Below GCSE	0.17	0.09	0.09	0.12	0.07
	(0.08)	(0.06)	(0.05)	(0.06)	(0.05)
GCSE/Equiv	0.14	0.09	0.10	0.10	0.06
	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)
A-level/Equiv	0.18	0.11	0.11	0.11	0.07
	(0.05)	(0.04)	(0.04)	(0.05)	(0.04)
Undergrad	0.22	0.14	0.15	0.14	0.09
	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)
Postgrad	0.13	0.08	0.10	0.06	0.03
	(0.06)	(0.05)	(0.04)	(0.05)	(0.04)
R^2	0.01	0.00	0.00	0.00	0.00
Adj. R^2	0.00	0.00	0.00	0.00	0.00
Num. obs.	3847	3847	3847	3847	3847

Table D17: Zero-Sum Libertarian-Authoritarian

	Raw	CFA1	CFA2	OCFA1	OCFA2
Intercept	3.05	2.58	2.51	2.62	2.35
	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
Below GCSE	−0.05	−0.05	−0.05	−0.06	−0.06
	(0.07)	(0.05)	(0.05)	(0.05)	(0.05)
GCSE/Equiv	−0.13	−0.09	−0.09	−0.11	−0.09
	(0.05)	(0.04)	(0.04)	(0.04)	(0.03)
A-level/Equiv	−0.45	−0.31	−0.30	−0.33	−0.28
	(0.05)	(0.04)	(0.04)	(0.04)	(0.03)
Undergrad	−0.76	−0.53	−0.52	−0.57	−0.47
	(0.05)	(0.04)	(0.03)	(0.04)	(0.03)
Postgrad	−1.15	−0.79	−0.76	−0.82	−0.67
	(0.06)	(0.04)	(0.04)	(0.04)	(0.04)
R^2	0.15	0.13	0.13	0.13	0.12
Adj. R^2	0.15	0.13	0.13	0.13	0.12
Num. obs.	4965	4965	4965	4965	4965

Education Recode Regression Results

Table D18: Empathy Libertarian-Authoritarian

	Raw	CFA1	CFA2	OCFA1	OCFA2
Intercept	3.11	2.60	2.44	2.59	2.29
	(0.05)	(0.03)	(0.03)	(0.04)	(0.03)
Below GCSE	−0.07	−0.03	−0.04	−0.06	−0.02
	(0.08)	(0.05)	(0.05)	(0.06)	(0.05)
GCSE/Equiv	−0.14	−0.09	−0.10	−0.11	−0.07
	(0.06)	(0.04)	(0.04)	(0.04)	(0.03)
A-level/Equiv	−0.53	−0.34	−0.34	−0.37	−0.27
	(0.06)	(0.04)	(0.04)	(0.04)	(0.03)
Undergrad	−0.77	−0.51	−0.50	−0.55	−0.41
	(0.05)	(0.04)	(0.04)	(0.04)	(0.03)
Postgrad	−1.24	−0.85	−0.82	−0.88	−0.67
	(0.06)	(0.05)	(0.04)	(0.05)	(0.04)
R^2	0.16	0.16	0.16	0.15	0.14
Adj. R^2	0.16	0.16	0.16	0.15	0.14
Num. obs.	3847	3847	3847	3847	3847

Table D19: Zero-Sum Left-Right Alternative

	Raw	CFA1	CFA2	OCFA1	OCFA2
Intercept	1.14	1.56	1.62	1.60	1.80
	(0.03)	(0.03)	(0.02)	(0.03)	(0.02)
GCSE/Equiv	0.07	0.05	0.04	0.06	0.04
	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
A-level/Equiv	0.16	0.09	0.08	0.10	0.07
	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
Undergrad	0.13	0.03	0.03	0.05	0.01
	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
Postgrad	0.12	−0.02	−0.01	0.00	−0.05
	(0.05)	(0.04)	(0.04)	(0.04)	(0.03)
R^2	0.00	0.00	0.00	0.00	0.00
Adj. R^2	0.00	0.00	0.00	0.00	0.00
Num. obs.	4965	4965	4965	4965	4965

Table D20: Empathy Left-Right Alternative

	Raw	CFA1	CFA2	OCFA1	OCFA2
Intercept	1.14	1.65	1.79	1.63	1.90
	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
GCSE/Equiv	0.08	0.06	0.06	0.05	0.04
	(0.05)	(0.03)	(0.03)	(0.04)	(0.03)
A-level/Equiv	0.12	0.07	0.08	0.06	0.04
	(0.05)	(0.03)	(0.03)	(0.04)	(0.03)
Undergrad	0.16	0.10	0.11	0.09	0.07
	(0.04)	(0.03)	(0.03)	(0.04)	(0.03)
Postgrad	0.07	0.04	0.06	0.01	0.00
	(0.06)	(0.04)	(0.04)	(0.05)	(0.04)
R^2	0.00	0.00	0.00	0.00	0.00
Adj. R^2	0.00	0.00	0.00	0.00	0.00
Num. obs.	3847	3847	3847	3847	3847

Table D21: Zero-Sum Libertarian-Authoritarian Alternative

	Raw	CFA1	CFA2	OCFA1	OCFA2
Intercept	3.03	2.56	2.50	2.60	2.33
	(0.03)	(0.03)	(0.02)	(0.03)	(0.02)
GCSE/Equiv	−0.11	−0.08	−0.07	−0.09	−0.07
	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
A-level/Equiv	−0.44	−0.29	−0.28	−0.31	−0.26
	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
Undergrad	−0.74	−0.52	−0.50	−0.55	−0.45
	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
Postgrad	−1.13	−0.77	−0.74	−0.80	−0.65
	(0.05)	(0.04)	(0.04)	(0.04)	(0.03)
R^2	0.15	0.13	0.13	0.13	0.12
Adj. R^2	0.15	0.13	0.13	0.13	0.12
Num. obs.	4965	4965	4965	4965	4965

Table D22: Empathy Libertarian-Authoritarian Alternative

	Raw	CFA1	CFA2	OCFA1	OCFA2
Intercept	3.08	2.59	2.42	2.57	2.28
	(0.04)	(0.03)	(0.03)	(0.03)	(0.02)
GCSE/Equiv	−0.12	−0.08	−0.08	−0.08	−0.06
	(0.05)	(0.03)	(0.03)	(0.04)	(0.03)
A-level/Equiv	−0.50	−0.33	−0.32	−0.35	−0.26
	(0.05)	(0.03)	(0.03)	(0.04)	(0.03)
Undergrad	−0.74	−0.50	−0.48	−0.53	−0.40
	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
Postgrad	−1.21	−0.84	−0.80	−0.86	−0.66
	(0.06)	(0.04)	(0.04)	(0.04)	(0.03)
R^2	0.16	0.16	0.16	0.15	0.14
Adj. R^2	0.16	0.16	0.16	0.15	0.14
Num. obs.	3847	3847	3847	3847	3847