Supplementary Information for:

Anchoring vignettes as a diagnostic tool for cross-national (in)comparability of survey measures: The case of voters' left-right self-placement

Appendix A: The survey and the sample

The survey was launched at the end of January 2020 and finished by February 2020, using internet panels of "respondi AG," an international survey company based in Germany. With the original questionnaire designed in English, each non-English survey was then translated by two certified translators at a translation company specialized in survey translations. We also provided additional information regarding our theoretical concepts to these translators to ensure that the questions are properly translated to reflect the concepts of our interest. Finally, the survey was implemented in each country's official language. A total of 18,051 respondents completed the survey (roughly 2,000 responses per country). Respondent quotas were set according to national demographic distributions for age and gender. In Tables A1 and A2, we show our sample characteristics in comparison with the Euro Stat 2019 data.

Table A1. Sample representativeness by female population (%) over 15-years-old

Country	Euro Stat 2019	Our Survey
France	52.26	52.26
Germany	50.97	51.00
Hungary	52.75	52.14
Italy	51.72	51.81
Netherlands	50.64	50.38
Poland	52.13	52.32
Spain	51.42	50.88
Sweden	49.99	50.88
UK	51.04	50.28

Table A2. Sample representativeness by age group (%)

Country	Age Group	Euro Stat 2019	Our Survey
France	15-39	36.41	32.88
France	40-59	31.74	32.38
France	60 and Above	31.85	34.73
Germany	15-39	34.09	32.61
Germany	40-59	33.32	34.34
Germany	60 and Above	32.59	33.06
Hungary	15-39	35.59	34.52
Hungary	40-59	33.46	36.17
Hungary	60 and Above	30.95	29.31
Italy	15-39	30.78	29.24
Italy	40-59	35.51	38.54
Italy	60 and Above	33.71	32.22
Netherlands	15-39	36.87	33.08
Netherlands	40-59	32.80	33.70
Netherlands	60 and Above	30.33	33.21
Poland	15-39	39.15	38.62
Poland	40-59	31.42	34.15
Poland	60 and Above	29.43	27.23
Spain	15-39	33.45	31.66
Spain	40-59	36.77	39.76
Spain	60 and Above	29.78	28.58
Sweden	15-39	38.39	33.47
Sweden	40-59	30.69	35.20
Sweden	60 and Above	30.92	31.32
UK	15-39	38.81	35.77
UK	40-59	31.96	33.35
UK	60 and Above	29.23	30.88

Appendix B: Diagnosis of the left-right vignettes

1. A test of unidimensionality

We examined whether our vignettes fall on a unidimensional scale as we intended to. To evaluate the extent to which respondents indeed perceive the scale of interest as unidimensional and whether they are able to place the vignettes on the same scale in the order scholars expected, a conventional way is to check the percentage of respondents who actually place these vignettes in the expected order (e.g., King et al. 2004, Bratton 2010, Lee et al. 2015; 2016, Bakker et al. 2014).

Table B1. Frequency of vignette ordering

Ordering	Frequency	Proportion	Order violation
A < B < C	7647	.42	0
A = B = C	1665	.09	0
B < A < C	1466	.08	1
A = B < C	1267	.07	0
C < B < A	1160	.06	3
C < A < B	864	.05	2
A < C < B	838	.05	1
B < C < A	620	.03	2
C < A = B	600	.03	2
A < B = C	573	.03	0
B = C < A	514	.03	2
A = C < B	445	.02	1
B < A = C	392	.02	1

Note: A denotes the leftist and C the rightist party vignette. A < B < C is the expected ordering.

In Table B1, we report how our respondents order the three vignettes. There are 13 orderings of the vignettes, including the orderings that we expect to be made by respondents and ones that go different from the expectation. Overall, 62% of our respondents place the party vignettes exactly in the order we expected with no-order-violation (vignette ties are not considered violation). Moreover, when we apply a strict rule by treating vignette ties as order violations, there are still 42% of the respondents who can place the vignettes in the correct order.[1] When we

[1] The random chance of a respondent placing all three vignettes in the correct order is 0.077 (1/13), when we apply the strict rule, i.e., treat ties as a violation) and 0.307 (4/13) when we

exclude the two Eastern European countries in the sample, the rates go up to 66% (naïve rule) and 46% (strict rule). For interested readers, we also provide country-specific numbers in Table B2.

Table B2. Performance of vignette placement across countries

	% Correct ordering, strict (ties treated as violation)	% Correct ordering, naïve (ties treated as non-violation)	Avg. Rate of Correctly Placing a Pair of Vignettes (ties treated as violation)
France	47.0	65.5	66.3
Germany	54.5	72.0	72.5
Hungary	28.4	50.9	49.1
Italy	27.4	60.4	53.2
Netherlands	44.0	60.8	63.4
Poland	27.1	46.6	47.2
Spain	46.9	65.9	65.9
Sweden	57.7	74.0	74.3
UK	45.5	58.4	63.4

To better understand whether we should be happy with or concerned about these numbers, we can put these results in context by comparing them to existing works in political science. For instance, in their work on questions about political interest, Lee et al. (2015, 2016) show that the percentage of respondents with no-order-violation was 67% and 72% given different country samples. Our analyses of data from other studies show some variances in that rate: 86% of no-order-violation in Bratton's (2010) studies how respondents evaluate democracy in Africa, and 88% of no-order-violation from Bakker et al.'s (2014) data on the comparability of left-right placement among political experts. In contrast, King et al.'s (2004) study of political efficacy has rates of no-order-violations at only 8%. We provide the performance of each of these benchmark studies in Table B3.

apply the naïve rule. Apparently, the proportion of correct orderings in our data shows much higher rates than these random probabilities.

Table B3. Performance of vignette placement in different studies

	% Correct ordering, strict (ties treated as violation)	% Correct ordering, naïve (ties treated non-violation)	Avg. Rate of Correctly Placing a Pair of Vignettes (ties treated as violation)	Number of Vignettes (pairs)
Political efficacy (King et al. 2004)	0.60	34.8	43.4	5 (10)
Democracy (Bratton 2010)	42.7	86.1	72.3	3 (3)
LR position (Bakker et al. 2014)	71.9	88.4	89.1	3 (3)
Political interest (Lee et al. 2015)	49.0	66.9	81.8	4 (6)
Political interest (Lee et al. 2016)	53.6	72.0	70.6	3 (3)
LR placement	42.4	61.8	61.7	3 (3)

More generally, our result falls somewhere in between other important political science measures that have been studied using the anchoring vignette method.[2] Still, we find that many respondents, even when encouraged to think of the left-right as a unidimensional policy dimension, fail to do so. This is perhaps not surprising given that a certain proportion of the population completely lacks the ability to make sense of the left-right in political terms. For example, in our other surveys of a representative sample of the population in Canada and the UK, we found that 18-23% of the respondents reported that they are not very familiar with the terms or never heard of the terms in politics. If we include those who reported a moderate level of familiarity (somewhat familiar) into this group, it is nearly 32-46% of the population that we could identify as not being so confident in their conceptualization or the use of the left-right terms in the respective countries, leaving us 54-68% of the population reporting their familiarity with the left-right terms, as what presented in Table B4.

[2] Of course, the proportion of no-order-violation may depend on the number of vignettes one includes in the survey. To account for this, we also calculate the average rate that a typical respondent is able to place a *pair* of vignettes correctly, and we obtain very similar results: 62% of our respondents can get the ordering of a given pair of vignettes correctly (equities excluded), 81% for Lee et al. (2015), 71% for Lee et al. (2016), 89% for Bakker et al. (2014), 72% for Bratton (2010) while only 43% for King et al. (2004).

Table B4. Self-reported familiarity with the left-right concept

	Canada 2017	Canada 2019	UK 2017	UK 2019	Germany 2018
Very Familiar	29%	33%	41%	37%	43%
Familiar	25%	26%	27%	28%	39%
Somewhat Familiar	23%	20%	14%	17%	12%
Heard of these terms before, but not familiar	19%	19%	16%	16%	5%
Never heard of the "left" and "right" in politics	4%	3%	2%	2%	1%
% Unfamiliar, bottom 2 categories	23	22	18	18	6
% Unfamiliar, bottom 3 categories	46	42	32	35	18

Note: Germany is known to report the highest level of left-right knowledge (e.g., Fortunato et al. 2016). In addition, our survey in Germany was in the field in late March and early April of 2018, just weeks after an unprecedented five months period of negotiations about the composition of the new government following the September 2017 elections -- a period of heightened political interest and awareness.

In addition, unlike the concept of political interest, which is a specific application (to politics) of a well-understood general concept (interest), the assessment of vignettes on the left-right requires the use of the terms "left" and "right" in a way that is fundamentally different from their use in general. Thus, it is unsurprising that this concept is less well understood than interest. What is more surprising is that this potentially complex concept performs as well as it does – and much better than similarly complex concepts like political efficacy.

While a lack of order violations is certainly important to examine, passing this test does not necessarily indicate a lack of CN-DIF. If vignettes are well-constructed and unidimensional, the typical respondent in two countries may well rank order the vignettes in the same way while differing substantially in the absolute placements of the vignettes on the scale. Thus, the diagnostic conclusions about CN-DIF in the existing literature have not been based on the frequency of order violations, but rather on cross-national comparisons of the vignette placements themselves (e.g., Bakker et al. 2014; Lee et al. 2015, 2016). We examine this below.

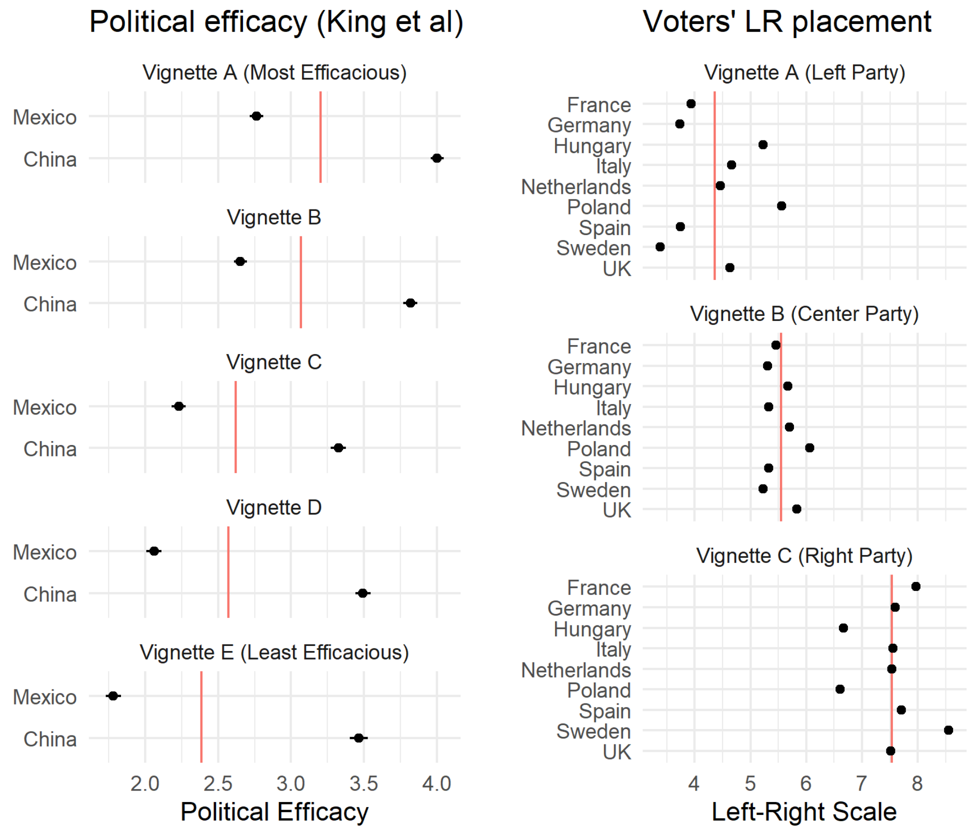
2. Examination of systematic differences in individuals' placement of vignettes on the left-right scale

Even when respondents place vignettes in the same order, respondents in some countries may systematically shift all vignettes to the left or right of the scale or use only part of the scale. For example, King et al.'s (2004) research on political efficacy found that respondents in China and Mexico used very different parts of a five-point efficacy scale to evaluate the same vignettes.

This is clearly illustrated in the left panel in Figure B1 (reproduced using original data) where we plot mean values of the respondents' placements for each vignette and 95% confidence intervals. Indeed, the average Mexican respondent's rating of the vignette describing the most politically efficacious individual in the study was actually lower than the rating the average Chinese respondent gave to the same vignette. More generally, the figure clearly demonstrates that for political efficacy, the cross-country variance in vignette placement is dramatically bigger than the variance across vignettes within countries. When this happens, it suggests very high levels of cross-national DIF.

What does this variation look like in our left-right policy placement data? Looking at the right panel in Figure B1, in contrast to the political efficacy case, there is clearly much more variation between vignettes within counties than there is across countries within each vignette. Most importantly, as what we will show in the paper, we can examine this impression by estimating random intercept models to parse the total variance depicted in these figures into so that we know exactly what proportion of the total variance comes from between-vignette variance and within-vignette variance (Lee et al. 2016). Our estimates suggest that cross-country variance (within-vignette) constitutes 11.4% of the total variance in our left-right placement data and 90% in King et al.'s political efficacy data. This again shows that the concept of left-right policy placement does not really suffer from the issue of CN-DIF.

Figure B1. Variation in vignette placements by country and vignette



3. Changes in country rankings after correcting the DIF

Our last examination of whether DIF is a severe problem in cross-country comparisons of ordinary citizens' left-right placements is to compare respondents' self-placements before and after any DIF is corrected (in the ways described in King et al. 2004). The idea of this task is straightforward. If DIF causes severe issues to the cross-country comparability, we should observe dramatic differences between the raw self-placements and the corrected measures.

For respondents who have ordered the vignettes correctly, with no ties, correcting the raw scores for DIF simply requires us to directly rescale the respondent's self-placement relative to the location of the set of vignettes (for details, see King et al. 2004). For the cases in which vignettes are rated by a respondent as the same category (tied) or mis-ordered, King et al. (2004) and King and Wand (2007) suggested several different methods, including omitting them and allocating the tied and mis-ordered cases to categories in some sensible way. King et al. (2004) simply assigned the respondent's self-placement to one of the corrected ranks with equal probability (uniform allocation), and later on, King and Wand (2007) improved on this by using a censored ordered probit model to estimate the probability of allocation to one of the possible ranks.

Figure B2. Relative country rankings by correction methods

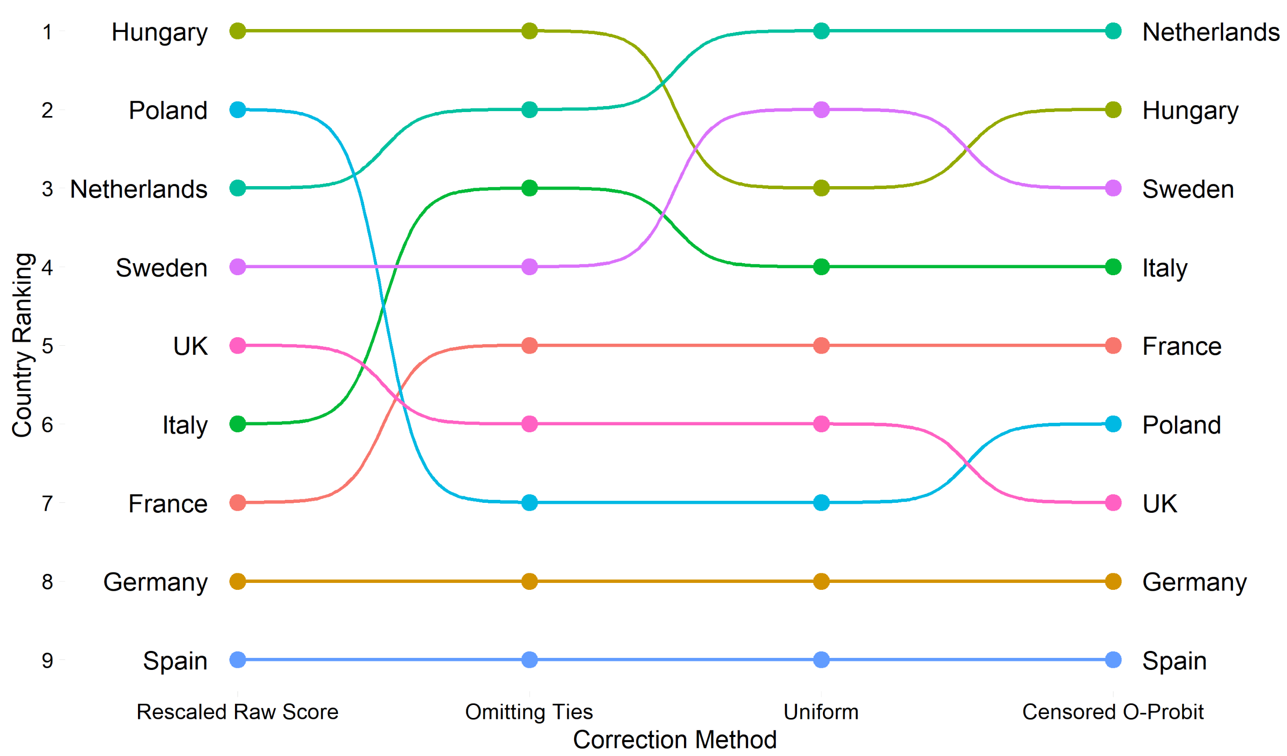


Figure B2 shows how the countries' relative ranking changes after correcting DIF by comparing the ranking based on the country mean of the raw left-right self-placements to the ranking based on three different correcting methods. Namely, when tied and mis-ordered answers are omitted ("Omitting Ties"), when those tied and mis-ordered cases are allocated to existing categories uniformly ("Uniform"), and when those incorrectly ordered answers are assigned to correct categories using a censored ordered probit model ("Censored O-Probit"). Figure B2 suggests that the rank order of countries before and after correction is mostly very similar. Most countries stay at the same level or vary by a rank or two, with Poland as the only exception, which moved from the rightest country in the sample with the raw score to a center or left one with the corrected score. This clearly provides another piece of evidence that is consistent with what we have seen from above analyses: the DIF issue may have some effect on respondents' left-right self-placement, but the effect is not large. All raw and corrected scores (country means) and rankings in greater detail are presented in Table B5.

Table B5. Mean and relative ranking of each country by correction methods

	Rescaled Raw*	Omitting Ties		Uniform		Censored O-Probit	
	mean (rank)	mean (rank)	Δrank	mean (rank)	Δrank	mean (rank)	Δrank
France	3.95 (7)	4.24 (5)	+2	4.17 (5)	+2	4.20 (5)	+2
Germany	3.58 (8)	3.95 (8)	0	3.95 (8)	0	3.94 (8)	0
Hungary	4.12 (1)	4.39 (1)	0	4.24 (3)	-2	4.29 (2)	-1
Italy	3.98 (6)	4.33 (3)	+3	4.19 (4)	+2	4.24 (4)	+2
Netherlands	4.11 (3)	4.37 (2)	+1	4.29 (1)	+2	4.34 (1)	+2
Poland	4.12 (2)	4.06 (7)	-6	4.05 (7)	-6	4.06 (6)	-5
Spain	3.35 (9)	3.80 (9)	0	3.81 (9)	0	3.76 (9)	0
Sweden	4.03 (4)	4.32 (4)	0	4.25 (2)	+2	4.29 (3)	+2
UK	4.00 (5)	4.08 (6)	-1	4.07 (6)	-1	4.06 (7)	-1
Mean	3.91	4.17		4.11		4.13	
Scale	1-7	1-7		1-7		1-7	

Note: Relative rankings are in parentheses (the rightest average placement is ranked 1).
*In the survey, respondents placed themselves on the left-right scale ranging from 0 to 10 (raw score) whereas the DIF-corrected scale ranges from 1 to 7 given the number of vignettes (2*N + 1, N= number of vignettes). To make the two scales comparable, we rescaled the 11-point raw scale to a 7-point one.

In addition to presenting the raw and corrected scores as well as rankings using our own data, we further make a comparison between our survey, which was fielded between January and February 2020, and the Eurobarometer surveys conducted in December 2019 (EB92) and August 2020 (EB93). As presented in Table B6, while the average left-right placement of respondents varies slightly across surveys, the results reveal some similarities in terms of country rankings.

Table B6. Comparing LR self-placement country rankings across surveys

Country	**EB92**		**EB93**		**Our Survey**	
	Mean	Rank	Mean	Rank	Mean	Rank
France	3.499	7	3.702	5	3.95	7
Germany	3.480	8	3.544	8	3.58	8
Hungary	4.321	2	4.204	2	4.12	1
Italy	4.027	3	4.087	3	3.98	6
Netherland	3.592	6	3.595	7	4.11	3
Poland	4.324	1	4.370	1	4.12	2
Spain	3.336	9	3.338	9	3.35	9
Sweden	3.721	4	3.724	4	4.03	4
UK	3.632	5	3.633	6	4.00	5

Note: EB data was rescaled from 1-10 to 1-7. EB92 was fielded November-December 2019, and EB93 was fielded July-August 2020. Our survey was fielded January-February 2020.

Appendix C: Technical details and robustness checks

1. Estimating $R_{CN\text{-}DIF}$

Formally, the multi-level model we estimate for vignette placements can be summarized as the following

$$(1)\quad y_{kji} = \alpha + u_k + u_j + u_{jk} + e_{jki}\,,$$

where y_{kji} denotes the placement of vignette j in country k by respondent i, α the constant, u_j the random intercepts for each of j = 1… J vignettes, u_k the random intercepts for k = 1…K countries, u_{jk} the random intercepts for J*K vignettes in countries, and e_{jki} the residual that captures the random effect on placements attributable to unmeasured factors idiosyncratic to the individual-country-vignette.[3] If we assume that each of the random effect terms are distributed independently normal with zero means, each will contribute a variance term (σ_k^2, σ_j^2, σ_{jk}^2, and σ_e^2) to the likelihood function and estimates of these variance terms can be used to produce direct measures of the proportion of variance attributable to vignettes vs. countries (as well as estimates in the uncertainty around this proportion).

Moreover, the measure of CN-DIF is calculated by:

$$(2)\quad R_{CN\text{-}DIF} = (\sigma_k^2 + \sigma_{jk}^2)/(\sigma_j^2 + \sigma_k^2 + \sigma_{jk}^2).$$

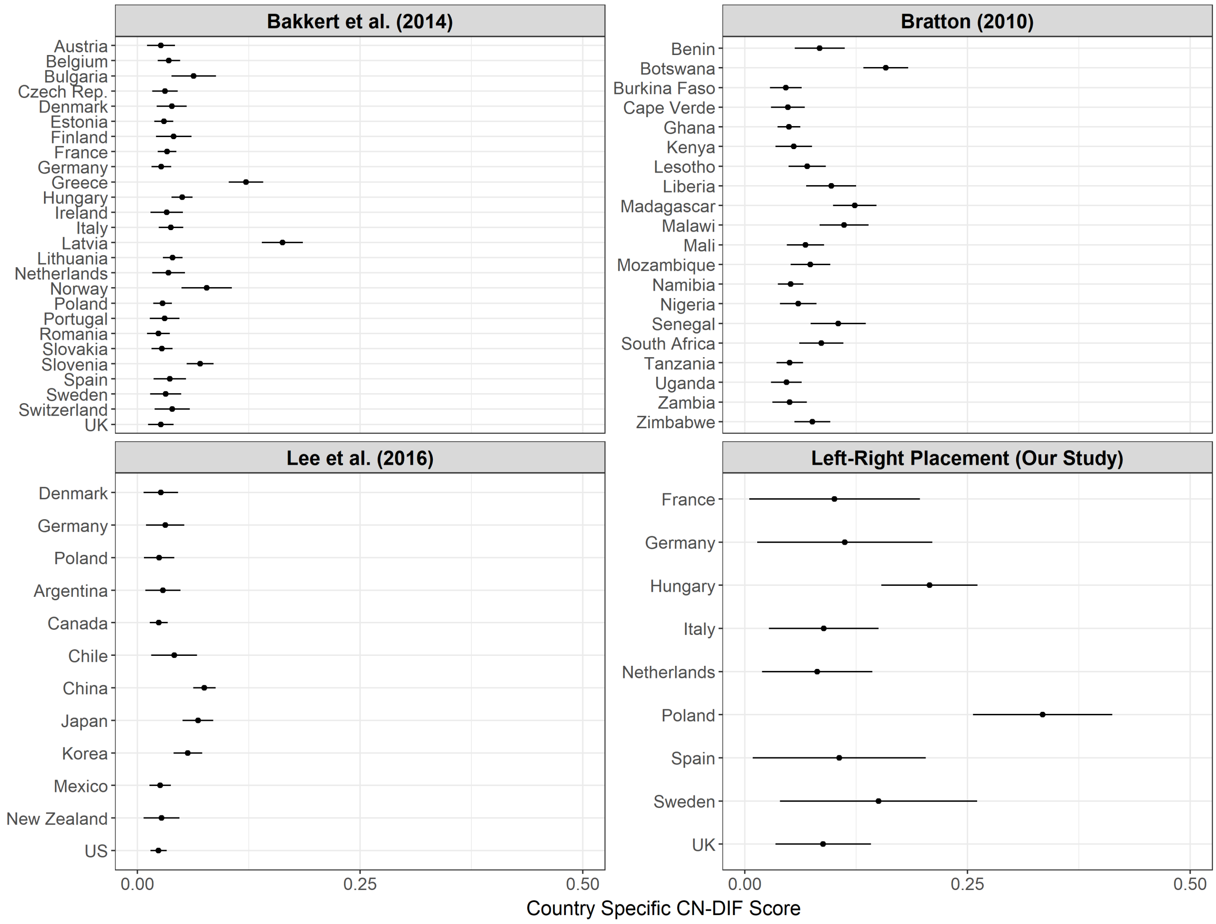
Note that the variance term for the country-vignette effects (i.e., σ_{jk}^2) is included in the numerator since we want to "count" it as part of the country effect. For instance, those effects could move all vignettes placements in the same country one way or the other. At the same time, they could move all placements of a given vignette in a country in one direction while possibly moving all placements of a different vignette (in the same country) in the opposite direction (e.g., a country effect that squeezed or expanded the placements of vignettes on the scale relative to a different country).

[3] In the example, we use a linear functional form but the same analysis applies to other forms. The only difference is that some models will not provide estimates for a residual variance. This term, however, plays no role in the diagnostic metrics we advocate here.

We do not include the estimate of the residual variance (i.e., σ_e^2) in the calculation since the inclusion of the estimate of the residual variance in the denominator would lead to the same conclusion about the proportion of variance associated with country relative to that associated with vignettes. Excluding the residual variance produces essentially the same estimates of this quantity that one would get from first collapsing the data to country-vignette means and then parsing the variance (and so that corresponds directly to the visual impression given in diagrams like those in Figure B1). However, including the residual variance can easily mask the importance of country effects when there is a lot of idiosyncratic, individual-level variation in the respondent ratings (i.e., that does not correspond to country or vignette). Suppose, for example, that 90% of the observed variance in placements was idiosyncratic, 5% was associated with vignettes, 3% with country, and 2% with country-vignette. In this case, the statistic for CN-DIF including idiosyncratic variance in the denominator would be 5%, while excluding it would be 50%. The latter case, 50% of the variance in placements explained by country is due to either vignette or nationality, is in our view a problem with CN-DIF. If one instead just reported the percent of the total variation in placements due to nationality (5%), one might conclude there was little CN-DIF.

When calculating the $R_{CN\text{-}DIF}$ score, there are two ways to estimate the necessary quantities from equation (1): 1) using the individual-level data, or 2) using the country-level data by collapsing vignette placements to country-vignette means. In fact, both approaches produce the same substantive message (as they must), but the advantage of using the individual-level data is that we can bootstrap standard errors for our metrics such that the uncertainty in the original country-vignette means is carried through the whole analysis. While such considerations will be inconsequential in large surveys with tiny standard errors on the country vignette averages, it could produce large confidence intervals for country-vignette means in studies that rely on a small number of homogeneous cases within a country. As we have demonstrated the estimated results using individual-level data in the main text, we present the estimates using country-level data as a robustness check in Figure C1. As one may see, doing so essentially yields the same substantive conclusion.

Figure C1. Country-specific CN-DIF scores in four studies (using aggregate level data)



2. Is our conclusion biased by an online sample with better-educated respondents?

As our sample has only been stratified by age and gender, one may concern that our sample is overrepresented with well-educated respondents and therefore our inference might be biased. In particular, respondents with lower education levels may find it difficult to understand policy terms such as "regulation" or "redistribution", making the validity of our conclusion challenged. Indeed, by comparing our data with the Euro Stat 2019 data, respondents in our survey tend to be better educated, as presented in Table C1. To investigate whether the sample bias may change our conclusion, we produce our country-specific CN-DIF estimates by dividing our samples into groups with higher and lower education levels. Respondents with higher levels of education are the ones who finished tertiary education. The estimated results are presented in Figure C2, which is essentially a replication of Figure 2 in main text. As one may see, all panels in Figure C2 show that Poland and Hungary stand out with greater CN-DIF scores, although this pattern is more pronounced among respondents with lower education levels. These results also suggest that underrepresentation of respondents with lower educational levels does not really change our substantive conclusion. Also, to assure the reader that our inference is not biased because of the under- or over- representation of a specific type of respondents, we produce similar tests based on age and gender sub-groups. The results are presented in Figure C3 and C4, and they suggest that our conclusion remain consistent.

Table C1. Sample representativeness by education level (%) over 15-years-old

Country	**Euro Stat 2019**		**Our Survey**	
	Upper- and post-secondary non-tertiary education, and less (levels 0-4)	Tertiary education (levels 5-8)	Upper- and post-secondary non-tertiary education, and less (levels 0-4)	Tertiary education (levels 5-8)
France	66.2	33.8	50.23	49.77
Germany	74.0	26.0	59.09	40.91
Hungary	77.5	22.5	61.14	38.86
Italy	82.6	17.4	68.02	31.98
Netherlands	65.2	34.8	59.45	40.55
Poland	71.8	28.2	48.39	51.61
Spain	64.9	35.1	41.87	58.13
Sweden	62.2	37.8	58.82	41.18
UK	59.3	40.6	32.82	67.18

Figure C2. Country-specific CN-DIF scores by education level

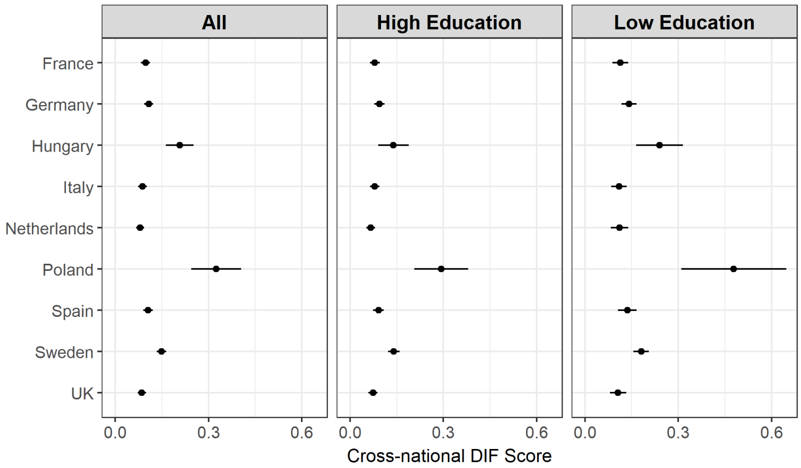


Figure C3. Country-specific CN-DIF scores by age group

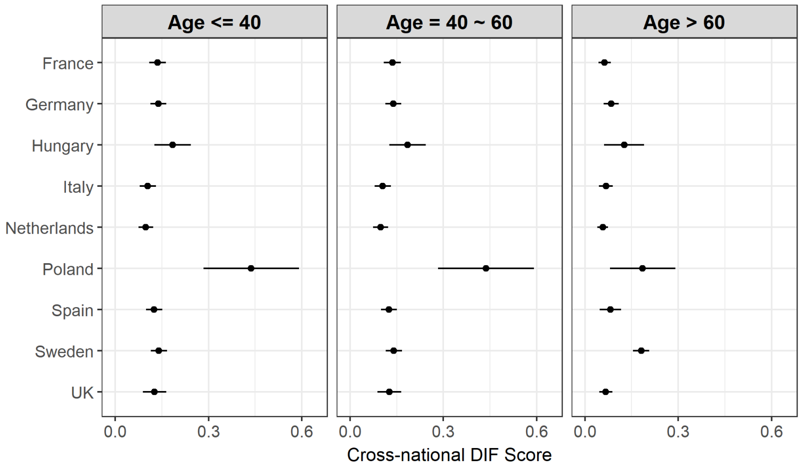
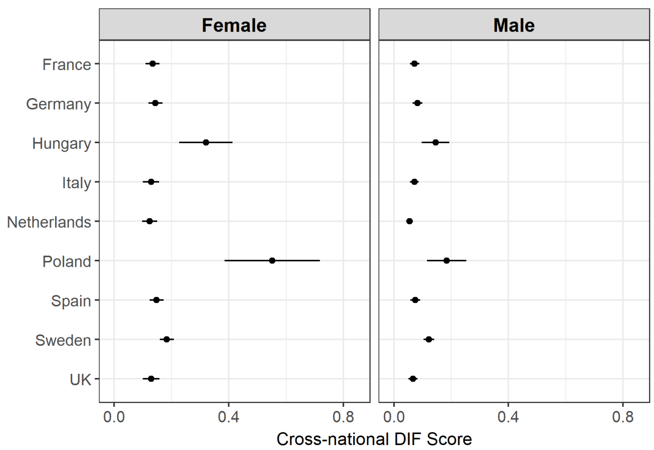


Figure C4. Country-specific CN-DIF scores by gender



Appendix D: Additional information for benchmark studies

1. Overview of benchmark studies examined

There is a small literature that has used anchoring vignettes to diagnose CN-DIF in other important political concepts, including political interest, political efficacy, and levels of democracy. King et al.'s (2004) original study attempts to explore CN-DIF in self-reported political efficacy in China and Mexico, and the levels of CN-DIF they identified are the highest of any published vignette-study of a political variable. Indeed, CN-DIF in this case is so high that there is no question of its analytic relevance (i.e., if one relied on uncorrected measures to make cross-national comparisons in this case, one would make grossly wrong inferences). Compared to that, Michal Bratton's (2010) study on citizen's perceptions of the extent of democracy in 19 African countries, Lee et al.'s (2015, 2016) two studies on self-reported political interest, and Bakker et al.'s (2014) vignette-based diagnosis of experts' left-right placements of political parties, all showed sufficiently low levels of CN-DIF that the authors ultimately encouraged the continued use of these measures (and the corresponding copra of existing data) to make cross-national comparisons.

Table D1. Overview of relevant studies in political science

Study	#Vignettes	#Countries	Level of DIF original study concluded
Political efficacy (King et al. 2004)	5	2	High
Democracy (Bratton 2010)	3	19	Low to Medium
Experts' LR Placement (Bakker et al. 2014)	3	26	Low
Political interest (Lee et al. 2015)	3	3	Low
Political interest (Lee et al. 2016)	3	12	Low

2. Analyses of the data from Bakker et al. (2014)

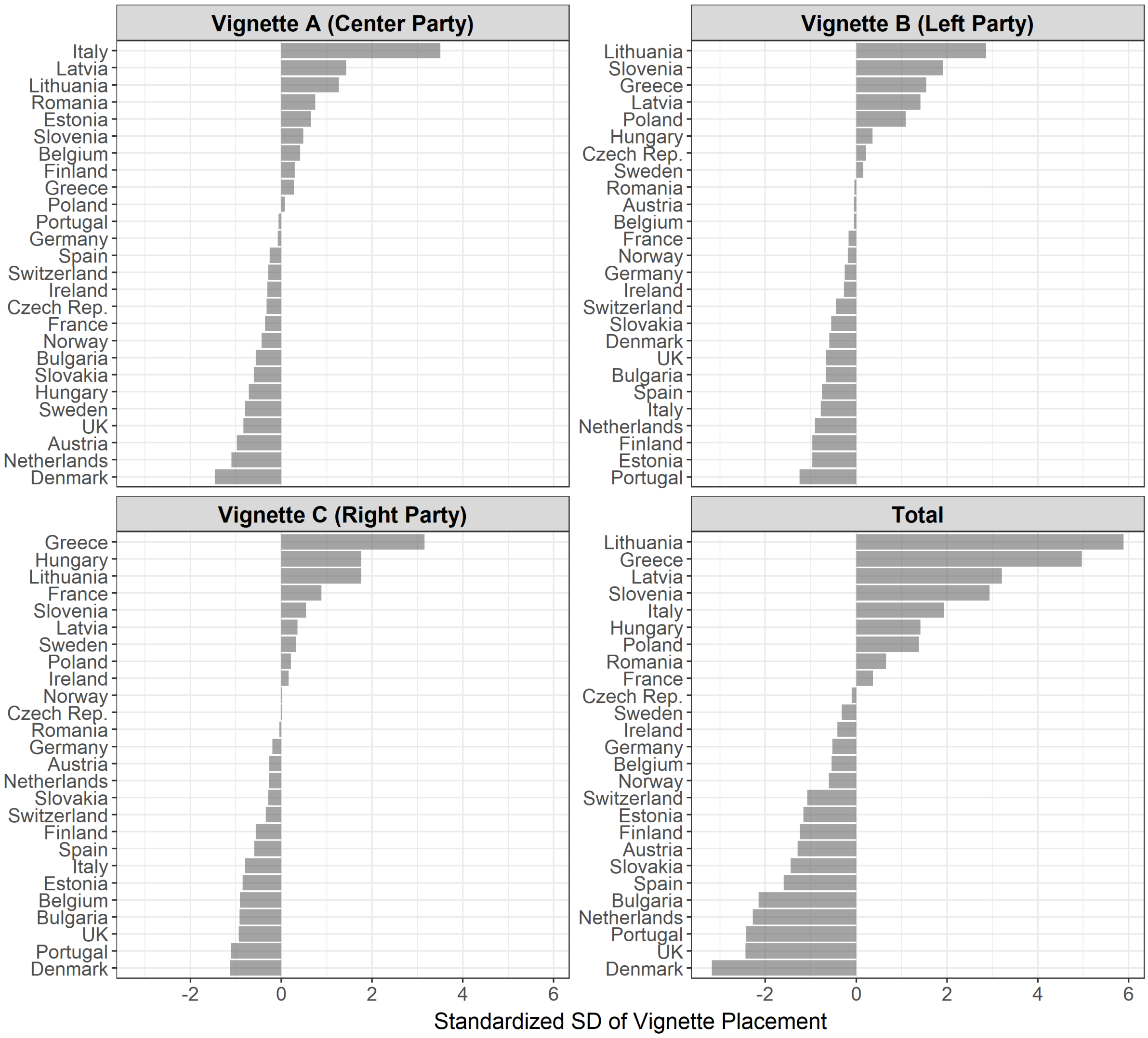
Compared to other benchmark studies, the work by Bakker et al. (2014) on left-right party placements made by political experts seems to be an interesting case. On the one hand, our estimated results in Figure 2 indicate that there is very little CN-DIF in experts' left-right placements as there is no country that gets a score that is statistically and significantly different from other countries. This is consistent with their own conclusion that party positions from the CHES expert data are generally comparable across countries.

On the other hand, the extremely wide 95% confidence intervals of the estimates for Greece, Latvia, Lithuania, and Slovenia warrant a further explanation. As we discussed in Appendix C, our estimation procedure begins with bootstrapping data at the individual level such that our country-specific metrics reflect uncertainty embedded in the original individual data. Large confidence intervals thus indicate the heterogeneity between individual experts (in terms of their placements of vignettes) within a country, particularly when the number of individual experts is only a dozen for each country. On average, there are only about thirteen experts who place parties in each country. This is obviously a much smaller sample size compared to other benchmark studies with ordinary citizens. The heterogeneity among these experts (in some countries) clearly stands out when we plot the standard error of the experts' placement of the vignettes.

In Figure D1, we calculate and plot the standardized standard error of the experts' vignette placements. Greece, Latvia, Lithuania, and Slovenia are the four countries that reveal more with-country heterogeneity than other countries. Therefore, our results suggest that researchers need to be vigilant when comparing experts' assessments of party positions of the above four countries with other countries. Two of these countries – Greece and Latvia – are described as being less comparable in Bakker et al.'s original work in which they stated that "while [the authors] find that experts in some countries conceive of left–right in slightly different ways than experts in other countries, this is largely limited to Bulgaria, Greece, Latvia, and Norway" (p.5).[4]

[4] In Bakker et al. (2014), this finding is based on the pair-wise comparison of the mean placements (of vignettes). Specifically, they found statistically significant differences in the ways assessing vignettes between those experts in Latvia and Norway (left party), those in Bulgaria and Greece (center party), and those in Greece and Bulgaria (right party).

Figure D1. Standardized SD of vignette placements in Bakker et al. (2014)



Note: Standardized $SD_{ij} = (SD_{ij} - M_j) / S_j$, where SD_{ij} indicates the value of SD for a country *i* in vignette *j*, M_j refers to the mean value of all country SDs within a vignette, and S_j denotes the standard error of all country SDs within the vignette.