**Supplementary Materials**

**Supplementary Methods**

**Measurement Invariance Test**

The most lenient invariance, configural invariance, only assumes the equal measurement model across different groups. It is deemed to be supported when the resultant fit indicators are found to be acceptable, RMSEA and SRMR < .08, CFI ≥ .90 (Hu & Bentler, 1999). Metric invariance additionally requires the equal factor loadings. To support scalar invariance, the equal intercept assumption should also be satisfied. Finally, the most restrictive invariance, residual invariance, is supported when the equal residuals are assumed. The latter three levels of invariance were tested in terms of changes in model fit indicators. In general, metric invariance is supported when ΔRMSEA and ΔSRMR < +.30, and ΔCFI ≥ -.01, and scalar and residual invariance are supported when ΔRMSEA and ΔSRMR < +.15, and ΔCFI ≥ -.01 after adding the additionally required constraint for the responsive level of the invariance (Putnick & Bornstein, 2016). Because scalar invariance is minimally required for cross-group comparisons, I focused on whether this level of invariance was supported in the present study. Because dichotomous variables were used, I employed the robust unweighted least squares (USLMV) estimator to minimize potential bias (Savalei et al., 2015).

**Differential Item Functioning Test**

I examined whether each item reported a significant result from the χ2 test. For each item, these three models were compared via χ2.

Model 1: logit Pr = α + β1 x ability (without any DIF)

Model 2: logit Pr = α + β1 x ability + β2 x group (with a uniform DIF)

Model 3: logit Pr = α + β1 x ability + β2 x group + β3 x group x ability (with a non-uniform DIF)

where Pr is a cumulative probability that one’s response to the tested item would be 1 (versus 0) given their ability, i.e., the postconventional reasoning, and group, i.e., gender, political or religious affiliation. The comparison of models 1 versus 3, models 1 versus 2, and models 2 verses 3 are associated with the total DIF, uniform DIF, and non-uniform DIF, respectively. Whether the *p*-values resulting from the three χ2 tests were smaller than .01 was examined following *lordif* guidelines (Choi et al., 2011). In addition to the *p*-values, *R2* values were also investigated. Because small *p*-values per se do not necessarily confirm presence of practically significant effects, I also examined whether *R2* values were .02 or higher. Finally, whether the change in the regression coefficient, Δβ1, was .10 or higher was investigated as well for effect size check.

These parameters for tests were generated through Monte Carlo simulations implemented in *lordif*. For each simulation, 1,000 replications were performed. To improve the speed of the simulation process, I employed an R routine for t. When all the above mentioned three conditions were satisfied, I assumed that there was a significant DIF across groups in the tested item, and thus, the item significantly favored a specific group in a biased manner.

**Supplementary References**

Choi, S. W., Gibbons, L. E., & Crane, P. K. (2011). lordif: An R Package for Detecting Differential Item Functioning Using Iterative Hybrid Ordinal Logistic Regression/Item Response Theory and Monte Carlo Simulations. *Journal of Statistical Software*, *39*(8), 1–30. https://doi.org/10.18637/jss.v039.i08

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, *41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

Savalei, V., Bonett, D. G., & Bentler, P. M. (2015). CFA with binary variables in small samples: A comparison of two methods. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01515

**Supplementary Figures**

**Figure S1**

*Differential Item Functioning Test Results for Heinz Dilemma (Dilemma 1) by Gender*



*Note*. Top: *p*-value from χ2 test for each item. Middle: *R2* value for each item. Bottom: Δβ1 for each item. For Top and Middle panels: Left: Model 1 vs. Model 2 (uniform DIF); Middle: Model 1 vs. Model 3 (total DIF); Right: Model 2 vs. Model 3 (non-uniform DIF).

**Figure S2**

*Differential Item Functioning Test Results for Newspaper (Dilemma 2) by Gender*



*Note*. Top: *p*-value from χ2 test for each item. Middle: *R2* value for each item. Bottom: Δβ1 for each item. For Top and Middle panels: Left: Model 1 vs. Model 2 (uniform DIF); Middle: Model 1 vs. Model 3 (total DIF); Right: Model 2 vs. Model 3 (non-uniform DIF).

**Figure S3**

*Differential Item Functioning Test Results for Escaped Prisoner (Dilemma 3) by Gender*



*Note*. Top: *p*-value from χ2 test for each item. Middle: *R2* value for each item. Bottom: Δβ1 for each item. For Top and Middle panels: Left: Model 1 vs. Model 2 (uniform DIF); Middle: Model 1 vs. Model 3 (total DIF); Right: Model 2 vs. Model 3 (non-uniform DIF).

**Figure S4**

*Differential Item Functioning Test Results for Heinz Dilemma (Dilemma 1) by Political Affiliation*



*Note*. Top: *p*-value from χ2 test for each item. Middle: *R2* value for each item. Bottom: Δβ1 for each item. For Top and Middle panels: Left: Model 1 vs. Model 2 (uniform DIF); Middle: Model 1 vs. Model 3 (total DIF); Right: Model 2 vs. Model 3 (non-uniform DIF).

**Figure S5**

*Differential Item Functioning Test Results for Newspaper (Dilemma 2) by Political Affiliation*



*Note*. Top: *p*-value from χ2 test for each item. Middle: *R2* value for each item. Bottom: Δβ1 for each item. For Top and Middle panels: Left: Model 1 vs. Model 2 (uniform DIF); Middle: Model 1 vs. Model 3 (total DIF); Right: Model 2 vs. Model 3 (non-uniform DIF).

**Figure S6**

*Differential Item Functioning Test Results for Escaped Prisoner (Dilemma 3) by Political Affiliation*



*Note*. Top: *p*-value from χ2 test for each item. Middle: *R2* value for each item. Bottom: Δβ1 for each item. For Top and Middle panels: Left: Model 1 vs. Model 2 (uniform DIF); Middle: Model 1 vs. Model 3 (total DIF); Right: Model 2 vs. Model 3 (non-uniform DIF).

**Figure S7**

*Differential Item Functioning Test Results for Heinz Dilemma (Dilemma 1) by Religious Affiliation*



*Note*. Top: *p*-value from χ2 test for each item. Middle: *R2* value for each item. Bottom: Δβ1 for each item. For Top and Middle panels: Left: Model 1 vs. Model 2 (uniform DIF); Middle: Model 1 vs. Model 3 (total DIF); Right: Model 2 vs. Model 3 (non-uniform DIF).

**Figure S8**

*Differential Item Functioning Test Results for Newspaper (Dilemma 2) by Religious Affiliation*



*Note*. Top: *p*-value from χ2 test for each item. Middle: *R2* value for each item. Bottom: Δβ1 for each item. For Top and Middle panels: Left: Model 1 vs. Model 2 (uniform DIF); Middle: Model 1 vs. Model 3 (total DIF); Right: Model 2 vs. Model 3 (non-uniform DIF).

**Figure S9**

*Differential Item Functioning Test Results for Escaped Prisoner (Dilemma 3) by Religious Affiliation*



*Note*. Top: *p*-value from χ2 test for each item. Middle: *R2* value for each item. Bottom: Δβ1 for each item. For Top and Middle panels: Left: Model 1 vs. Model 2 (uniform DIF); Middle: Model 1 vs. Model 3 (total DIF); Right: Model 2 vs. Model 3 (non-uniform DIF).