**Supplementary Materials 2**

To investigate how robust the obtained findings are to changes in the construction of the dependent variable, regressions were run using three alternative constructions – one where, for parties with missing expert ideology data, the mean of the estimates given by *all* participants within the electoral sample, rather than just those with university degrees, is used (‘all-estimate’), one where participant-party utility scores are calculated without factor weights (‘unweighted’), and one where the participant-party ‘ideology’ utility factor is determined by the modulus of the difference between ideology scores rather than directional differences (‘proximity’). In all other respects, each construction is identical to the main construction. The results of multi-level regression analyses identical in design to those already reported conducted on the full waves 1-5 sample with the continuous measure of consistency, for each construction, are reported in Table 2.

**Table 2.** *Regression models with alternatively-constructed Continuous Consistency*

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Factor | | | | All-Estimate | | Unweighted | | Proximity | |
| *Fixed Effects* | | | | | |  | |  | |
| *Individual-level predictors* | | | | | |  | |  | |
|  | | Intercept | | .675 (.009), **p < .001\*\*\*** | | .617 (.010),  **p < .001\*\*\*** | | .676 (.008), **p < .001\*\*\*** | |
|  | | Age | | .001 (<.001),  **p < .001\*\*\*** | | .001 (< .001),  **p < .001\*\*\*** | | .001 (<.001), **p < .001\*\*\*** | |
|  | | Political sophistication | | .006 (.002), **p < .001\*\*\*** | | .010 (.002), **p < .001\*\*\*** | | .011 (.002),  **p < .001\*\*\*** | |
|  | | Efficacy | | .105 (.007), **p** **< .001\*\*\*** | | .132 (.007), **p < .001\*\*\*** | | .105 (.006),  **p < .001\*\*\*** | |
| *Institutional-level predictors* | | | | | |  | |  | |
|  | | Log years of democracy | | -.005 (.009), p = .546 | | .008 (.010),  p = .444 | | <.001 (.009), p = .978 | |
|  | | Media | | -.004 (.010), p = .698 | | -.010 (.012), p = .413 | | .008 (.009),  p = .388 | |
|  | | Polarization | | .003 (.007), p = .703 | | -.004 (.010),  p = .668 | | .002 (.007),  p = .831 | |
|  | | Clear lines | | .021 (.007), **p = .003\*\*** | | .013 (.009), p = .176 | | .014 (.007),  p = .051 | |
|  | | ENEP | | -.133 (.007),  p = .061 | | -.016 (.010), p = .112 | | -.009 (.007),  p = .214 | |
|  | | Personal vote | | -.163 (.008), **p = .030\*** | | -.005 (.009), p = .576 | | -.014 (.007),  **p = .035\*** | |
|  | | Disproportionality | | .010 (.006), p = .108 | | -<.001 (.008),  p = .989 | | .005 (.006), p = .330 | |
|  | | | | | |  | |  | |
|  | | Log years of democracy \* age | | <.001 (<.001),  p = .311 | | -<.001 (<.001) p = .371 | | <.001 (<.001),  p = .653 | |
|  | | Efficacy \* media | | .010 (.007), p = .136 | | .015 (.007),  **p = .032\*** | | .005 (.006),  p = .385 | |
| Factor | | | All-Estimate | | Unweighted | | Proximity | |
| *Individual-level variance components* | | | | |  | |  | |
|  | Intercept | | .006 (.001) | | .009 (.001) | | .005 (.001) | |
|  | Age | | <.001 (<.001) | | <.001 (<.001) | | .001 (.001) | |
|  | Political sophistication | | <.001 (<.001) | | <.001 (<.001) | | <.001 (<.001) | |
|  | Efficacy | | .003 (.001) | | .005 (.001) | | .003 (.001) | |
|  | Residual | | .113 (.003) | | .116 (.003) | | .105 (.003) | |

The only notable discrepancy relative to the main results already reported is that disproportionality is null in all three models. However, the finding that, in countries with both SMD and MMD elections, continuous consistency was higher in the MMD elections, was replicated in all three constructions (all-estimate: .80 (.34) vs .78 (.35), *F*(1, 62528)=69.715, *p<*.001*, f* = .03; unweighted: .76 (.35) vs .75 (.36), *F*(1, 62753)=38.286, *p<*.001*, f* = .03; proximity: .81 (.33) vs .79 (.33), *F*(1, 62528)=47.297, *p<*.001*, f* = .03). This suggests the whilst the difference in proportionality across systems is sensitive to the analysis strategy, the impact of proportionality within the same system is more robust.

Aside from disproportionality, the initial pattern of results was robust across all three models for the individual-level predictors, years of democracy, media density, party polarization, ENEP, and the years of democracy \* age interaction. The effect of personal vote incentives was replicated in two of three constructions, being non-significant in the unweighted construction. The same can almost be said for clear lines of responsibility, though its effect is at *p* = .051 in the proximity construction. In the aggregate then, the evidence suggests that both personal vote incentives and clear lines of responsibility likely have real effects, since across five models using a continuous measure of vote consistency, the former is significant in four and the latter significant in three, though these effects are also clearly sensitive to the weighting of the factors. We would contend that weighting provides the best model of consistency, but future work could certainly test the robustness of a wider range of more sophisticated weighting schemes. The significant efficacy \* media interaction in the unweighted model is likely spurious given its lack of significance elsewhere.