**Creating the “Programmatism” Indices**

**A New Memo**

**(HK 2016/02/13-15)**

**Introduction and Overview**

I will propose a stepwise procedure to (re)generate the Programmatism indices Kent Freeze and I produced in 2010, as well as to make possible efficient modifications in light of criticisms.

The step-wise procedure allows different branches at each and every step:

***Step 1: Means-Correcting the experts’ judgments***

Purpose: to get country-based anchor points out of the picture, or at least make them less relevant. Reasonable minds might disagree…. So let’s produce two datasets:

**DATASET A**: all subsequent calculations based on “means” data.

**DATASET B**: all subsequent calculations based on “natural, raw” data, as supplied by the expert scores.

***Steps 2-4: Preparing the individual “attributes” of programmatism for each party/issue/country:***

**Unit of analysis and construction**: each issue i for each party j in country k, separately

Step 2: Attribute Co: “cohesion;”

Step 3: Attribute Sa: “salience”

Step 4: Attribute Po: “polarization” ….. The most complicated to create is the polarization score/party level.

🡺 of these, the most difficult one is step 4, polarization at the individual party level

***Step 5: “Normalizing” the Parties’ Scores on the Attributes***

* We discussed this extensively in October 2015, see attachment 1 with these e-mails;

***Step 6: First Stage “Aggregation” of scores: Individual Parties’ Issue Programmatism.***

Create issue i based “programmatism” variables as multiplication of 0-1 standardized programmatism attitudes for each party j in country k:

CoSal\_ijk(n) = Co X Sal (mnemonics: cohesion X salience for issue i in party j, either with

means-corrected or with raw dataset)

CoSalPo\_ijk(n) = Co X Sal X Po (analogous mnemonics, for triple multiplication)

***Step 7: Second Stage “Aggregation:” Parties’ Cross-Issue Overall Programmatism Scores.***

This is a ***critical step where theory and mathematical logic provide little help***.

* Should issues be aggregated that generate the highest PR-2x or PR-3x products ***for each party*** or ***for the country***, on average?
* PR-2xPjk (mnemonics: Pr for program 2x for multiplication of two attributes of party Pj in country k)
* PR-3xPj (mnemonics: analogous for triple interaction of partyj in country k);
* Should just the issues with the highest PR scores be aggregated or should substantive constraints be imposed on the choice of aggregation, e.g. restrictions on the N of issues that are from individual country questions (d7-19), or scope imperatives requiring that issues integrated have to come from different theoretically apriori dimensions of preference/policy variance, as in 3G (grid: d1-3; group: d4; grid: d5), supplemented by country-specific questions that may “load” on any of the 3G? Creating multiple different indices?

***Step 8: Third Stage Aggregation: Countries’ Cross-Party Overall Programmatism Scores***

Weighted means of the individual party CoSalPo Scores.

🡺 different from Kitschelt/Freeze 2010 maybe in two regards:

1. The systemic polarization index is a means of the individual-level party polarization scores. In KF 2010, a separate systemic polarization score was calculated, the p21/electoral size-weighted standard deviation of the individual parties’ mean scores on each policy issue d1 … dn.
2. The issues selected for CoSalPo aggregation across parties were not taken from INDIVIDUAL parties’ top programmatism issues, but NATIONAL AVERAGE most programmatic issues.

**Step 1: Means-Correcting the experts’ judgments**

to get country-based anchor points out of the picture, or at least make them less relevant. Reasonable minds might disagree…. So let’s produce two datasets:

**DATASET A**: all subsequent calculations based on “means” data.

**DATASET B**: all subsequent calculations based on “natural, raw” data, as supplied by the expert scores.

**Step 2: “Cohesion” of issue i for party j in country k**

**The critical steps are outlined in the Kselman memo 6/2015, but have to be supplemented by a few further operational rules:**

* **What to do, if standard deviations have to be calculated with N <= 5 expert observations?** Then almost inevitably these sd’s become volatile and sometimes hugely large. Moreover, given that there are so few responses, it’s likely to indicate that the issue is irrelevant for this party.

***RULE 1: If N of judgments 5 or less, then code ‘cohesion’ as ZERO (0).***

* **In the cohesion index, with a scale running from 0 to 10 for expert judgments, high cohesion yields small numbers** (with sd’s around or less than 2.0)**, low cohesion larger number** (perfect 0-10 randomness yields an sd ~3.3, extreme polarization of 10 respondents in 5 responses with score 0 and 5 responses with score 10 yields an sd ~5.3, depending on the N of respondents)

**🡺 let the attribute scoring, and especially later the normalization (step 5) not be driven by a few extreme outliers that could generate extreme sd’s (e.g. > 3.5);**

***RULE 2: IF a raw computation of sd yields a score of >3.5, recode it as 3.5.***

* **For the later construction of programmatism scores, to be intuitive, so that greater programmatism yields higher scores, while low programmatism yields lower scores.**

***RULE 3: Recalculate all cohesion scores as 3.5 – x, with x being the actual score of the standard deviations.***

**Note that this will make all CoSal and CoSalPo scores that contain a 0-cohesion score 0 altogether. But that’s analytically intended, as parties with positions “all over the map” should not display any score of cohesion at all.**

**From Kselman 6/2015 memo: Cohesion**

This measure is designed to capture the *Cohesion* of expert scores for  on each of the issues. More specifically, we will capture cohesiveness of expert scores by using their standard deviation from the party’s mean score on that issue across all experts. Formally, for each party  on programmatic issue , we need to generate a standard deviation of expert scores:

,

where  is the individual expert’s placement and  is party ’s mean position on ideological dimension . One important note about the denominator: it should include ***only the number of experts who actually score a party on dimension ***; conversely, the count should not include experts who left the item blank (missing data) or who chose ‘Don’t Know’ or ‘Party has no Clear Position’.

 can be generated using the ‘collapse’ command in Stata. The most straight-forward, brute force way of doing this is in two-steps. The first step will be generating the expert mean . For example, begin with issue position D1, and use the collapse to command to generate a new data set with the 506 party means across on experts on this issue position. Then, use the ‘merge’ command in Stata to reintroduce this new variable back into the raw data set, so that each raw observation has both the expert’s score  and the party’s average score across experts . Finally, use these two measures to generate the summed of squared distances of each expert is from the mean, and then once again collapse the data set down to 506 by party to generate the ‘mean’ squared deviation (i.e. standard deviation) for each party.

A more elegant, less labor-intensive approach would use more advanced coding in with the ‘collapse’ command to have Stata generate , the standard expert deviation, with a in a single ‘collapse’ step. This would be quicker and more painless, but also requires a bit more coding comfort in Stata.

Both options will yield N=506 observations of  on survey item D1. Please repeat this process for all of the 5-N dimensions in a country, as well as the ‘left-right’ index DW. Each party will have at least 6 scores, the 5 common to all countries and the ‘left-right’ dimension which is also common to all countries.

**It is important to be cautions in the collapsing process, and to do some brute force double checking of the work. Especially common are mistakes with either missing data or answers such as ‘Don’t Know’. The key is that Stata not treat these observations as ‘0’s, nor anything else but missing data.**

We will also need to repeat this process using a ***mean-correction procedure*** on expert scores. This involves first calculating , the average score expert  assigns to all of a country’s parties on a particular issue . To do this use the same ‘collapse’ command in Stata to generate the mean value  that expert *E* assigned to parties on issue . Then, as above, merge this variable mean back into the raw data set such that each expert/party observation has the measure , the expert’s [placement of partyon issue , as well as the expert’s average placement of parties on that issue . Finally, generate a new variable by subtracting the second form the first, i.e.  - , yielding the expert’s ‘mean-corrected’ placement of party on issue .

Once you have these mean-corrected scores, you will repeat identically the process described above to generate the standard deviation, except in this case using mean-corrected expert scores instead of their raw scores .

**Step 3: “Salience” of issue i for party j in country k**

**Kselman memo:**

Here we are looking for the percentage of experts who chose to score a party on any particular issue variable. *The choice to not score a party on a particular measure* can take one of three forms:

* first of all ***any missing data*** for party  on issue  is considered not scoring a party; Iss
* second of all choosing the ***option ‘Don’t Know’*** on an issue  is considered not scoring a party;
* third choosing ‘***Party has no Clear Position’*** is considered not scoring a party on issue .

To generate the % of experts which choose to score a party, create for each issue dummy variable which codes whether an expert gave the party a score (assigned the value 1) or did not score the party (assigned the value 0). Then, to create the SAL score for the 506 parties, simply collapse these new dummy variables by party to generate party means, i.e. the total number of 1’s as a % of all experts on the issue. This will yield new SAL variables, which (once again…) not all parties will be score on.

**Implication of the Kselman memo:**

* Each policy issue i (d1, d2….dn) will have a score by party/country: SALijk
* There will be *as many salience attribute variables as there are issues i in a country*.
* Issue dimensions on which experts were NOT ASKED to score parties in a country, typically some or all of the d7, d8… and higher variables, will have ***no salience attribute variables***. 🡺 delete

**Step 4: “Polarization” of issue i for party j in country k**

This variable is clearly the hardest to generate at the party level.

The basic idea is that distances of policy positions on issue i are calculated for all party dyads. Each party j is involved in j-1 dyads. Each country has j \* (j-1)/2 dyads) (n = observed parties).

Because some countries have more parties (dyads) than others, we need to divide the party dyad scores by the number of parties (j-1/j) to obtain a party j’s polarization score.

**Compared to cohesion and salience, the peculiarity of the polarization score is that it depends not just on the position of the individual party j on issue i, but also on that of all the other parties’ positions on i in country k. It is a *relational score*, only partially controlled by the choices of the individual party.**

**I will propose below another NEW scoring technique based on the “eccentricity” (see Laver and Sergenti 2011) of a party, that is also still somewhat relational, but less starkly than the polarization score.**

A tricky question is the **weighting** of small and large parties in each dyad, especially since extreme parties that potentially contribute a great deal to polarization scores tend to be small:

* Does a moderate distance between two large parties contribute the same to each party’s overall summed-up polarization score than a large distance between a small extreme party and a large moderate party?
* So should we ***weight*** the dyad distances by the sum of the dyad parties’ relative electoral size (p31, two recent elections)?
* In order to further reduce the influence of extreme outliers, should we ***log-transform the polarization scores*** for each party? Past explorations yield that ***log-transformation makes hardly any difference from the non-log scores and only adds another complication to our computation that requires justification***.

It may be theoretically adequate, therefore, to weight the distances between small and large parties in each dyad by the average electoral performance of the two parties j1 and j2 in the two most recent legislative elections (p31 j1, j2).

**Step 4.1. Two Ways to Compute the Differentiation/Polarization Scores, as Originally Conceived (Kitschelt/Freeze 2010)**

**4.1.1. Operationally Coding Differentiation/Polarization: The “simple” method (HK)**

* On each issue i for party j, use the experts’ mean position d ijk (use dataset A: mean-corrected).
* Then compute the ***distance dyads*** between a party j’s mean position and each of the other parties j’s in a party system on issue i. (j-1 dyads to compute mean polarization for party j).
* ***Divide the sum of dyad distance scores for each party j by the j-1 number of dyads*** to obtain party j’s differentiation or polarization score = POLijk.

**4.1.2. Operationally Coding Differentiation/Polarization: The “complex” method (DK)**

**C.) Generating POL**

Here we are looking for the average ‘Differentiation’ of party’s programmatic positions from the positions of other parties in the country. This should be done two ways: first calculating this at the expert level for each party, and then calculating this at the country level. ***In this first round of data reconstruction, I suggest we stick to the former***. To calculate this at the expert-level, we will want the following measure:

,

To begin, we are thus looking for the total distance expert  observes between  and all other parties on issue . To calculate this measure, create variables for each expert/party observation that capture the distance between  and some other party  on the programmatic issue. For each issue, there will thus be an additional  variables, where represents the number of parties in any given country. With these  difference variables we can then create for each expert/party observation a Total Difference on issue , which is simply the sum of the  individual distances on issue . Finally, once we have this Total Difference measure for each expert/party observation, we can collapse across parties to find the mean Total Difference (formula above).

**Step 4.2.: Eccentricity Score Rather than Polarization Score?**

Here is an alternative way to calculate what Laver would call the “eccentricity” of a political party j on an issue i in country k. This doesn’t quite tell us how much “differentiation” there is in a party system, but it tells us whether there are “extremists” of some electoral weight in the target group that is investigated:

* Create the ***mean scores of all parties j on issue i*** (from d1, d2, …dn) in country k (means corrected or raw score doesn’t matter)d ijk.
* ***Weight*** the parties’ d ijk issue positions *by each party j’s electoral size (p31)* d ijkw.
* Calculate for an issue i the mean position of ALL the parties j in polity k (d ik): the ***weighted mean issue position of the party system on issue i as a whole (d ikw)***.
* Calculate the “eccentricity” of an individual party j’s d ijk as the ***absolute distance*** of the focal party j’s position on issue i (d ijk) from the weighted mean issue position of the party system (d ikw):

**|d ijk – dikw| = eccentricity of a party’s position on issue i.**

***I ONLY MENTION THE “ECCENTRICITY SCORE” HERE AS AN ALTERNATIVE.***

***BELOW, I ASSUME THAT WE’LL ONLY USE THE DIFFERENTIATION/POLARIZATION SCORES FOR THE AGGREGATION PROCESS OF CoSalPo. WE COULD, HOWEVER, CREATE AN ADDITIONAL CoSalPo DATASET WITH ECCENTRICITY SUBSTITUTING FOR POLARIZATION.***

**Step 5: Normalization of the Attributes of Programmatic Appeal (cohesion, salience, polarization), Once Created from the Raw Data (see appendix)**

🡺 see appendix 1 for e-mail discussion DK/HK October 2-4, 2015

So we’ll create the normalization of the variable scores in the following fashion:

1. **FIXED and UNIVERSAL calculation of normalization for each ATTRIBUTE (co, sal, po): *normalization calculated in the same metric for all issues i (d1, d2 …dn) and all parties j (n=506) in all countries k.***
2. COHESION: The actual scores were ***inverted and truncated*** already (see above), and can arithmetically run from 0.0 (no cohesion) to 3.5 (full cohesion), with ***actually observed scores*** between 0 and no higher than 2.5.
3. SALIENCE: This variable is ***naturally normalized***, as the % of respondents who score a party j on issue i. ATTENTION: Since in cohesion we set parties = 0, if there are 5 or fewer respondents, making whatever programmatism score that is multiplicative =0, if cohesion is in it, the de facto range of salience scores that matters is probably between 0.4 or 0.5 (depending on what percentage of total experts in a country is accounted for by 6 respondents) to 1.0 (when all experts score a party).
4. DIFFERENTIATION/POLARIZATION: Have not inspected the actual range of the observations, when calculated according to any of the three methods, but would think, it’s somewhere between 0 and ~ 3.
5. **BOUNDED OR LIMITATIONAL SCALE? …. That’s the question for cohesion and polarization**:

* Will the normalization set 0 = lowest ***observed level*** of cohesion (not quite zero), salience (zero), and polarization (probably near zero), or the lowest mathematically feasible scale value which is always 0?
* Will the normalization set 1 = ***highest observed*** individual ijk value, or the ***highest mathematically possible value*** (i.e. 3.5 on cohesion, don't know what is mathematically highest possibility on polarization)?

**Pragmatically, I would opt for the following BOUNDED normalization:**

**Lowest value 0 = lowest observed value** … probably zero on all three attributes (co, sal, po).

**Highest value 1 = highest observed value** (probably 2.5 on co, 1.0 on sal, and ~3 on the polarization scores).

**We’ll then end up with the following disaggregate normalized dataset(s) with ijk values:**

|  |  |  |
| --- | --- | --- |
|  | **Mean corrected d1, d2 … dn scores as input** | **Raw d1, d2, …dn scores as input** |
| cohesion | Mean corrected cohesion scores for each ijkn (n=normalization) | Raw cohesion scores for each ijk |
| salience | No difference between raw and mean corrected | |
| Polarization 4.1.1: simple HK | Mean corrected simple differentiation | Raw simple differentiation (should be same as mean-corrected?) |
| Polarization 4.1.2: complex DK  …mathematically equivalent to 4.1.1.? | Mean corrected simple differentiation | Raw simple differentiation (should be same as mean-corrected?) |
| Party Eccentricity 4.2. | No difference between raw and mean corrected …but index implementation not pursued in the aggregation process | |

Now the task is to reassemble the .ijk attribute observations stepwise:

1. Step 6: Aggregation across attributes for each ijkn: Co X Sal; Co X Sal X Pol;
2. Step 7. Aggregation across issues creating jk CoSal/CoSalPo scores for each party.
3. Step 8. Aggregation across countries

**Step 6: First Stage “Aggregation” of scores: Individual Parties’ Individual Issue Programmatism**

**(PR-CoSal\_ijk) and PR-CoSalPo\_ijk)**

Step 6 is entirely straight-forward. It involves only the ***multiplication of normalized cohesion scores*** with normalized other attributes (salience; salience and polarization/differentiation), done separately with either means corrected (m) or raw data:

**CoSal\_ijkm: CoSal = Co X Sal (means corrected dataset)**

**PR-CoSal\_ijk: CoSal = Co X Sal (raw score dataset)**

(mnemonics: Pr for program 2x for multiplication of two attributes of issue i in party j, n of normalized data are being used)

**CoSalPo\_ijkm: CoSalPo = Co X Sal X Po (means corrected dataset)**

**PR-CoSalPo\_ijk: CoSalPo = Co X Sal X Po (raw score dataset)**

(analogous mnemonics, for triple multiplication)

**TOTAL OF FOUR COMPLETE DATASETS AT THE LEVEL OF PROGRAMMATISM.**

**Step 7: Second Stage “Aggregation:” Parties’ Cross-Issue Overall Programmatism Scores**

**(CoSal\_X-jkm; CoSal\_X-jk; CoSalPo\_X-jkm, CoSalPo\_X-jk)**

**CHALLENGE:** Which issues i should be aggregated (averaged) to obtain a ***summary index of a party j’s programmatism in different countries k***?

* **Theoretical Consideration 1:**  A party cannot be scored as highly programmatic, if it’s a sharply contoured appeal on just ONE issue. It’s a niche party then. It’s likely to suffer tremendous programmatic diffuseness in its practical operation in legislatures, given that there are many issues on which its representatives have to vote that are dealing with other, often orthogonal issues that nevertheless affect the quality of life and interests of its supporters profoundly. *Programmatism doesn’t depend just on what parties ‘want’ to represent, but also on what they are ‘forced’ to represent, given that they are not autonomous in political agenda setting.*

**RULE 1: The relevant programmatism index to assess individual parties must incorporate more than one issue dimension. If there is a single over-riding issue, however, its “weight” can still be coming to the fore by limiting the range of issue dimensions that go into the ijk programmatism score 🡺 have 3 or 4 issues, but no more.**

* **Theoretical Consideration 2: How shall the multiple issues that contribute to a party’s composite programmatism score be selected?**

Here there are several conceivable logics of programmatic political appeal, and each of them should be embodied in a CoSal/CoSalPo ijk index to explore the robustness of arguments based on programmatism.

* Simply take the ***four issues d1, d2…dn for each party that yield the highest programmatism scores (“MAX”)***, ***regardless*** of whether they were in the “core” d1-5 schedule of issues scored by experts in all countries, or the country-specific questionnaires.

**Scoring Instruction:** For each party j in country k, take the four (4) issue dimensions d1 … dn that yield the highest CoSal/CoSalPo scores (“Max”) and calculate their averages, yielding indices notated in the following way:

**CoSal\_Max-jkm (means corrected data based)**

**CoSal\_Max-jk (raw data based)**

**CoSalPo\_Max-jkm (means corrected data based)**

**CoSalPo\_Max-jk (raw data based)**

* One may argue that because parties do not control the issue agenda of legislatures and face great uncertainty as to the issues that will come up and on which they will have to take a stance, ***the party-level index of programmatism must take all conceivable issue dimensions (3G: greed, grid, group) into account***, yielding a minimum of three issue components. In order to recognize the individual strategic approach of each party, emphasizing some issues more than others, a fourth issue then should be added, selecting whatever is the highest scoring ijk programmatism issue still available, once the first three have been incorporated. This yields the familiar CoSal/CoSalPo\_4.ijk index.

**Scoring Instruction:** For each party j in country k, take

1. the highest scoring of the three GREED scales d1, d2, d3
2. the party’s GRID scale programmtism d5
3. the party’s GREED scale programmatism d4
4. whichever fourth issue—whether one of the untapped GREED scales or any of the country-specific scales d7, d8 … dn.

Then calculate the average programmatism score of these four issues, yielding

indices notated in the following way:

**CoSal\_4-jkm (means corrected data based)**

**CoSal\_4-jk (raw data based)**

**CoSalPo\_4-jkm (means corrected data based)**

**CoSalPo\_4-jk (raw data based)**

* Conversely, one might argue that in all political conflicts the overriding dimension of political struggle is “who gets what when and how.” As a consequence, what ultimately matters primarily is where parties stand on questions of distributive conflict. As a consequence, we should only take GREED issues into account in the construction of the parties’ overall programmatism scores. If economic-distributive programmatism of parties and party systems is low, then they are simply not programmatic….

**Scoring Instruction:** For each party j in country k, take the parties’ programmatism ijk scores ***on the three economic GREED variables (d1-d3)*** to calculate their overall programmatic appeals, yielding indices notated in the following way:

**CoSal\_Econ-jkm (means corrected data based)**

**CoSal\_Econ-jk (raw data based)**

**CoSalPo\_Econ-jkm (means corrected data based)**

**CoSalPo\_Econ-jk (raw data based)**

**Step 8: Third Stage Aggregation: Countries’ Cross-Party Overall Programmatism Scores, Weighted by Party Size (p31)**

**(CoSal\_X-km; CoSal\_X-k; CoSalPo\_X-km, CoSalPo\_X-k)**

This step is again very straight-forward. We want to know what the “average” programmatic appeal of parties is within a polity k. Of course, small parties should matter less in determining that average. Therefore, we are weighting by parties’ electoral size (p31).

For each country, the average CoSal or CoSalPo index value is the p31-weighted average of all the scored parties’ CoSal or CoSalPo indices, yielding four indices (N=88 observations) for each of the three issue composites going into the Max, 4 and Econ indices:

**CoSal\_Max-km (means corrected data based)**

**CoSal\_Max-k (raw data based)**

**CoSalPo\_Max-km (means corrected data based)**

**CoSalPo\_Max-k (raw data based)**

**CoSal\_4-km (means corrected data based)**

**CoSal\_4-k (raw data based)**

**CoSalPo\_4-km (means corrected data based)**

**CoSalPo\_4-k (raw data based)**

**CoSal\_Econ-km (means corrected data based)**

**CoSal\_Econ-k (raw data based)**

**CoSalPo\_Econ-km (means corrected data based)**

**CoSalPo\_Econ-k (raw data based)**

**9. Methodological Self-Criticism of the Programmatism Indices**

**9.1. Measurement Level**

The algebraic operations creating the programmatism indices assume naively or force (by normalization) the presence of a very high measurement level, namely metric scaling, where additions, subtractions, multiplications, and divisions are meaningful.

If we assume more conservatively that experts’ measures are at best ordinally scaled, none of these indices make any sense.

RESPONSE? None, really. Maybe: The proof of the pudding is in the eating…If one can generate interesting results, it’s plausible that the measures represent something real (“construct” validation).

Maybe others can use the dataset and create more sophisticated, refined indicators that assume more modest measurement levels. I doubt it though…

**9.2. Anchor Point Problem: Comparability across Countries and Parties**

Each individual expert judge, or at least each country’s expert judges, may have different referents in mind when they try to assess the substantive meaning of the endpoints of the d1, d2 …dn issue scales and then score parties relative to each other. Because anchor points differ across judges and settings, the resulting scores are not strictly comparable.

The critique is correct, and I am incompetent when it comes to scaling theory and strategies to deal with “differential item functioning” (DIF) (Aldrich/McKelvy 1975) across contexts. This is a fiendishly complex field (see work by Saeigh). I just hope that results will stand up, even if probed by methodological experts.

….and there are too many other sources of measurement error and inferential error. Shall we really worry too much about DIF? …. Defense:

1. means-corrected expert scores address one source of anchor point variability;
2. the comparisons often involve within-country differences (e.g. polarization) across issues or countries; scores of parties are then not directly compared to each other, but set in context.
3. Contextual explanations are a sort of theoretical exploration of anchor point consequences on party assessments.

**9.3. Ceiling Problem in Computing Index Values**

As we know from Cees Van der Eijk’s work, a ***measure of programmatic cohesion*** based on the simple calculation of standard deviations suffers from ceiling effects. He encourages scholars, therefore, to implement (Cees Van der Eijk 2001) measure of agreement A1 in order to get around this (see my memo of March 18, 2015, p. 2). Of course, this would add another complication, with possibly limited payoffs, see pp.2-3 of my memo re Somer-Topcu’s corrections.

**APPENDIX STEP 1: Normalizing the Experts’ Party Policy Scores**

**(2015)**

**Memo HK 2015/10/02:**

Dan,

let me run this by you:

The critical question appears to be *how to determine the end-points of the normalized 0-1 scale*.

* In "our" (Freeze and myself) version, the 0-1 scale is *FIXED for all issues/parties and BOUNDED:* there can be some parties that have issue cohesion lower (higher) than what is defined as the minimum on the normalized scale (the lowest mean cohesion score for all issues of a party).
* *Moreover, I believe "our" score is UNIVERSALLY calculated OVER ALL 506 PARTIES/ALL COUNTRIES.* With "your" (DMK & Haohan), the 0-1 normalization is *distinct for each issue dimension d1*…., so the numerically same "raw," pre-normalized issue score of a party j on d1 and d2 may translate into a different normalized score, depending on what the minimum and maximum party cohesion score on that issue in a specific country is. Moreover, I am not sure whether you determine "minimum" and "maximum" over the entire set of N=506 parties, or by individual country?

So, in the mathematical operation called "normalization," three decisions have to be made:

(1) **FIXED or CONTINGENT 0/1 scale**: Will each issue have the same or a different scale?

(2) **PARTICULAR or UNIVERSAL 0/1 scale**: Will the normalized scale be calculated over the parties of a single country or all countries?

(3**) BOUNDED OR LIMITATIONAL SCALE**: Will 0/1 defined as the lowest (highest) MEAN cohesion score of a party over all issues or simply as the LOWEST (HIGHEST) score *any* party reaches over all issues? In the first calculation, a party may have a cohesion score on a specific issue d that is lower than the lowest mean score over all issues, yet this difference will not be registered in the normalized scale and

Now, upon reflection, I come to the following views, and please give me a reality check here:

There should be a *universally common yardstick* (I.e., interpretation of the 0/1 extremes) across all parties and all issues. So a specific "raw cohesion score," like that of 1.744 that both Kent and Haohan generate for the standard deviation of the Turkish AKP on d2, regardless for whatever party and on whatever issue in whatever country, should always translate into the same normalized cohesion score. So the scale should be FIXED and UNIVERSAL.

What is debatable, however, is whether the normalization should be *bounded or limitational*. The bounded strategy resulted from the concern *not to let the 0/1 normalization be driven by strange outliers.* But I don't even know whether this is empirically a problem. And the *boundedness* of the measure *loses information* (by setting everything below (above) the lowest mean cohesion score of any party over all the issues equal to 0 (1)). But if we don't do this in a bounded way, we run into the following problem, as *Kent points out in one of his memos: The standard deviations may be extremely volatile and very large, if we have very few respondents. The scale limitation must then be set to apply as the lower bound to parties where at least 5 or 6 expert scores are available … and even then there may be some weird outliers yielding very high standard deviations*.

==> The fact that Haohan's calculations tend to be higher than Kent's may have to do with the fact that the limitational low point = 0 (defined by the highest extreme of standard deviations) is lowered by such outlier parties with few respondents, awarding scores to opposite extremes, and thus driving the raw deviations to extremely high scores above 3.0, thus squeezing most parties closer together in the upper range of cohesion.

SO, WHAT WE HAVE, APPEARS TO BE THE FOLLOWING:

Freeze/Kitschelt 2010: normalization is fixed (all dimensions) and universal (all parties/all countries), but bounded.

Kselman/Chen 2015: normalization is contingent (each dimension separately) and either particular or universal (can't tell: by country or for all?), but universal (no correction for outliers).

WHAT I THINK IS THE CORRECT WAY TO GO, MAY BE THE FOLLOWING "hybrid" between Freeze/Kitschelt and Kselman/Chen:

The normalization should be *fixed across all issue dimensions*: the 0/1 normalized end points should be transformations of the same raw standard deviations.

The normalization should apply the *same scaling to all parties and all issues across N=506 parties in N=88 countries.*

More debatable: Should the normalization be *bounded or limitational?* I think this has to be contingent upon whether we find a different way to "solve" the outlier problem than the mathematical path chosen by Kent (take as endpoints lowest (highest) AVERAGES of raw cohesion scores across issues for parties, not lowest scores for individual issues by party.

For example, could one log-transform the raw standard deviations, thus making extremely high cases less extreme? Would this be better than Kent's approach?

In any case, parties/issue standard deviations based on fewer than a minimum of 5 experts should not be entered to compute the scales. With so few respondents, it's anyway futile/irrelevant to calculate cohesion, if we employ cohesion only in one of our interactive formulations (CoSal, let alone CoSalPo). In these, the low N of responses would immediately show up in a very low salience score, regardless of how high or low the cohesion (polarization) term of the multiplication would be.

I don't want to send this communication (yet) to Haohan or Kent for their review, as I might simply mathematically misunderstand from the cryptic description you cite in your 2015-08-07 memo from Freeze's and my paper(attached here so that you don't have to dig around in your computer to find the file), conveying what we have done, or the more transparent normalization procedure you lay out in that very same memo

So, please think about these issues, as they are critical: Without resolving this, we can't take a step forward.

And, again, it appears to me that what Kent and I generated has a great deal more empirical plausibility, and it may, in fact, be driven by the decision to "rein in" on extreme outliers by (1) normalizing over ALL issue dimensions, (2) over all countries/parties and (3) by eliminating the influence of extreme outliers through boundedness….

Best,

Herbert

**RESPONSE DK 2015/10/04:**

Dear Herbert,

Thanks for this, and for leading the charge getting this data into a shape where we can begin confidently finalizing our papers for submission!

After reading your email carefully, I agree on the FIXED and UNIVERSAL criteria: our final indicator should normalize ACROSS dimensions as well as ACROSS countries. Indeed, all of the indicators that Haohan and I created were UNIVERSAL (normalized across countries).

Regarding the cross-issue question, what I in fact asked Haohan to do was generate data using both FIXED and CONTINGENT normalizations (which are now both in the data set), and was hoping that Haohan's FIXED normalizations would match Kent's. I now see that the issue of BOUNDEDNESS may be generating the differences between Haohan and Kent's FIXED normalization scores.

My hunch would be that we ought to generate both non-BOUNDED and BOUNDED scales. The latter, which Kent generated, behave nicely in analyses...but the particular endpoints chosen could be criticized as 'ad hoc'. So, I would suggest that first Haohan regenerate Kent's scores (so that we satisfy the replicability criterion) but that he then also generate unbounded measures.

Another obvious issue will be which dimensions we choose to include in the aggregate indices. Here again I like the way Kent's measures behave, but also worry that the selection criteria documented in your original memo could be criticized as 'ad hoc'...though I understand their logic: we do not want to generate differences which are driven by whether a country had extra, country-specific issues included in its survey. Here again I would generate BOTH Kent's original measure AND an unfiltered measure which takes a country's top 4 scores on ANY of the possible dimensions (even if they all happen to country-specific...).

Finally, please tell Haohan to include all 506 parties ENTIRE NAMES in the data set...his coding in 'R' cut some of the names short, which means that in certain countries (e.g. Colombia...) it is impossible to tell the parties apart (because their names all begin with the same 8-9 letter...).

Also, if possible it would be great it Haohan could order the observations in the same way as found in party-level 21. We want to try and avoid if at all possible uncomfortable merging and sorting processes when combining data sets.

I am trying to finish my primaries paper this weekend and next, and then can turn to finishing these papers.

Best,

dan