Online appendix

In this appendix I further compare the proposed measure of elite collective action to other measures of elite influence in the literature. I first compare the data they use, and then their parametrization.

Table 1A summarizes the different variables that each of the measures of elite influence use. For the ECA measure, I focus only on the power-sharing dimension. This is the most readily comparable quantity to existing work, since past research has not focused on the cohesion of the elites. There is substantial overlap when it comes to power-sharing. ECA and GS both use Svolik's coding of military involvement in government. GWF use a somewhat similar coding to capture whether the sitting autocrat controls the military or if the military functions as an independent force. The same holds for the second row, where GS and ECA use essentially the same variable capturing whether the autocrat owes its position to a collective or not. GWF focuses less on the origins of the dictator and more on whether rulers currently control the party. The next couple of variables in all measures try to capture the independence or power of the party relative to the executive. ECA and GS also overlap in their attention to the cabinet. Particularly, on whether the autocrat has shuffled members of the cabinet. GWF somewhat similarly consider whether members can climb the party ladder by their own means or if they depend on the whims of the dictator.

Key areas of difference relate to legislatures and purges. GS and GWF do not consider legislatures as potential power-sharing spaces. For some cases, like some communist countries or autocracies with rubber stamp legislatures, this omission is likely justified. However, in some cases legislatures do behave as genuine counterbalances. The RI model ECA uses allows it to consider legislatures without making the assumption that they are always (or never) spaces of power-sharing. Purges, military or otherwise, are seemingly clear indicators of a lack of power-sharing. However, while purges do diminish power-sharing, research is less clear about what is it that purges tell us about the overall level of elite power. Some dictators may purge only when they know the elite is weak enough to not be able to resist it, while others may do it when they fear the elite has amassed too much influence (Sudduth 2017). Without clear theoretical expectations, ECA opts to not consider purges as an indicator of power-sharing. Moreover, by excluding purges from the measure of power-sharing, it is possible to assess the empirical relationship between the two without circularity.

Table 1A: Comparison of power-sharing/personalism indicators

ECA Power-sharing	GS Personalism	GWF Personalism
Military involvement	No military involvement	Executive controls security apparatus
Executive is selected	Executive is not a collective	Executive controls party
Single party	No regime party	Party serves as rubber stamp
Hegemonic party	Party created by executive	Party created by executive
Presence of a legislature	Multiple parties	Loyalty-based access to party
Stable cabinet	Cabinet change	Loyalty-based promotion in party
Stable legislature		
Autonomous legislature		
Local government		
	Military purges	Military purges
	Civilian purges	
	Executive's family in power	
	Executive has no title*	
	Dictator before transition	
		Executive controls paramilitaries

*Gandhi and Sumner include three variables: executive has not official position, executive has two positions, executive has three official positions.

Comparison of standard 2-parameter and random item IRT models.

This section goes presents in further detail the differences between the two parameter IRT model (2PL) and the random items model (RI). The 2PL model is the most common in political science, and is also the one used in current measures of autocratic elite influence (Gandhi and Sumner 2020; Geddes, Wright, and Frantz 2018, 82). The RI model is a more general version of the 2PL in that it does not assume measurement invariance. In other words, it recognizes that the items perform differently in different contexts. The traditional 2PL model specifies the probability of a positive response by individual *i* to item *k* as:

$$P(y_{ik} = 1|\theta_i, \alpha_k, \beta_k) = \left(1 + exp(\alpha_k\theta_i - b_k)\right)^{-1}$$

Where $\theta_i, \alpha_k, \beta_k$ denote the latent ability, the item discrimination, and the item difficulty, respectively. The model is usually estimated through a latent continuous variable Z_{jk}.

$$Z_{ik}|y_{ik}\ , \theta_i, \alpha_k, \beta_k \sim \Phi(\alpha_k\theta_i - \beta_k,\ 1) \quad \text{Eq. 1}$$

The random items model lets α_k β_k to vary by group *j*. Equation 1 then becomes

$$Z_{ijk}|y_{ijk}\ , \theta_i, \alpha_{jk}, \beta_{jk} \sim \Phi(\alpha_{jk}\theta_i - \beta_{jk},\ 1) \quad \text{Eq. 2}$$

where $\alpha_{jk} = \alpha_k + a_{jk}$ *and* $\beta_{jk} = \beta_k + b_{jk}$. Additionally, the RI model also treats the latent abilities as the result of a fixed and random component such that $\theta_i = \mu_\theta + \sigma_\theta\epsilon_\theta$. As reported by Fox (2010, 200), the mean of the latent continuous variable is then:

$$\mathrm{Z_{ijk}} = (\alpha_k + a_{jk})(\mu_\theta + \sigma_\theta\epsilon_\theta) - \left(\beta_k + b_{jk}\right) + \epsilon_{ijk}$$

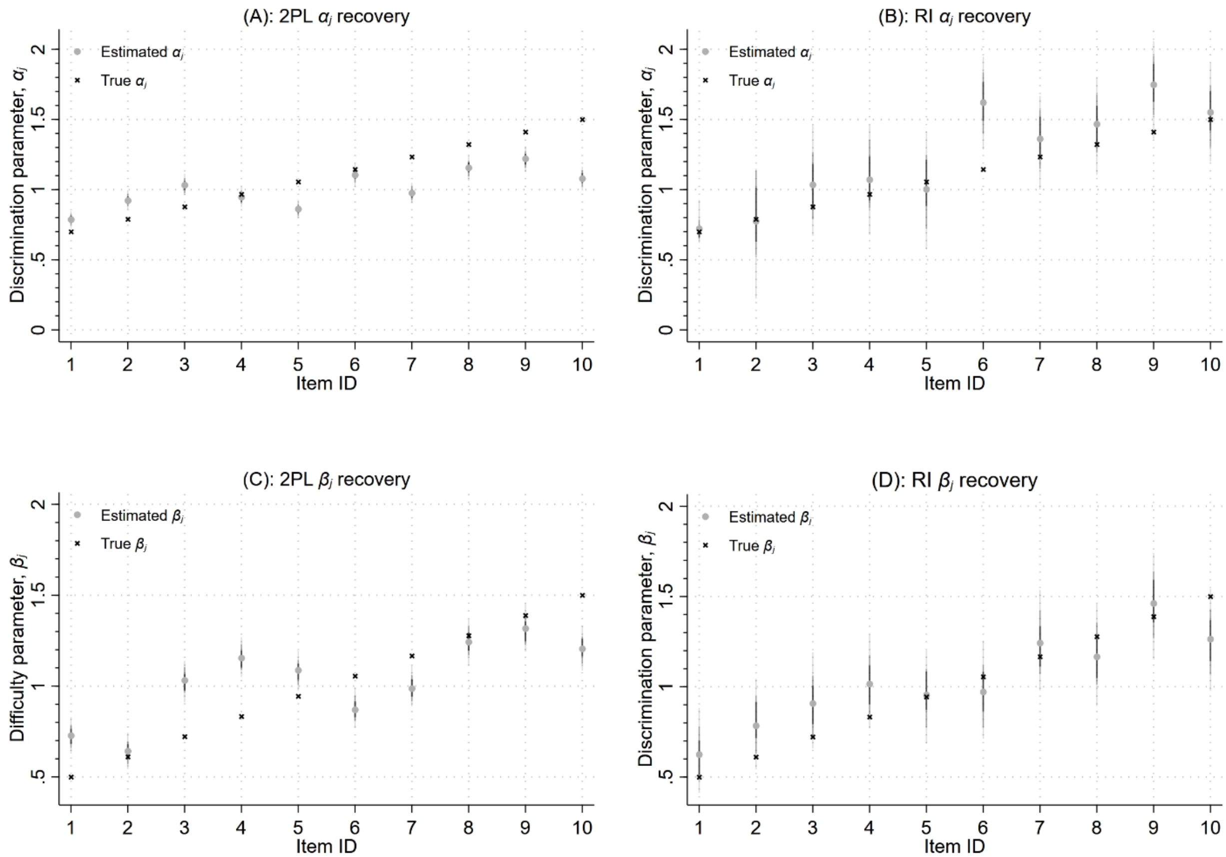
$$Z_{ijk} = \alpha_k\mu_\theta - \beta_k + \alpha_k\sigma_\theta\epsilon_\theta + a_{jk}\sigma_\theta\epsilon_\theta + a_{jk}\mu_\theta - b_{jk} - \epsilon_{ijk} \quad \text{Eq. 3}$$

The results reported in the manuscript and in this appendix reflect hierarchical versions of the 2PL and RI models. That is to say that the ability parameters have a hierarchical structure so that regime-year observations are nested in regimes (Fox, 2010, 203). While the hierarchical version of the 2PL model is not standard in the literature, I use it here to make it more comparable to the RI model.

Simulation performance

The main text compared how well the 2PL and RI models recovered the item parameters α_k and β_k examining their RMSE. Figure 1A instead shows the posterior predictive checks for each of the 10 simulated items. Recall that α_k and β_k capture the mean discrimination and difficulty. In addition, the simulated data contains deviations from the mean denoted as a_{jk} and b_{jk}. For the first few items, a_{jk} and b_{jk} are close to zero but get farther from it for the later items. That is, the later items have more variance across countries. Panel A shows that the 2PL model approximates the items with small a_{jk} and b_{jk} well enough. However, the later half of items the model performs consistently worse. In contrast, Panel B shows that the RI items performs well regardless of the degree of item variance. When item variance is low, the RI model gives similar answers to the 2PL model, but it gives more accurate answers as item variance increases. The difficulty parameters show a similar pattern.

Figure 1A: Item parameter recovery for 2PL and RI models



Results from a simulation study with 10 items, 50 groups and 40 years. The credible intervals shown are for 90% (light) 70% and 50% (dark).

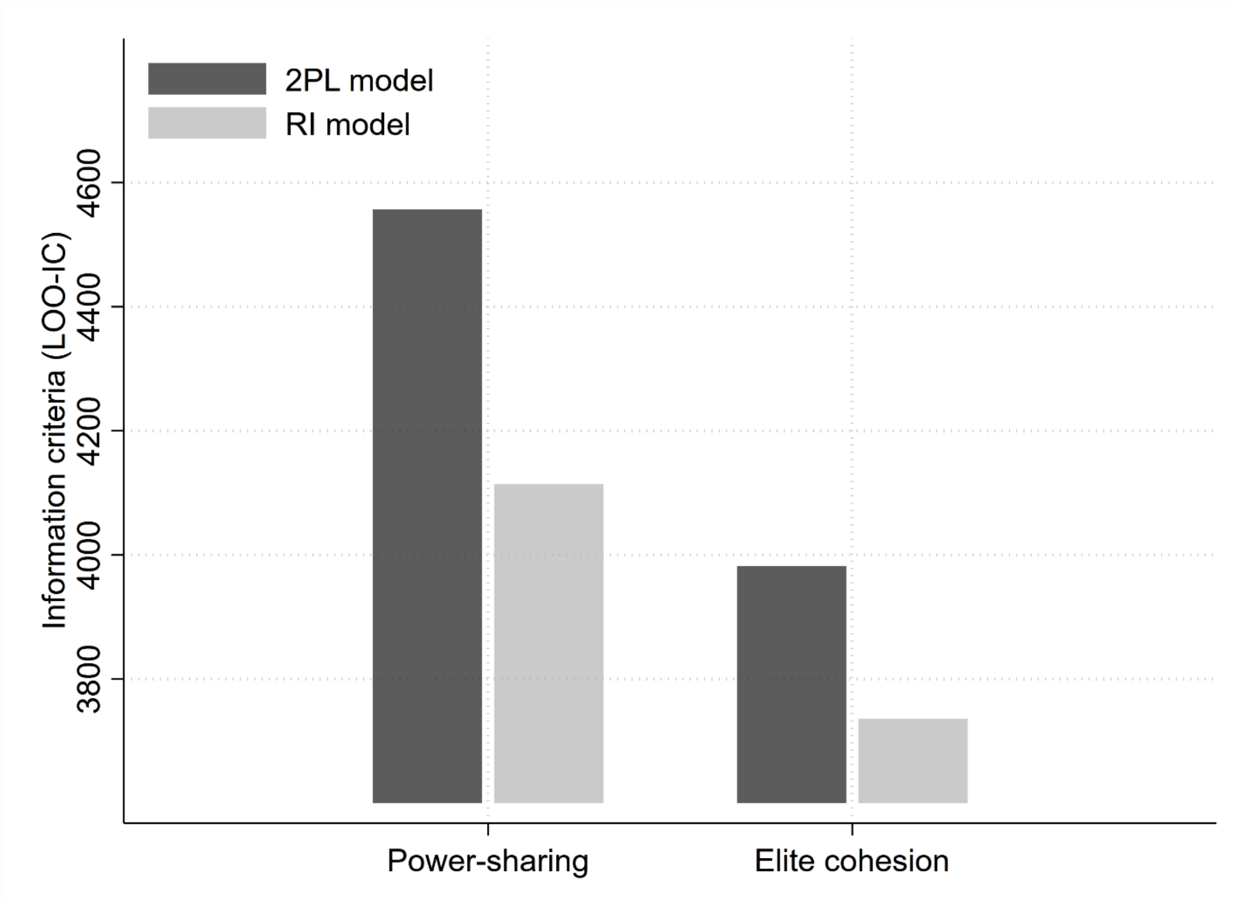
It is worth noting the much wider credible intervals of the RI model. While there is more uncertainty at the parameter level with the RI model, there is less uncertainty at the level of the prediction. This standard behavior models arises from the propagation of uncertainty from the random effects, and has the advantage of offering estimates that do not undersell the uncertainty (McElreath 2015). Nevertheless, the results do suggest that maximizing the number of repeated observations (i.e. years for each country) would be beneficial when using an RI model. The ECA measures uses covers 71 years instead of the 40 in the simulation. Hence the ECA measure should alleviate some of the uncertainty inherent with random effects models.

Performance in elite collective action measure

I now compare the performance of the ECA measure reported in the manuscript using a RI model, and an alternative specification using the 2PL model. I compare them using leave-one-out cross-validation. In this process, all the data except for one datapoint is used to train the model. Then the resulting parameters are used to predict the observation that was left out. This process is then repeated for all datapoints in the dataset. The results are summarized by leave-one-out information criteria, with smaller scores representing less error.

Figure 2A shows the information criteria for the RI and 2PL version of the ECA measure. It disaggregates power-sharing and cohesion components. In both cases, the RI version has considerably less error. A difference of four is usually taken as evidence of better fit (Vehtari, Gelman, and Gabry 2017). In this case, the difference is 236 and 443 for power-sharing and elite cohesion, respectively. This suggests that the RI model matches the data better than the 2PL model.

Figure 2A: Error in recovery of latent dimensions



Stan code

The models mentioned here were fitted within the R package for Stan (Carpenter et al. 2017). The following is the Stan code for the 2PL and RI models.

2PL model

```
data{
 int<lower=1> K;             // # items
 int<lower=1> J;             // # countries
 int<lower=1> JK;            // # of random effects
 int<lower=1> I;             // # country-years
 int<lower=1> N;             // # observations
 int<lower=1> kk[N];         // index for items
 int<lower=1> jj[N];         // index for countries
 int<lower=1> ii[N];         // index for country-years
 int<lower=1> jjkk[N];       // index random effects
 int<lower=0> y[N];          // response variable
 int<lower=1> nn[N];         // index for observations
 int<lower=1> jb[I];         // index for countries for beta parameter
}
parameters{
 vector [I] e_ij;
 vector [N] e_ijk;
 vector[J] mu_0j;               //Spread for mean country abilities
 vector<lower=0>[K-1] free_a_k;   // mean discrimination  of item k
 vector<lower=0>[K-1] free_b_k;   // mean difficulty  of item k
}
transformed parameters{
 vector [I] theta_ij;
 real mu_theta;
 real<lower=0> sigma_theta;
 vector[K] a_k;                // mean discrimination  of item k, complete vector
 vector[K] b_k;                // mean difficulty  of item k, complete vector

 a_k[1:K-1]=free_a_k;
 a_k[K]=1/prod(free_a_k);        // Constrain a_k to multiply to 1 for identification
 b_k[1:K-1]=free_b_k;
 b_k[K]=1/prod(free_b_k);        // Constrain a_k to multiply to 1 for identification
 mu_theta=0;                   // Fix scale of theta, mean=0
 sigma_theta=1;                 // Fix scale of theta, sd=1
 theta_ij[ii]=mu_0j[jj]+ e_ij[ii];
}
model{
 vector[N] eta;
 e_ij~normal(0, 1.5);
 e_ijk~normal(0,1);
 mu_0j~normal(0, 1.5);
 a_k~normal(0,1);
 b_k~normal(.5, 1);
 for (n in 1:N){
  eta[n] = a_k[kk[n]]*mu_theta-b_k[kk[n]]+
  a_k[kk[n]] *sigma_theta .*theta_ij[ii[n]]+
```

```
   e_ijk[n];
  }
  y ~ bernoulli_logit(eta);
}
generated quantities{
  vector[N] eta;
  vector[N] log_lik;        //Log likelihood to estimate deviance and LOO-IC
  real dev;
  dev = 0;
    for (n in 1:N){
   eta[n] =  a_k[kk[n]]*mu_theta-b_k[kk[n]]+
   a_k[kk[n]] *sigma_theta .*theta_ij[ii[n]]+
   e_ijk[n];
    log_lik[n] = bernoulli_logit_lpmf(y[n] | eta[n]);
  }
  dev = dev + (-2)*bernoulli_logit_lpmf(y | eta);
}
```

Random items model

```
data{
  int<lower=1> K;            // # items
  int<lower=1> J;            // # countries
  int<lower=1> JK;           // # of random effects
  int<lower=1> I;            // # country-years
  int<lower=1> N;            // # observations
  int<lower=1> kk[N];        // index for items
  int<lower=1> jj[N];        // index for countries
  int<lower=1> ii[N];        // index for country-years
  int<lower=1> jjkk[N];      // index random effects
  int<lower=0> y[N];         // response variable
  int<lower=1> nn[N];        // index for observations
  int<lower=1> jb[I];        // index for countries for beta parameter

}
parameters{
  vector [I] e_ij;
  vector [N] e_ijk;
  vector[J] mu_0j;               //Spread for mean country abilities
  vector<lower=0>[K-1] free_a_k;    // mean discrimination  of item k
  vector<lower=0>[K-1] free_b_k;    // mean difficulty  of item k
  vector[K-1] e_a_kj_raw[J];      // discrimination of item
  vector[K-1] e_b_kj_raw[J];       // difficulty of item

}
transformed parameters{
  vector [I] theta_ij;
  real mu_theta;
  real<lower=0> sigma_theta;
  vector[K] a_k;                 // mean discrimination  of item k, complete vector
```

```
 vector[K] b_k;                 // mean difficulty  of item k, complete vector
 vector[K] e_b_kj[J];
 vector[K] e_a_kj[J];


 a_k[1:K-1]=free_a_k;
 a_k[K]=1/prod(free_a_k);        // Constrain a_k to multiply to 1 for identification
 b_k[1:K-1]=free_b_k;
 b_k[K]=1/prod(free_b_k);        // Constrain b_k to multiply to 1 for identification

 for (j in 1:J) {             // Contrain random effects to sum to 1
  e_a_kj[j] = append_row(e_a_kj_raw[j], -sum(e_a_kj_raw[j]));
 }
 for (j in 1:J) {
  e_b_kj[j] = append_row(e_b_kj_raw[j], -sum(e_b_kj_raw[j]));
 }

 mu_theta=0;                  // Fix scale of theta, mean=0
 sigma_theta=1;                // Fix scale of theta, sd=1
 theta_ij[ii]=mu_0j[jj]+ e_ij[ii];

}
model{
 vector[N] eta;
 e_ij~normal(0, 1);
 e_ijk~normal(0,1);
 mu_0j~normal(0, 1);
 a_k~lognormal(0, .3);
 b_k~normal(1, .5);

 for (k in 1:J) {
  e_b_kj[k] ~ normal(0, 1);  // prior on  beta random effects
 }
  for (k in 1:J) {
  e_a_kj[k] ~ normal(0, 1);  // prior on alpha random effects
 }

 for (n in 1:N){
  eta[n] = a_k[kk[n]]*mu_theta-b_k[kk[n]]+
   a_k[kk[n]] *sigma_theta .*theta_ij[ii[n]]-
   e_a_kj[jj[ii[n]], kk[n]]*mu_theta+
   e_a_kj[jb[ii[n]], kk[n]]*sigma_theta*theta_ij[ii[n]]-
   e_b_kj[jb[ii[n]], kk[n]]+e_ijk[n];
 }
 y ~ bernoulli_logit(eta);
}
generated quantities{
 vector[N] eta;
 vector[N] log_lik;
 real dev;
```

```
    for (n in 1:N){
   eta[n] = a_k[kk[n]]*mu_theta-b_k[kk[n]]+
    a_k[kk[n]] *sigma_theta .*theta_ij[ii[n]]+
    e_a_kj[jj[ii[n]], kk[n]]*mu_theta+
    e_a_kj[jb[ii[n]], kk[n]]*sigma_theta*theta_ij[ii[n]]-
    e_b_kj[jb[ii[n]], kk[n]]+e_ijk[n];
    log_lik[n] = bernoulli_logit_lpmf(y[n] | eta[n]);
  }
  dev = -2 * sum(log_lik);
}
```

References

Carpenter, Bob et al. 2017. “Stan: A Probabilistic Programming Language.” *Journal of Statistical Software* 76(1): 1–32.

Fox, Jean-Paul. 2010. *Bayesian Item Response Modeling: Theory and Applications*. New York: Springer.

Gandhi, Jennifer, and Jane Sumner. 2020. “Measuring the Consolidation of Power in Non-Democracies.” *Journal of Politics*.

Geddes, Barbara, Joseph Wright, and Erica Frantz. 2018. *How Dictatorships Work: Power, Personalization, and Collapse*. Cambridge, U.K.: Cambridge University Press.

McElreath, Richard. 2015. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. Boca Raton: Chapman and Hall/CRC.

Sudduth, Jun Koga. 2017. “Strategic Logic of Elite Purges in Dictatorships.” *Comparative Political Studies* 50(13): 1768–1801.

Vehtari, Aki, Andrew Gelman, and Jonah Gabry. 2017. “Practical Bayesian Model Evaluation Using Leave-One-out Cross-Validation and WAIC.” *Statistics and Computing* 27(1): 1413–32.